Year : 2022

Model Uncertainty, Tail Risk and Structured Reinsurance

Vincent Léonard

Vincent Léonard, 2022, Model Uncertainty, Tail Risk and Structured Reinsurance

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_F966BE70780B1

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

Model Uncertainty, Tail Risk and Structured Reinsurance

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences actuarielles

par

Léonard VINCENT

Directeur de thèse
Prof. Hansjörg Albrecher

Jury

Prof. Rafael Lalive, Président
Prof. François Dufresne, expert interne
Prof. Andreas Tsanakas, expert externe
Prof. Tim J. Boonen, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DE SCIENCES ACTUARIELLES

Model Uncertainty, Tail Risk and Structured Reinsurance

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences actuarielles

par

Léonard VINCENT

Directeur de thèse
Prof. Hansjörg Albrecher

Jury

Prof. Rafael Lalive, Président
Prof. François Dufresne, expert interne
Prof. Andreas Tsanakas, expert externe
Prof. Tim J. Boonen, expert externe

LAUSANNE
2022

UNIL | Université de Lausanne
HEC Lausanne
Le Décanat
Bâtiment Internef
CH-1015 Lausanne

I M P R I M A T U R

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Léonard VINCENT, titulaire d'un bachelor en Economie Politique de l'Université de Lausanne, titulaire d'un master en Sciences Actuarielles de l'Université de Lausanne, en vue de l'obtention du grade de docteur ès Sciences actuarielles.

La thèse est intitulée :

MODEL UNCERTAINTY, TAIL RISK AND STRUCTURED REINSURANCE

Lausanne, le 04 juillet 2022

La Doyenne

Marianne SCHMID MAST

Members of the Jury

Prof. **Rafael Lalive**
President of the Jury,
Department of Economics, University of Lausanne, Switzerland.

Prof. **Hansjörg Albrecher**
Thesis supervisor,
Department of Actuarial Science, University of Lausanne, Switzerland.

Prof. **François Dufresne**
Internal expert,
Department of Actuarial Science, University of Lausanne, Switzerland.

Prof. **Andreas Tsanakas**
External expert,
Faculty of Actuarial Science and Insurance, Bayes Business School,
United Kingdom.

Prof. **Tim J. Boonen**
External expert,
Faculty of Economics and Business, University of Amsterdam,
the Netherlands.

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Léonard VINCENT

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: June 14, 2022

Prof. Hansjoerg ALBRECHER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Léonard VINCENT

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: 14 June 2022

Prof. François DUFRESNE
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Léonard VINCENT

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: _14 June 2022_____

Prof. Andreas TSANAKAS
External member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Actuarial Science

I hereby certify that I have examined the doctoral thesis of

Léonard VINCENT

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature:  Date: ___June 14, 2022_____

Prof. Tim BOONEN
External member of the doctoral committee

Acknowledgements

This thesis was a long journey, throughout which I had the chance to be supported and helped by great people, and I want to express them my sincere gratitude and appreciation here.

First, I am profoundly grateful to my supervisor, Professor Hansjörg Albrecher. His help and involvement made me achieve work which I am very proud of. Moreover, his curiosity, intellectual rigour, enthusiasm and knowledge made our discussions very pleasant and rewarding experiences, which have substantially influenced and stimulated me.
I would like to also express my deep gratitude to Professor François Dufresne, who hired me for this PhD thesis, and helped me all the way, with valuable advices and benevolent support.

I would like to thank my PhD committee members, Professors Andreas Tsanakas and Tim J. Boonen for their helpful advices and valuable suggestions. Professor Tsanakas deserves a special word of gratitude for having given me useful references while I was working on Chapter 2, which significantly helped me.

I am grateful to all my colleagues at the Department of Actuarial Science, and I would like to especially thank Arian, Charbel, Dina and Maïssa, for the great moments spent together.

Last but not least, I would like to thank my family and friends, for their love and support. A warm and special thanks goes to my parents Christine and Paolo, my sister Lucie, and my brother Jonas, for their encouragements and unfailing support. My love, Lysiane, deserves my greatest thanks, for all the support she provided me, and so much more.

Contents

Chapter 1

Introduction

In many aspects of our everyday lives, uncertainty is at play, and when this concerns the occurrence of an event that may have adverse consequences, one talks about risk. Once identified, a risk can be managed. This process is called risk management, and in broad terms it consists of:

(a) Quantifying the likelihood and consequences of the risk.

(b) Responding accordingly, by either avoiding, mitigating, transferring or accepting the risk.

In the insurance and banking industry, within large companies and among certain public administration sectors, this process is performed explicitly and extensively, with a variety of tools, which often includes a probabilistic analysis. The present thesis follows this wake, and contributes to topics related to both (a) and (b), with a focus on so-called tail risk.

Hereafter, we recall some concepts that will be useful in the remainder, and we summarize the main contributions of this thesis.

1.1 Risk modelling

In risk management, a probabilistic analysis usually starts with treating the risk under consideration as a random variable, say X, which one then assumes to follow some probability distribution or model $F(x) = \mathbb{P}(X \leq x)$. In the remainder, larger values of X denote more adverse scenarios, so for instance X may represent some annual insurance claim amount, the loss of value of an asset, or the severity level of a natural disaster, such as the magnitude of an earthquake or a wind-speed measure during a hurricane.

The distribution of a risk can then serve to calculate some measures of risk and other summarizing quantities (such as the value-at-risk, expected shortfall, or simply the mean, variance or skewness), which are subsequently used to take decisions on pricing, mitigation or transfer of that risk, to perform an allocation of resources (capital), or to compare the risk with some other risks.

1.1.1 Model selection

In most cases the distribution of a risk is unknown, and for determining (estimating) it one may consider a variety of sources of information, such as

- The knowledge of the random mechanism generating the risk. For example, if the risk is known to result from the aggregation (multiplication) of a sufficiently large number of independent random factors, then, under some conditions, as a result of the central limit theorem, its distribution will be reasonably well approximated by a (Log)Normal distribution.
- Some characterizing properties. For instance, if the risk is continuous (discrete) and satisfies the memoryless property $\mathbb{P}(X > x + u|X > u) = \mathbb{P}(X > x)$ for all $x, u > 0$ ($x, y \in \mathbb{N}_0$), then it necessarily follows an Exponential (Geometric) distribution. Another example of a characterizing property would be invariance of the distribution under a change of scale (i.e. scale invariance), which points towards power laws, such as the Pareto distribution.
- Some available data (observed past realizations of the risk), which may be analysed with statistical tools such as quantile-quantile or mean-excess plots, and used to fit a suitable distribution (see e.g. the books by McNeil et al. [62], Beirlant et al. [11] and Albrecher et al. [1] for details).

For illustration purposes, we recall some well-known continuous and discrete distributions in Table 1.1 and 1.2.

1.1.2 Estimated and predictive distributions

When resulting from a finite amount of data, an estimated (fitted) distribution will likely differ from the true distribution. From a risk management perspective, the potential error in the estimation of a distribution constitutes an additional source of risk (epistemic uncertainty), which can be treated as different from the original risk (aleatoric uncertainty) and shall be considered (for some background on the dichotomy between aleatoric and epistemic uncertainty, see e.g. Der Kiureghian and Ditlevsen [29], Aven and Zio [7] and Scherer and Stahl [72]). For instance, assume that limited resources must be allocated to two independent risks with identical estimated marginal distributions, but one of them has been obtained with significantly less data than the other. In this situation, the epistemic uncertainty inherent to each of these two risks evidently differs, and it is intuitive that one shall allocate unequal resources to them, despite their estimated marginal distributions being identical.

There exist several methods allowing to build a distribution that accounts for the epistemic uncertainty relative to a risk. In the Bayesian framework, a distribution of this type is referred to as a predictive distribution (or posterior predictive distribution), which can be seen as an "average distribution" (the average of several possible distributions, with averaging probabilities being influenced by the data), by contrast to an estimated (fitted)

Distribution	Probability density function $f(x) = \frac{\mathrm{d}}{\mathrm{d}x}F(x)$	Parameters
Uniform(a, b)	$\frac{1}{b-a},\ x \in [a, b]$	$a, b \in \mathbb{R}$, with $a < b$
Normal(μ, σ)	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},\ x \in \mathbb{R}$	$\mu \in \mathbb{R}$, $\sigma > 0$
LogNormal(μ, σ)	$\frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\ln(x)-\mu}{\sigma})^2},\ x > 0$	$\mu \in \mathbb{R}$, $\sigma > 0$
Exponential(λ)	$\lambda e^{-\lambda x},\ x > 0$	$\lambda > 0$
Pareto(τ, α)	$\alpha\tau^{\alpha} x^{-\alpha-1},\ x > \tau$	$\tau, \alpha > 0$

Table 1.1: Some well-known continuous distributions.

distribution, which tends to rather be a "modal distribution" (the "most likely" true distribution given the data). For contributions on the relevance of predictive distributions in the context of risk management, see e.g. Cairns [19], Gerrard and Tsanakas [47] and Bignozzi and Tsanakas [12, 13].

1.1.3 Model and parameter uncertainty

In practice, the estimation of a distribution is often performed according to a parametric approach, which can be summarized as follows. In a first step the risk under consideration is assumed to follow a distribution (or model) that belongs to a family of distributions indexed by a parameter (possibly a vector ; see Table 1.1 and 1.2 for examples of parametric distributions). The value of that parameter is then estimated from the data, and one checks if the corresponding distribution provides a satisfying fit. If so, then that distribution is considered as the "correct" one, and otherwise the process may be repeated with another parametric distribution family. As a result, under a parametric approach the epistemic uncertainty can be decomposed into the potential error in the selection of the model (model uncertainty) and in the estimation of the parameter (parameter uncertainty). In many situations this distinction turns out to be useful (notably when building a predictive distribution), e.g. because it allows to reflect the eventual different degree of uncertainty one may have regarding the selected model and estimated parameter.

1.2 The body and the tail of a distribution

When modelling risks, it is often relevant to distinguish the body (or bulk, central part) of the distribution, from its tails (left and right). Loosely speaking, the body of a distribution F consists of a (limited) interval of values within which that distribution concentrates most of its probability mass, and hence where most occurrences of the risk X will fall. By contrast, the left-tail and right-tail of a distribution constitute less likely values of the risk, which deviate (potentially by far) from the body, by being respectively smaller and larger. Since here larger values of X denote more adverse scenarios and we adopt a risk management standpoint, we will focus on the right tail of F, which we will simply refer to as the tail of the distribution.

Distribution	Probability mass function $\mathbb{P}(X = x)$	Parameters
Bernoulli(p)	$p^x(1-p)^{1-x},\ x \in \{0,1\}$	$p \in [0,1]$
Binomial(n,p)	$\binom{n}{x}p^x(1-p)^{n-x},\ x \in \{0,...,n\}$	$n \in \mathbb{N}_0, p \in [0,1]$
NegativeBinomial(r,p)	$\binom{x+r-1}{x}p^x(1-p)^r,\ x \in \mathbb{N}_0$	$r \in \mathbb{N}_+, p \in [0,1]$
Poisson(λ)	$\frac{\lambda^x e^{-\lambda}}{x!},\ x \in \mathbb{N}_0$	$\lambda > 0$

Table 1.2: Some well-known discrete distributions.

There are several reasons for distinguishing the body and tail of a distribution, which include:

- In practice, the body and tail of the distribution of a risk often turn out to exhibit behaviours of a different kind, which may therefore be better modelled separately. For instance, most data points of an observed sample may cluster according to some Normal shape, and at the same time the largest observations (order statistics) may signal some heavy-tail behaviour.
- The data will by definition be scarcer in the tail of a distribution than in its body, and different statistical methods may therefore be appropriate for these two regions. For example, one may be reluctant to extend a fitted distribution into the tail, beyond the largest observation, and therefore rely on some alternative technique to model that part of the distribution.
- The purpose of risk modelling may be the calculation of some quantity that depends solely on the tail of distribution, such as the value-at-risk (quantile) or expected shortfall of X at some high level.

1.2.1 Types of tails

The tail of a distribution encapsulates crucial information regarding the behaviour of a risk. The study of the tail behaviour is therefore an important topic in risk management, and one may distinguish several categories.

Heavy- and light-tailed distributions

Let $\overline{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ be the survival function of X. Then X is said to have a heavy tail if its survival function decreases slower than that of an Exponential random variable, i.e.

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{e^{-\lambda x}} = \infty, \text{ for all } \lambda > 0.$$

Otherwise X is said to have a light tail. Heavy-tailed distributions have the ability to produce much larger deviations than light-tailed distributions. For instance, if X has a heavy

tail, then its expectation $\mathbb{E}[X]$ may or may not be finite, but if it has a light tail, then $\mathbb{E}[X]$ is necessarily finite.

Another important quantity is the so-called mean-excess function. Let x_F be the right endpoint of F (note that since the right endpoint of the Exponential distribution is infinite, by the above definition a heavy-tailed distribution F implies $x_F = \infty$). The mean-excess function of X is defined as $e_X(u) = \mathbb{E}[X|X > u]$, for $u < x_F$, i.e. the expectation of X given that X exceeds the threshold u. If X has a heavy tail and $\mathbb{E}[X] < \infty$, there exists some threshold $x^* < x_F$ such that $e_X(u)$ is strictly increasing for all $u > x^*$, meaning that at some point, the larger the threshold u, the larger X is expected to be. In terms of risk, this opens the door to catastrophes ("the worse it has been, the worse it is expected to be"). By contrast, if X is light-tailed, then there exists a $x^* < x_F$ such that $e_X(u)$ is strictly decreasing for all $u \in (x^*, x_F)$, and the mean-excess function of an Exponential random variable is constant. Note that $\mathbb{E}[X] = \infty$ implies $e_X(u) = \infty$, for all u.

In Table 1.1, the LogNormal and Pareto distributions are heavy-tailed, and the Exponential and Normal distributions are light-tailed. The Uniform distribution is also light-tailed, as any other distribution F with $x_F < \infty$.

Sub-exponential distributions

Another way to classify distributions according to their tail behaviour is the concept of sub-exponentiality. Let $X_1, ..., X_n$ be $n \geq 2$ mutually independent and identically distributed (iid) positive random variables with common distribution F and $x_F = \infty$, and consider the sum $S_n = X_1 + ... + X_n$ for $n \geq 1$. The distribution of S_n is

$$F^{*n}(x) = \mathbb{P}(S_n \leq x) = \int_0^x F(x-y)\,\mathrm{d}F^{*n-1}(y),$$

with $F^{*1} = F$, and F^{*n} is the n^{th} convolution of F with itself.

The sum S_n may be large because two or more X_i are relatively large, or only one (i.e. its maximum $M_n = \max\{X_1, ..., X_n\}$) is large. When the latter option is far more likely than the former, i.e. $\mathbb{P}(S_n > x) \sim \mathbb{P}(M_n > x)$ as $x \uparrow \infty$, S_n shall have roughly n times more chances to exceed some large threshold than an individual X_i. That property can be formalized in mathematical terms as

$$\lim_{x\to\infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n, \tag{1.1}$$

in which case F is said to belong to the class $\mathcal{S}$ of sub-exponential distributions (Teugels [76]), and we write $F \in \mathcal{S}$. Chistyakov [23] proved that (1.1) for $n = 2$ already implies $F \in \mathcal{S}$.

It is clear that from a risk management viewpoint, the tail of a distribution $F \in \mathcal{S}$ can be considered dangerous, since it has the ability to produce not so rarely some realizations that exceed by far most others. Also, note that all sub-exponential distributions are heavy-tailed, but the converse it not true, see e.g. Embrechts et al. [32].

Extreme value index

Another approach to characterize the tail of a distribution is provided by extreme value theory. Let $X_1, ..., X_n$ be n iid random variables X with marginal distribution F. The latter has right endpoint $x_F \leq \infty$ and is assumed to be ultimately continuous. Consider now the maximum $M_n = \max\{X_1, ..., X_n\}$. In that setting, the Fisher-Tippett-Gnedenko theorem [40, 48] states that if one can find some normalising constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n\to\infty} F(c_n x + d_n)^n = G(x)$$

for some non-degenerate distribution G, then G is necessarily of one of the following forms:

$$\text{Fréchet:} \quad \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \end{cases} \quad \alpha > 0,$$

$$\text{Gumbel:} \quad \Lambda(x) = \exp(-e^{-x}),\ x \in \mathbb{R}, \tag{1.2}$$

$$\text{Weibull:} \quad \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x \leq 0, \\ 1, & x > 0, \end{cases}, \quad \alpha > 0,$$

and we say that F is in the max-domain of attraction (MDA) of G, which we write $F \in \text{MDA}(G)$.

The Fisher-Tippett-Gnedenko theorem is one of the two main theorems of extreme value theory, the other being the Pickands-Balkema-de Haan theorem [8, 69], which concerns the limit distribution of $(X - u|X > u)$ as $u \to x_F$. For classical textbooks on that topic, we refer to Embrechts et al. [32], Beirlant et al. [11] and de Haan and Ferreira [28]. Since we are interested in the classification of tails here, it is sufficient to notice that the Fisher-Tippett-Gnedenko theorem establishes a class of distributions (those resulting in a non-trivial asymptotic behaviour of $\lim_{n\to\infty} M_n$ after a proper linear normalisation), which splits into the Fréchet, Gumbel and Weibull MDAs. Those three sub-classes can then be shown to have some specific properties (see [32] for details):

- $F \in \text{MDA}(\Phi_\alpha) \Leftrightarrow x_F = \infty$ and $\overline{F}$ can be written as $\overline{F}(x) = \ell(x)x^{-\alpha}$, where $\ell(x)$ is a slowly varying function, i.e.

  $$\lim_{x\to\infty} \frac{\ell(tx)}{\ell(x)} = 1, \text{ for all } t > 0.$$

  So if F is in the Fréchet MDA, then it has an infinite right endpoint and a Pareto-type tail, with parameter α.

- $F \in \text{MDA}(\Lambda) \Leftrightarrow x_F \leq \infty$ and there exists some $z < x_F$ such that $\overline{F}$ can be written as

  $$\overline{F}(x) = c(x) \exp\left(-\int_z^x \frac{g(t)}{a(t)}\,\mathrm{d}t\right),\ z < x < x_F,$$

where $\lim_{x \to x_F} c(x) = c > 0$ and $\lim_{x \to x_F} g(x) = 1$, and $a(x)$ is a positive and continuous function such that $\lim_{x \to x_F} a'(x) = 0$. Note that the above representation is not unique, and a possible choice of $a(x)$ is $a(x) = e_X(x)$, i.e. the mean-excess function of X. A distribution F that belongs to the Gumbel MDA may therefore have a finite or infinite right endpoint, and its asymptotic tail behaviour mixes that of the Exponential distribution, with the respective behaviours of functions $a(x)$, $c(x)$ and $g(x)$.

- $F \in \text{MDA}(\Psi_\alpha) \Leftrightarrow x_F < \infty$ and $\overline{F}(x_F - x^{-1}) = \ell(x) x^{-\alpha}$, i.e. if F is in the Weibull domain, then it necessarily has a finite right endpoint, and in the limit its tail behaves like a bounded power law.

After up to a multiplication of their argument by a constant, the three distributions in (1.2) can be assembled into the so-called Generalized Extreme Value (GEV) distribution

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \text{if } \xi \neq 0, \\ \exp(-\exp(-x)), & \text{if } \xi = 0, \end{cases}$$

where $1 + \xi x > 0$, and $\xi \in \mathbb{R}$ is called the extreme value index. It can then be shown that (again, see [32])

$$\begin{aligned} F \in \text{MDA}(\Phi_\alpha) &\Leftrightarrow F \in \text{MDA}(H_\xi), \text{ with } \xi = \alpha^{-1} > 0, \\ F \in \text{MDA}(\Lambda) &\Leftrightarrow F \in \text{MDA}(H_\xi), \text{ with } \xi = 0, \\ F \in \text{MDA}(\Psi_\alpha) &\Leftrightarrow F \in \text{MDA}(H_\xi), \text{ with } \xi = -\alpha^{-1} < 0, \end{aligned}$$

so the GEV puts distributions in the Fréchet, Gumbel and Weibull MDA into a continuum, where they are characterized by their extreme value index ξ. The latter index may then finally be interpreted as ranking those distributions according to their "dangerousness", since

- $\xi > 0$: F has a Pareto-type tail with parameter $\alpha = \xi^{-1}$, and the larger ξ, the heavier the tail,
- $\xi \leq 0$: F has a tail that is lighter than a Pareto-type tail, and for $\xi < 0$, the smaller ξ, the lighter the tail.

However, the Gumbel domain ($\xi = 0$) includes a large variety of distributions, with both heavy tails (e.g. the LogNormal distribution) and light tails (such as the Normal distribution, or any distribution with $\xi = 0$ and $x_F < \infty$). Also, some distributions with $\xi = 0$ and $x_F < \infty$ may have a lighter tail than some distributions with $\xi < 0$.

1.3 Risk measures

In risk management, the purpose of a probabilistic analysis is often to summarize the risk under consideration through a scalar quantity, which is referred to as a risk measure.

That quantity then typically serves to calculate some performance measure, an amount of resource (e.g. capital) to hold for backing the risk, or to compare risks among each other. The common notation for a risk measure associated with a random variable X is $\rho[X]$, where ρ denotes a functional mapping the distribution of X to $\mathbb{R}$.

1.3.1 Coherent risk measures

In their influential paper, Artzner et al. [4] established the following list of four desirable properties of risk measures (and call a risk measure satisfying all of them, a *coherent* risk measure):

- Translation invariance: $\rho[X + b] = \rho[X] + b$, for any constant $b \in \mathbb{R}$. Adding a constant value to the risk shall change its measure by the same amount.
- Positive homogeneity: $\rho[aX] = a\rho[X]$, for any constant $a \geq 0$. The measure of a risk shall be proportional to its size.
- Monotonicity: $\mathbb{P}(X_1 \leq X_2) = 1 \Rightarrow \rho[X_1] \leq \rho[X_2]$. A risk X_1 that almost surely has a smaller realization than another risk X_2 shall be measured smaller.
- Sub-additivity: $\rho[X_1 + X_2] \leq \rho[X_1] + \rho[X_2]$. Two risks held together shall be measured smaller than separated, i.e. ρ shall reflect the diversification effect of pooling risks together.

Note that some of these properties may not be desirable in some contexts. For instance, in insurance the positive homogeneity property may not always be justified (scaling up risks could lead to non-linear changes for a risk measure, see e.g. the discussion on premium principles in Albrecher et al. [1]). Dhaene and al. [31] pointed out that under sub-additive risk measures, merging risks leads to potentially larger shortfall, which is not so desirable from a regulatory viewpoint, when the risk measures serve to calculate capital requirements.

1.3.2 Value-at-risk and expected shortfall

We now recall and discuss two risk measures that are frequently considered in risk management: the value-at-risk (VaR) and the expected shortfall.

The VaR of a random variable X at level p is defined as

$$\text{VaR}_p[X] = \inf\{x \in \mathbb{R} : F(x) \geq p\},\ p \in [0, 1],$$

and it thus corresponds to the quantile of X at level p. Note that the VaR is not a coherent risk measure, because it fails to satisfy the sub-additivity property in general (which does not prevent it to be sub-additive in many instances).

The VaR arises naturally in case one is interested in the event of a risk exceeding some threshold. For example, when X denotes the loss of an entity (e.g. an insurance company) and k its capital, one is typically interested in the ruin event $X > k$. Accordingly, one may be willing to determine the amount of capital required to keep the probability of a ruin at some low target level β, in which case the capital shall be set at $k = \text{VaR}_{1-\beta}[X]$: If F is continuous at $\text{VaR}_{1-\beta}[X]$, then by definition the resulting ruin probability is $\mathbb{P}(X > \text{VaR}_{1-\beta}[X]) = \beta$. If not, then $\text{VaR}_{1-\beta}[X]$ is the smallest capital amount that makes the ruin probability smaller than β. In fact, this approach has even been made compulsory in the European Union, since the solvency capital requirement of an insurance complying with Solvency II regulation is defined as the VaR at level $p = 99.5\%$ of its net asset position under a one-year time horizon [34].

It is worth mentioning that VaR as a risk measure has been criticised by several authors. For instance, Artzner et al. [4] argued that since VaR is not sub-additive in general, it might discourage diversification. Also, VaR focuses on one point of the distribution, and hence it may measure two risks equally despite their respective tails being possibly very different. In not well regulated or constrained environments, this may create incentives for worsening the potential shortfall ("Après nous, le déluge"). See e.g. Embrechts et al. [33] for a recent contribution on that, in a context of solvency capital optimization.

The expected shortfall at level β of a random variable X is defined as

$$\text{ES}_p[X] = \frac{1}{1-p}\int_p^1 \text{VaR}_\gamma[X]\,\mathrm{d}\gamma,\ p \in [0,1).$$

It is interesting to note that it can alternatively be expressed as

$$\text{ES}_p[X] = \mathbb{E}[X|X > \text{VaR}_p[X]] + \frac{\mathbb{P}(X = \text{VaR}_p[X])}{1-p}\text{VaR}_p[X],\ p \in [0,1), \tag{1.3}$$

and hence for F being continuous at $\text{VaR}_p[X]$ we have

$$\text{ES}_p[X] = \mathbb{E}[X|X > \text{VaR}_p[X]],\ p \in [0,1),$$

the right term being the so-called conditional tail expectation at level p, which is yet another risk measure. Also, from Equation (1.3) we can further write

$$\text{ES}_p[X] = \text{VaR}_p[X] + e_X(\text{VaR}_p[X]) + \frac{\mathbb{P}(X = \text{VaR}_p[X])}{1-p}\text{VaR}_p[X],\ p \in [0,1),$$

which exhibits the connection between the expected shortfall and the mean-excess function (see e.g. Pflug and Romisch [68]).

The expected shortfall is a coherent risk measure, and in contrast to the VaR, it accounts for the whole tail behaviour of X. This is usually seen as an advantage, but it also makes the expected shortfall being potentially more sensitive to the choice of the model, and hence less robust against model uncertainty. Also, note that $\mathbb{E}[X] = \infty$ yields $\text{ES}_p[X] = \infty$, for all $p \in [0,1)$. The expected shortfall is therefore not useful to measure risk for infinite mean models.

1.4 Reinsurance

Reinsurance is a particular type of risk-transfer mechanism, under which an insurer (the reinsured, or first-line insurer, cedent) cedes to another insurer (the reinsurer) a part of his risks, and pays in exchange a reinsurance premium. There are several motivations for an insurer to buy reinsurance, such as

- Reducing the probability to suffer losses that are hard to digest.
- Stabilizing business results.
- Reducing required capital.
- Increasing underwriting capacity.
- Accessing benefits from larger diversification pools.

For a detailed discussion of the above motivations, and more generally, an extensive treatment of topics related to reinsurance, we refer to Albrecher et al. [1].

1.4.1 Choice of reinsurance

From a mathematical point of view, the relationship between an insurer and a reinsurer can be represented (and simplified) as follows: The insurer sells one or more insurance contracts to its policyholder, after which he bears the total risk X, but has collected the total premium P_X. The insurer may then purchase a reinsurance cover, under which he will cede the portion $R = r(X)$ (the ceded loss) of X to the reinsurer, and in turn pay the corresponding reinsurance premium $P_R = \pi[R]$. The function r denotes the ceded loss function, i.e. the pre-defined rule that determines the amount to be paid by the reinsurer, depending on the realization of X. And the functional π is the premium principle, i.e. the risk measure that the reinsurer applies to R for determining the reinsurance premium P_R. Under a reinsurance cover, the part of X and P_X being retained by the insurer are thus $D = X - R$ (the retained loss) and $P_D = P_X - P_R$ (the retained premium), and the reinsurer has the net position $P_R - R$.

The insurer can usually choose r among a set of candidates proposed by the reinsurer, say $\mathcal{C}$, and each choice of r will then result in a potentially different reinsurance premium $\pi[r(X)]$, the premium principle π being chosen by the reinsurer. In this setting, the best choice of r arises as an interesting problem, which can typically be tackled by considering some optimization problem, e.g.

$$r^* = \arg\max_{r \in \mathcal{C}} \varphi[r(X)], \tag{1.4}$$

where φ is some functional representing an appropriate objective for the insurer.

For instance, let $L_r = D - P_D = X - r(X) - P_X + \pi[r(X)]$ be the potential loss of the insurer under ceded loss function r (so $-L_r$ is the corresponding potential profit).

Let further $\rho[L_r]$ be the solvency capital that the insurer must hold to cover L_r. Then examples of the objective $\varphi[r(X)]$ are

- Total cost or total risk exposure (Cai and Tan [18]): $\varphi[r(X)] = \rho[L_r + P_X]$, so in that case $\varphi[r(X)]$ is to be minimized.
- Expected profit (Kull [59] and Albrecher and Cani [2]): $\varphi[r(X)] = \mathbb{E}[-L_r] - \frac{i}{1-i}\rho[L_r]$, for ρ being translation invariant, and i denotes some cost-of-capital rate.
- Return on risk adjusted capital (RORAC): $\varphi[r(X)] = \mathbb{E}[-L_r]/\rho[L_r]$.

Many variants of problem (1.4) have been considered, e.g. with a budget constraint for the reinsurance premium (Cheung et al. [20]), with some non-insurable (background) risk standing besides X (Dana and Scarsini [26]), when the insurer and reinsurer have heterogeneous beliefs regarding the distribution of X (Boonen [16]), and when the choice of reinsurance is to be made simultaneously for several risks (Zhu et al. [83]), to cite a few (for an overview, see Albrecher et al. [1] and Cai and Chi [17]).

1.4.2 Reinsurance forms and premium principles

For illustration purposes, we list three examples of ceded loss functions that are often considered, both in practice and in the actuarial literature:

- Quota-Share (QS): $r(x) = ax$, where $a \in (0, 1)$ denotes the proportionality factor. Under a Quota-Share, the insurer thus shares a portion a of X to the reinsurer, i.e. $R = aX$.
- Stop-Loss (SL): $r(x) = (x - d)_+$, where $d \geq 0$ is called the deductible. So under a SL cover, only the part of X exceeding d is ceded to the reinsurer.
- Bounded Stop-Loss (BSL): $r(x) = \min\{(x - d)_+, \ell\}$, with $d \geq 0$ being again the deductible, and $\ell > 0$ is called the limit or layer. As its name indicates, a BSL is simply the bounded version of a SL, i.e. the reinsurer will not pay more than ℓ.

Finally, we further mention three instances of premium principles:

- Expected value principle: $\pi[R] = (1 + \theta)\mathbb{E}[R]$, where $\theta \geq 0$ is some safety loading factor.
- Standard-deviation principle: $\pi[R] = \mathbb{E}[R] + \beta\mathbb{S}[R]$, where $\beta > 0$ is again some loading factor, and $\mathbb{S}[R]$ denotes the standard-deviation of R.
- Risk-adjusted or Wang principle: $\pi[R] = \int_0^\infty w(\mathbb{P}(R > x))\,\mathrm{d}x$, with $w(u)$ being a non-negative increasing and concave function such that $w(0) = 0$ and $w(1) = 1$.

For more examples of both reinsurance forms and premium principles and instructive discussions about these topics, we again refer to Albrecher et al. [1].

1.5 Contributions of this thesis

This thesis contains results in the field of risk modelling, both with and without parameter and model uncertainty, as well as reinsurance, with applications that are mostly related to tail risk. The article on which Chapter 5 is based has already been published in *Insurance: Mathematics and Economics*.

In Chapter 2, we consider the problem of risk estimation under parameter and model uncertainty. For so-called transformed location-scale distribution families, we prove that when only parameter uncertainty is involved, a special type of predictive distribution can be built and used to calculate a capital that yields a targeted solvency probability, despite parameter uncertainty. This extends some of the results in Gerrard and Tsanakas [47] to a more general framework. These results are then used to derive an explicit and simple analytical formula for the capital estimator of a Pareto random loss, which can be used even in the particular case where the data consists of consecutive order statistics rather than a complete random sample. Afterwards, we test the robustness of this capital estimator against potential model misspecification, for distributions with a Pareto-type tail. As a by-product, we also provide a new explanation of the connection between Bayesian and fiducial inference.

By accounting for epistemic uncertainty, predictive distributions tend to reflect risk better than estimated distributions. This naturally makes them more conservative, which is sometimes viewed as a drawback: At the end of the risk assessment process, limited resources (e.g. capital) are available, and by being too conservative one may not be effective. While this objection is a valid concern, we believe that it does not account for an important (and often implicit) aspect of many risk management decision processes: The assessment of risk often serves to ultimately perform some kind of trade-off between several competing risks for allocating limited resources, rather than determining absolute resource amounts to isolated risks. In such a situation, since anyway limited resources are available, if each risk is modelled with a predictive distribution, the conservative aspect likely ceases to be an issue. Correspondingly, since predictive distributions tend to reflect risk better than estimated distributions, they may allow for potentially better resource allocation decisions. In Chapter 3, we illustrate this by considering a setting where limited resources must be allocated separately to two risks, and parameter uncertainty is involved for each of them. Through a simple example, we show the ability of predictive distributions to indeed yield better resource allocations than a certain type of estimated distributions. This chapter therefore contributes to the existing literature advocating in favour of predictive distributions in risk management problems, see e.g. Cairns [19], Gerrard and Tsanakas [47] and Bignozzi and Tsanakas [12, 13].

The Fisher-Tippett-Gnedenko theorem provides the only possible non-degenerate limit distributions for the linearly normalised maximum of a large sample of independent and identically distributed random variables, and it is one of the foundational results of extreme value theory. That theorem naturally compares with the central limit theorem, which establishes the normality of linearly normalised sample sums, under some mild

conditions. However, unlike the central limit theorem, the Fisher-Tippett-Gnedenko theorem can be applied in various ways in a given situation, and this potentially results in different limit distributions for the considered sample maximum. In Chapter 4, we use this aspect of the Fisher-Tippett-Gnedenko theorem to propose an alternative family of extreme value distributions, which contains both the Weibull and Fréchet distributions, but where the Gumbel distribution is modified, and ends up being endowed with a shape parameter. Also, we mention several directions for extending this work, which may result in interesting new results in extreme value theory.

In Chapter 5 we finally study a special type of reinsurance cover, for which the ceded loss function not only depends on the loss of the cedent, but is also contingent upon his performance, for instance measured in terms of his loss ratio compared to the average loss ratio of the market. We show that this type of reinsurance may be efficiently used to manage risk in the presence of financial distress cost, when the cover is provided to a cohort of insurers with positively correlated loss experience. In addition to theoretical results we quantitatively illustrate the potential performance improvement in a numerical example Interestingly, this can lead to an improvement for the insurer and reinsurance at the same time, which is not so common in this type of literature. This chapter thus contributes to the literature on the optimal choice of reinsurance (see Albrecher et al. [1] and Cai and Chi [17] for an overview), as well as potentially to initiatives in the reinsurance industry.

Chapter 2

Risk estimation under parameter and model uncertainty

This chapter is based on L. Vincent (2022): *Risk estimation under parameter and model uncertainty*. Preprint, University of Lausanne [79].

Abstract. This chapter considers the problem of risk estimation under parameter and model uncertainty, when a random loss must be backed with some capital, with the objective to reach a target solvency probability. We prove that for a large class of distributions that are frequently used in risk management, when only parameter uncertainty is involved, a (Bayesian) predictive distribution can be built, which allows to calculate a capital that can be expected to yield the target solvency probability, despite parameter uncertainty. This extends some of the results in Gerrard and Tsanakas [47] to a more general framework. We then use these results to derive a simple analytical formula for the capital estimator of a Pareto random loss, which can be used even in case the data consists of consecutive order statistics, rather than a complete random sample. Afterwards, we test the robustness of this capital estimator against potential model misspecification, for distributions with a Pareto-type tail. As a by-product, we also provide a new explanation of the connection between Bayesian and fiducial inference.

2.1 Introduction

2.1.1 Aleatoric and epistemic uncertainty

In the modern practice of risk management, one often seeks to represent risks (random outcomes that may have adverse consequences) with probability distributions (or simply distributions), which one then uses to compute risk measures and other useful quantities. However, in most cases the distribution of a risk (assuming it exists) is unknown and must therefore be estimated. For doing so, one relies on some estimation technique, which typically consists to fit a distribution to the data (observed past realizations of the risk). Throughout the years, statisticians have developed numerous such fitting techniques, with a variety of potentially desirable properties. Still, as good as those techniques might be, when performed under a limited amount of data, the resulting estimated distribution is

likely to differ from the true one.

From a risk management perspective, the potential error in the estimation of a distribution constitutes an additional source of risk (epistemic uncertainty), which can be treated as different from the original risk (aleatoric uncertainty ; for some background on the dichotomy between aleatoric and epistemic uncertainty, see e.g. Der Kiureghian and Ditlevsen [29] and Aven and Zio [7]). While the manifestation of epistemic uncertainty tends to depend on the particular technique by which the distribution of a risk is estimated, it can be considered as always present when that estimation is performed under a limited amount of data, and it shall be considered. For instance, assume that limited resources must be allocated to two independent risks with identical estimated marginal distributions, but one of them has been obtained with significantly less data than the other. In this situation, the epistemic uncertainty inherent to each of these two risks evidently differs, and it is intuitive that one shall allocate unequal resources to them, despite their estimated marginal distributions being identical. In the risk management literature, it has been shown multiple times and in various frameworks that failure to account for epistemic uncertainty yields incorrect (and usually too optimistic) risk assessment, see for instance Gerrard and Tsanakas [47], Fröhlich and Weng [43] and Pitera and Schmidt [70].

There exist several methods that allow to build a distribution reflecting both the aleatoric and epistemic uncertainty relative to a risk. In the Bayesian framework, a distribution of this type is referred to as a predictive distribution (or posterior predictive distribution). For calculating it (more details are provided in Section 2.5.2), one must start by assuming that the true but unknown distribution of the risk belongs to a set of known candidate distributions (or distribution family), which are thus possible descriptions of its aleatoric uncertainty. Before the data is taken into account, the uncertainty (or belief) that one has about which distribution is the true one is quantified in terms of prior probabilities. Using the data, these prior probabilities are then updated to posterior ones (according to Bayes' formula), and the latter are taken as a quantitative representation of epistemic uncertainty. Afterwards, all possible distributions are averaged according to their respective posterior probabilities, and the resulting distribution, termed the predictive distribution, serves to describe the risk under consideration.

The predictive distribution of a risk is thus an average distribution, which by construction indeed accounts for both its aleatoric and epistemic uncertainty. Also, it generally differs from an estimated (fitted) distribution, just as the mean of a random variable typically differs from its mode. To say it differently, an estimated distribution aims at answering an ontological question ("What is the true distribution of the risk?"), where a predictive distribution reflects the risk from a rather personal viewpoint ("How does this risk appear to me, i.e. considering the potentially limited knowledge I have of it?").

2.1.2 Model and parameter uncertainty

In practice the estimation of a distribution is often done according to a parametric approach, which can be summarized as follows. In a first step the risk under consideration

is assumed to follow a distribution (or model) that belongs to a family of distributions indexed by a parameter (possibly a vector). The value of that parameter is then estimated from the data, and one checks if the corresponding distribution provides a satisfying fit. If so, then that distribution is taken as the estimated one, and otherwise the process may be repeated with another parametric distribution family. As a result, under a parametric approach the epistemic uncertainty can be decomposed into the potential error in the selection of the model (model uncertainty) and in the estimation of the parameter (parameter uncertainty). Using that distinction, several authors have considered the special case where the distribution of a risk belongs to a known parametric family, but the value of the parameter is not known, i.e. there is no model uncertainty and the epistemic uncertainty therefore consists of only parameter uncertainty. They proposed several methods for that situation, which allow to account for both the aleatoric and parameter uncertainty of the risk.

For instance, Gerrard and Tsanakas [47] considered the case of a risk whose distribution belongs to a transformed location-scale family, which refers to a distribution family that can be expressed as a strictly increasing transformation of a location-scale family and include several models being frequently used in risk management (see Section 2.3.2). In that setting, they proposed two different approaches for calculating a threshold value (capital amount), which is expected to yield a given target failure (ruin) probability, despite the parameter of the distribution being unknown: (i) Using an estimated distribution (the one corresponding to the maximum likelihood estimate of the parameter), but with a modified target failure probability. Under this approach, the estimated distribution thus represents the aleatoric uncertainty, and the modification of the target failure probability accounts for parameter uncertainty. (ii) Using a predictive distribution, together with the initial target failure probability. Fröhlich and Weng [43] then obtained a similar results, but using fiducial distributions, which can be seen as predictive distributions, but the selection of the prior is done indirectly, through the so-called fiducial argument (see Section 2.9). Bignozzi and Tsanakas [12, 13] considered the notion of residual risk, which corresponds to a risk measure (such as the value-at-risk and expected shortfall) applied to the difference between a risk (in terms of a random loss) and the corresponding capital estimator. They shown that capital estimators can be built, using either parametric bootstrapping or a predictive distribution, which lower the impact of parameter uncertainty on the residual risk (and even fully compensate it for some risk measures and location-scale families). Pitera and Schmidt [70] revisited the notion of unbiasedness in the context of financial risk estimation and backtesting procedures, and used it to compare various estimators, for several risk measures (mainly the value-at-risk and expected shortfall), both theoretically (under the assumption of a normally distributed risk) and with NASDAQ data.

2.1.3 The present contribution

In this chapter, we consider the setting where the risk represents a future random loss, whose parametric distribution belongs to a transformed location-scale family. The loss must be backed with some capital, which, ideally, yields some target solvency probability. That loss could for instance represent the aggregate loss of an insurance company

complying with Solvency II regulation, who must therefore hold a capital that leaves it with a solvency probability of at least 99.5% over a one-year time horizon. Other examples are public infrastructures such as dykes or nuclear plants, which are often supposed to be built so as to cope with various types of extreme natural events that are expected to occur, for instance, once every $10'000$ years (in that case the target annual solvency probability would be 99.9%). Note that for those other examples the loss and capital would typically not be in monetary units.

However, in that setting, we assume that parameter uncertainty is at play (and possibly model uncertainty too), which prevents to calculate the desired capital amount, and hence the latter estimated from some available data. Our setting is thus closely related to that of Gerrard and Tsanakas [47] (but our problem is formulated in terms of solvency rather than failure or ruin), and our main contributions consist of:

(a) Firstly, we extend the above-mentioned approach (ii) of Gerrard and Tsanakas [47] by relaxing some of their assumptions (see Remark 2.5.2), and show that a predictive distribution can also be built in that more general framework, which allows to calculate a capital that yields a target expected solvency probability, despite parameter uncertainty.

(b) Secondly, we use the above to obtain an explicit and simple formula for calculating the capital of a random loss following a Pareto distribution. Thanks to (a), that formula can not only be used when the data is a Pareto random sample, but also when it consists of consecutive order statistics of a larger Pareto random sample.

(c) Thirdly we consider the situation where one ignores the overall model of the random loss, but one knows that its distribution has a Pareto-type tail. That situation can be seen as a particular type of model uncertainty, and we propose addressing it by using the formula obtained in (b), to calculate the capital using only the largest order statistics of the data. This approach is then illustrated through a numerical example.

The result in (a) relies on a lemma that we prove in this chapter, and from which incidentally an interesting discussion arises on the connection between Bayesian and fiducial inference. Yet, since that discussion relates to statistical inference rather than risk management, it is not linked with the core topic of this article, and we moved it to appendix.

The rest of the chapter is organized as follows. In Section 2.2, we introduce our notation and provide some useful reminders. In Section 2.3, we describe the setting and assumptions that we consider in the sequel. In Section 2.4, we introduce and discuss the notion of expected solvency probability. In Section 2.5, we first recall and illustrate the result according to which, when parameter uncertainty is involved, capital calculation via simple maximum likelihood estimation does not provide the targeted solvency probability in general. Afterwards, we derive the results (a), and provide some calculation details and examples. In Section 2.6 we consider the Pareto case, and hence treat (b). In Section

2.7, we address (c), i.e. capital calculation when both parameter and model uncertainty are involved. Section 2.8 then concludes, and in the appendix we discuss the connection between Bayesian and fiducial inference in Section 2.9, whereas the proofs are provided in Section 2.10.

2.2 Preliminaries

In the remainder, all random variables are defined on the common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Also, all random variables and random vectors are written in capital letters, and all vectors are written in bold letters. We denote the set of real numbers by $\mathbb{R} = (-\infty, +\infty)$, and that of positive and strictly positive real numbers respectively by $\mathbb{R}_0 = [0, +\infty)$ and $\mathbb{R}_+ = (0, +\infty)$.

Let $\mathbf{Q} = (Q_1, ..., Q_d)$ be a continuous random vector of dimension $d \geq 1$ (note that a random variable can be defined as a random vector of dimension 1, so the notation relative to random vectors that is hereafter introduced also applies to random variables). The i^{th} greatest element (order statistics) of $\mathbf{Q}$ are denoted by $Q_{i:d}$, for each i, so $Q_{1:d} \leq ... \leq Q_{d:d}$. The joint cumulative distribution function (cdf) and joint probability density function (pdf) of $\mathbf{Q}$ at $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$ are written as $F_{\mathbf{Q}}(\mathbf{x}) = \mathbb{P}(Q_1 \leq x_1, ..., Q_d \leq x_d)$ and $f_{\mathbf{Q}}(\mathbf{x}) = \frac{\mathrm{d}}{\mathrm{d}x_1} \cdots \frac{\mathrm{d}}{\mathrm{d}x_d} F_{\mathbf{Q}}(\mathbf{x})$. The copula and copula density of $\mathbf{Q}$ at $\mathbf{u} = (u_1, ..., u_d) \in [0,1]^d$ are denoted by $C_{\mathbf{Q}}(\mathbf{u}) = \mathbb{P}(F_{Q_1}(Q_1) \leq u_1, ..., F_{Q_d}(Q_d) \leq u_d)$ and $c_{\mathbf{Q}}(\mathbf{u}) = \frac{\mathrm{d}}{\mathrm{d}u_1} \cdots \frac{\mathrm{d}}{\mathrm{d}u_n} C_{\mathbf{Q}}(\mathbf{u})$. The version of $\mathbf{Q}$, given that another random vector $\mathbf{R}$ has value $\mathbf{r}$, is denoted by $(\mathbf{Q}|\mathbf{R} = \mathbf{r})$, and we write the corresponding conditional cdf and pdf by $F_{\mathbf{Q}|\mathbf{R}}(\cdot|\mathbf{r})$ and $f_{\mathbf{Q}|\mathbf{R}}(\cdot|\mathbf{r})$. If the distribution of $\mathbf{Q}$ depends on some parameter $\boldsymbol{\theta}k$, then depending on the context we may write $\mathbf{Q}$ as $\mathbf{Q}(\boldsymbol{\theta})$, and its cdf, pdf, copula and copula density are denoted by $F_{\mathbf{Q}}(\cdot;\boldsymbol{\theta})$, $f_{\mathbf{Q}}(\cdot;\boldsymbol{\theta})$, $C_{\mathbf{Q}}(\cdot;\boldsymbol{\theta})$ and $c_{\mathbf{Q}}(\cdot;\boldsymbol{\theta})$.

Consider some mapping $s(\cdot) : \mathbb{R}^d \to \mathbb{R}^z$, where $z \geq 1$. Then the resulting random vector $\mathbf{S} = s(\mathbf{Q}) = (S_1, ..., S_z)$ is said to be a statistic of $\mathbf{Q}$. If the distribution of $\mathbf{Q}$ depends on a parameter $\boldsymbol{\theta}$, so does the distribution of any statistic $\mathbf{S} = s(\mathbf{Q}) = (S_1, ..., S_z)$, and hence we may write $\mathbf{S}$ and each S_i as $\mathbf{S}(\boldsymbol{\theta})$ and $S_i(\boldsymbol{\theta})$.

There exist various types of statistics, and for later use we recall that of a sufficient statistic.

Definition 2.2.1. *Let $\mathbf{S} = s(\mathbf{Q})$ be a statistic. If $f_{\mathbf{Q}}(\mathbf{x};\boldsymbol{\theta})$ can be decomposed as*

$$f_{\mathbf{Q}}(\mathbf{x};\boldsymbol{\theta}) = \gamma_1(s(\mathbf{x});\boldsymbol{\theta})\gamma_2(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathbb{R} \text{ and } \boldsymbol{\theta} \in \mathcal{P},$$

where $\gamma_1(\cdot;\boldsymbol{\theta})$ and $\gamma_2(\cdot)$ are two non-negative functions, and $\gamma_2(\cdot)$ does not depend on $\boldsymbol{\theta}$, then $\mathbf{S}$ is said to be a sufficient statistic of $\mathbf{Q}$.

Finally, the value-at-risk at level $p \in [0,1]$ of a random variable Q is written as $\text{VaR}_p[Q] = \inf\{x \in \mathbb{R} : F_Q(x) \geq p\}$. Also, we recall that for any increasing and continuous function $\zeta(\cdot)$ and $p \in [0,1]$, we have

$$\text{VaR}_p[\zeta(Q)] = \zeta(\text{VaR}_p[Q]). \tag{2.1}$$

For a proof, see e.g. Theorem 1 in Dhaene et al. [30].

2.3 The setting

2.3.1 General setting

Let Y be a continuous random variable representing a future loss, whose distribution depends on a two-dimensional parameter $\boldsymbol{\theta} = (\theta_1, \theta_2)$. We ignore the value of $\boldsymbol{\theta}$, but we know the distribution family $D_Y = \{f_Y(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathcal{P}\}$, with $\mathcal{P} = \mathbb{R} \times \mathbb{R}_+$ being the set of possibles values for $\boldsymbol{\theta}$, or parameter space for short. Also, we observed the realization of sample $\mathbf{X} = (X_1, ..., X_n)$, where $X_1, ..., X_n$ are $n > 1$ continuous random variables. The marginal distribution of each X_i depends on the same parameter $\boldsymbol{\theta}$ as Y, and each marginal family $\mathcal{D}_{X_i} = \{f_{X_i}(\cdot;\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathcal{P}\}$ is assumed to be known. Note that since the distribution of each X_i depends on $\boldsymbol{\theta}$, so does the distribution of $\mathbf{X}$.

It is then assumed that

$$\mathbf{X}(\boldsymbol{\theta}) \text{ and } Y(\boldsymbol{\theta}) \text{ are independent, for all } \boldsymbol{\theta} \in \mathcal{P}. \tag{A1}$$

Regarding random variables $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$, they may or may not be mutually independent and identically distributed (iid). For example, $\mathbf{X}$ can be a random sample of size n (iid case), or the n largest order statistics of a larger random sample (non-iid case). However in any case, we assume that the dependence structure of $\mathbf{X}(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$, which in terms of the copula $C_{\mathbf{X}}(\cdot;\boldsymbol{\theta})$ means that

$$C_{\mathbf{X}}(\cdot;\boldsymbol{\theta}) \text{ is independent of } \boldsymbol{\theta}, \tag{A2}$$

and we thus write $C_{\mathbf{X}}(\cdot;\boldsymbol{\theta}) = C_{\mathbf{X}}(\cdot)$. Moreover, we assume that $C_{\mathbf{X}}(\cdot)$ is known. Therefore, since on the one hand each distribution family $\mathcal{D}_{X_i}$ is known, and on the other hand by Sklar's theorem [75] for any $\boldsymbol{\theta} \in \mathcal{P}$ we have

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) = c_{\mathbf{X}}(F_{X_1}(x_1;\boldsymbol{\theta}), ..., F_{X_n}(x_n;\boldsymbol{\theta})) \prod_{i=1}^{n} f_{X_i}(x_i;\boldsymbol{\theta}), \text{ for all } \mathbf{x} \in \mathbb{R}^n,$$

the pdf $f_{\mathbf{X}}(\cdot;\boldsymbol{\theta})$ is also known, for all $\boldsymbol{\theta} \in \mathcal{P}$.

The respective domains of $\mathbf{X}(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ may or may not depend on $\boldsymbol{\theta}$, and we write them as $\mathcal{X}(\boldsymbol{\theta}) = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) > 0\}$ and $\mathcal{Y}(\boldsymbol{\theta}) = \{y \in \mathbb{R} : f_Y(y;\boldsymbol{\theta}) > 0\}$. Finally, the domains of $\mathbf{X}$ and Y induced by $\mathcal{P}$ are denoted by $\mathcal{X} = \cup_{\boldsymbol{\theta}\in\mathcal{P}}\mathcal{X}(\boldsymbol{\theta})$ and $\mathcal{Y} = \cup_{\boldsymbol{\theta}\in\mathcal{P}}\mathcal{Y}(\boldsymbol{\theta})$.

2.3.2 Transformed location-scale families

Each marginal distribution family $\mathcal{D}_{X_i}$ is assumed to be a transformed location-scale family. So, by definition, for each i we can find a strictly increasing function $g_i(\cdot) : \mathbb{R} \to \mathcal{X}_i \subseteq \mathbb{R}$, such that $X_i(\boldsymbol{\theta})$ can be written as

$$X_i(\boldsymbol{\theta}) = g_i(\theta_1 + \theta_2 U_i), \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.2}$$

where, letting $g_i^{-1}(\cdot)$ being the inverse of $g_i(\cdot)$, and $\mathbf{e} = (0, 1)$, the random variable U_i is defined as $U_i = g_i^{-1}(X_i(\mathbf{e}))$, and hence it is continuous and its distribution does not depend on $\boldsymbol{\theta}$. It may be recognized that the distribution family of random variable $\theta_1 + \theta_2 U_i$ is by definition a location-scale family, with location and scale parameters θ_1 and θ_2. Therefore, since $\mathcal{D}_{X_i}$ is the distribution family of the transformed random variable $X_i(\boldsymbol{\theta}) = g_i(\theta_1 + \theta_2 U_i)$, it is said to be a transformed location-scale family. In Example 2.3.1 the Pareto distribution is shown to define a transformed location-scale family, and other examples are listed in Table 2.1.

Example 2.3.1. Let U_i be an Exponential(1) random variable. Let further $g_i(x) = e^x$, and consider the resulting random variable $X_i(\boldsymbol{\theta}) = g_i(\theta_1 + \theta_2 U_i) = e^{\theta_1 + \theta_2 U_i}$, with $\boldsymbol{\theta} \in \mathcal{P}$. It may be checked that $X_i(\boldsymbol{\theta})$ has pdf

$$f_{X_i}(x; \boldsymbol{\theta}) = \theta_2^{-1} (e^{\theta_1})^{\theta_2^{-1}} x^{-\theta_2^{-1} - 1} \mathbb{1}\{x > e^{\theta_1}\},$$

where $\mathbb{1}\{E\}$ is the indicator function of the event E. While unusually parametrized, this turns out to be the pdf of a Pareto(τ, α) random variable, where $\tau = e^{\theta_1}$ and $\alpha = \theta_2^{-1}$. Also, since here $g_i(\cdot)$ is a strictly increasing function, the distribution family of $X_i(\boldsymbol{\theta})$ is by construction a transformed location-scale family. Therefore, the Pareto distribution defines a transformed location-scale family.

Define the mappings $g(\cdot)$ and $g^{-1}(\cdot)$, such that

$$g(\mathbf{x}) = (g_1(x_1), ..., g_n(x_n)), \text{ for all } \mathbf{x} \in \mathbb{R}^n,$$

and

$$g^{-1}(\mathbf{x}) = (g_1^{-1}(x_1), ..., g_n^{-1}(x_n)), \text{ for all } \mathbf{x} \in \mathcal{X},$$

and note that from Assumption (A2), random vectors $\mathbf{X}(\mathbf{e})$ and $\mathbf{X}(\boldsymbol{\theta})$ have the same copula, for all $\boldsymbol{\theta} \in \mathcal{P}$. Therefore, letting $\mathbf{U} = g^{-1}(\mathbf{X}(\mathbf{e})) = (U_1, ..., U_n)$, Assumption (A2) and Equation (2.2) allow to write $\mathbf{X}(\boldsymbol{\theta})$ as

$$\mathbf{X}(\boldsymbol{\theta}) = g(\theta_1 + \theta_2 \mathbf{U}), \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.3}$$

where $\theta_1 + \theta_2 \mathbf{U} = (\theta_1 + \theta_2 U_1, ..., \theta_1 + \theta_2 U_n)$, by convention. The domain of $\mathbf{U}$ is denoted by $\mathcal{U}$.

The distribution family $\mathcal{D}_Y$ is also assumed to be a transformed location-scale family, and hence as for each X_i, we can find a strictly increasing function $h(\cdot) : \mathbb{R} \to \mathcal{Y} \subseteq \mathbb{R}$ such that $Y(\boldsymbol{\theta})$ can be written as

$$Y(\boldsymbol{\theta}) = h(\theta_1 + \theta_2 V), \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.4}$$

where $V = h^{-1}(Y(\mathbf{e}))$.

We now illustrate the above overall setting with an example, which will then reappear throughout this article.

Distribution family	$g_i(x)$	$f_{U_i}(x)$	$f_{X_i}(x;\boldsymbol{\theta})$
Uniform(α,β) $\alpha,\beta\in\mathbb{R}$, and $\alpha>\beta$	x	$1,\ x\in(0,1)$ Uniform$(0,1)$	$\frac{1}{\beta-\alpha},\ x\in(\alpha,\beta)$ $\theta_1=\alpha,\ \theta_2=\beta-\alpha$
Normal(μ,σ) $\mu\in\mathbb{R},\ \sigma>0$	x	$\varphi(x),\ x\in\mathbb{R}$ Normal$(0,1)$	$\frac{1}{\sigma}\varphi\big(\frac{x-\mu}{\sigma}\big),\ x\in\mathbb{R}$ $\theta_1=\mu,\ \theta_2=\sigma$
Log-Normal(μ,σ) $\mu\in\mathbb{R},\ \sigma>0$	e^x	$\varphi(x),\ x\in\mathbb{R}$ Normal$(0,1)$	$\frac{1}{\sigma x}\varphi\big(\frac{\ln(x)-\mu}{\sigma}\big),\ x>0$ $\theta_1=\mu,\ \theta_2=\sigma$
Pareto(τ,α) $\tau>0,\ \alpha>0$	e^x	$e^{-x},\ x>0$ Exponential(1)	$\alpha\tau^{\alpha}x^{-\alpha-1},\ x>\tau$ $\theta_1=\ln(\tau),\ \theta_2=\alpha^{-1}$
Weibull(λ,κ) $\lambda>0,\ \kappa>0$	e^x	$e^{x-e^x},\ x\in\mathbb{R}$ $-$Gumbel$(0,1)$	$\frac{\kappa}{\lambda}(\frac{x}{\lambda})^{\kappa-1}e^{-(x/\lambda)^{\kappa}},\ x>0$ $\theta_1=\ln(\lambda),\ \theta_2=\kappa^{-1}$

Table 2.1: Examples of transformed location-scale families. The function $\varphi(\cdot)$ denotes the pdf of a Normal$(0,1)$ random variable.

Example 2.3.2. Let $U_1,...,U_n$ and V be mutually independent Exponential(1) random variables, with $g_i(x)=e^x$, for each i, and $h(x)=e^x$. Then as seen in Example 2.3.1, for $\boldsymbol{\theta}\in\mathcal{P}$, each $X_i(\boldsymbol{\theta})=e^{\theta_1+\theta_2U_i}$ and $Y(\boldsymbol{\theta})=e^{\theta_1+\theta_2V}$ are Pareto$(e^{\theta_1},\theta_2^{-1})$, so each distribution family $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ is a transformed location-scale family. Moreover, since $U_1,...,U_n$ and V are mutually independent, so are $X_1(\boldsymbol{\theta}),...,X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$, and Assumption (A1) and (A2) are thus satisfied. Also, note that each $g_i(\cdot)$ and $h(\cdot)$ map $\mathbb{R}$ to $\mathbb{R}_+$. As a result, we have $\mathcal{X}_i=\mathbb{R}_+$, for each i, and $\mathcal{Y}_i=\mathbb{R}_+$, and finally $\mathcal{X}=\mathbb{R}_+^n$.

2.3.3 Equivariant statistic

Consider a pair of functions $t_1(\cdot)$ and $t_2(\cdot)$, each mapping $\mathcal{X}$ to a subset of $\mathbb{R}$, and satisfying the equivariance properties

$$t_1(g(\theta_1+\theta_2 g^{-1}(\mathbf{x})))=\theta_1+\theta_2 t_1(\mathbf{x}),\ \text{for all } \mathbf{x}\in\mathcal{X} \text{ and } \boldsymbol{\theta}\in\mathcal{P}, \tag{P1}$$

and

$$t_2(g(\theta_1+\theta_2 g^{-1}(\mathbf{x})))=\theta_2 t_2(\mathbf{x}),\ \text{for all } \mathbf{x}\in\mathcal{X} \text{ and } \boldsymbol{\theta}\in\mathcal{P}. \tag{P2}$$

For the ease of notation, we may then assemble $t_1(\cdot)$ and $t_2(\cdot)$ into the single mapping $t(\cdot)=(t_1(\cdot),t_2(\cdot))$.

There exist many such pairs of functions. For instance, the average and standard-deviation of $g^{-1}(\mathbf{x})$ can be shown to satisfy Properties (P1)-(P2), respectively. Hereafter we provide another illustration, with more details.

Example 2.3.3. Consider the setting of Example 2.3.2, so that $g_i(x) = e^x$ and $g_i^{-1}(x) = \ln(x)$, for each i, and the i^{th} order statistic of $g^{-1}(\mathbf{x})$ is thus $g_{i:n}^{-1}(x_{i:n}) = \ln(x_{i:n})$. Then the pair of functions

$$t_1(\mathbf{x}) = g_{1:n}^{-1}(x_{1:n}) = \ln(x_{1:n}), \text{ for all } \mathbf{x} \in \mathcal{X},$$

and

$$t_2(\mathbf{x}) = \sum_{i=1}^{n}(g_{i:n}^{-1}(x_{i:n}) - g_{1:n}^{-1}(x_{1:n})) = \sum_{i=1}^{n}(\ln(x_{i:n}) - \ln(x_{1:n})), \text{ for all } \mathbf{x} \in \mathcal{X},$$

respectively satisfy Properties (P1)-(P2): Consider the transformation $g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))$, for $\mathbf{x} \in \mathcal{X}$. Since each $g_i(\cdot)$ is strictly increasing, if $\boldsymbol{\theta} \in \mathcal{P}$, then the transformation $g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))$ preserves the order of $g^{-1}(\mathbf{x})$ and we can thus write the i^{th} order statistic of $g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))$ as $g_{i:n}(\theta_1 + \theta_2 g_{i:n}^{-1}(x_{i:n})) = e^{\theta_1 + \theta_2 \ln(x_{i:n})}$, for each i. This yields $t_1(g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))) = \theta_1 + \theta_2 \ln(x_{1:n}) = \theta_1 + \theta_2 t_1(\mathbf{x})$ and $t_2(g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))) = \sum_{i=2}^{n}(\theta_2 \ln(x_{i:n}) - \theta_2 \ln(x_{1:n})) = \theta_2 t_2(\mathbf{x})$, for all $\mathbf{x} \in \mathcal{X}$ and $\boldsymbol{\theta} \in \mathcal{P}$, so Properties (P1)-(P2) are indeed both satisfied.

For an arbitrary pair of such functions, consider the resulting statistic $\mathbf{T} = t(\mathbf{X}) = (T_1, T_2)$, where thus $T_1 = t_1(\mathbf{X})$ and $T_2 = t_2(\mathbf{X})$. Then Equation (2.3) and Properties (P1)-(P2) make $\mathbf{T}$ being an equivariant statistic, in the sense that

$$\mathbf{T}(\boldsymbol{\theta}) = (\theta_1 + \theta_2 W_1, \theta_2 W_2), \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.5}$$

with $\mathbf{W} = t(g(\mathbf{U}))$, and hence $W_1 = t_1(g(\mathbf{U}))$ and $W_2 = t_2(g(\mathbf{U}))$.

The domain of $\mathbf{W}$ is denoted by $\mathcal{W}$, and we can for instance write it as $\mathcal{W} = \{t(g(\mathbf{u})), \mathbf{u} \in \mathcal{U}\}$, so $\mathcal{W}$ depends on the distribution of $\mathbf{U}$, via $\mathcal{U}$, as well as on the functions $t_1(\cdot)$ and $t_2(\cdot)$, via $t(\cdot)$. In the remainder, we assume that, among all the pairs $t_1(\cdot)$ and $t_2(\cdot)$ satisfying Properties (P1)-(P2), the distribution of $\mathbf{U}$ allows to find at least one such that

$$\mathcal{W} \subseteq \mathcal{P}. \tag{A3}$$

Accordingly, from now on, the functions $t_1(\cdot)$ and $t_2(\cdot)$ are considered to satisfy Assumption (A3), in addition to Properties (P1)-(P2).

Example 2.3.4. Consider the setting of Example 2.3.2, and let $t_1(\cdot)$ and $t_2(\cdot)$ be as in Example 2.3.3. Then, after a few calculation steps, we obtain $T_1(\boldsymbol{\theta}) = \theta_1 + \theta_2 U_{1:n}$ and $T_2(\boldsymbol{\theta}) = \theta_2 \sum_{i=2}^{n}(U_{i:n} - U_{1:n})$, so $W_1 = U_{1:n}$ and $W_2 = \sum_{i=2}^{n}(U_{i:n} - U_{1:n})$. For Assumption (A3) to be satisfied, the domain of W_1 needs to be a subset of $\mathbb{R}$, and that of W_2 a subset of $\mathbb{R}_+$. Under the setting of Example 2.3.2, $U_1, ..., U_n$ are mutually independent Exponential(1) random variables. The domain of W_1 is thus $\mathbb{R}_+$, and it then only remains to check that of W_2. But, beforehand, we would like to recall the well-known Rényi representation [71], which for a sequence of mutually independent Exponential(1) random variables $U_1, ..., U_n$ allows to write

$$(U_{1:n}, ..., U_{n:n}) \stackrel{d}{=} \left(\sum_{j=1}^{1} \frac{U_j^*}{n-j+1}, ..., \sum_{j=1}^{n} \frac{U_j^*}{n-j+1} \right), \tag{2.6}$$

where $\stackrel{d}{=}$ denotes the equality in distribution and $U_1^*, ..., U_n^*$ are again mutually independent Exponential(1) random variables. The latter will be useful again later in this article, and here it allows to write

$$W_2 \stackrel{d}{=} \sum_{i=2}^{n} \sum_{j=2}^{i} \frac{U_j^*}{n-j+1} = \sum_{j=2}^{n} \sum_{i=j}^{n} \frac{U_j^*}{n-j+1} = \sum_{j=2}^{n} U_j^*. \tag{2.7}$$

The random variable W_2 thus follows an Erlang$(n-1, 1)$ distribution, which can be defined with either domain $\mathbb{R}_0$, or $\mathbb{R}_+$. In this article we follow the second option, leading to $\mathcal{W} = \mathbb{R}_+^2 \subset \mathcal{P}$. In conclusion, we have shown that under the setting of Example 2.3.2, the functions $t_1(\cdot)$ and $t_2(\cdot)$ introduced in example 2.3.3 ensures Assumption (A3) to be satisfied.

The domain of $\mathbf{T}(\boldsymbol{\theta})$ is defined as $\mathcal{T}(\boldsymbol{\theta}) = \{\mathbf{t} \in \mathbb{R}^2 : f_{\mathbf{T}}(\mathbf{t}; \boldsymbol{\theta}) > 0\}$. Note that here and throughout this article $\mathbf{t}$ denotes a vector $\mathbf{t} = (t_1, t_2)$, where t_1 and t_2 are two quantities, which must not be confused with the functions $t_1(\cdot)$ and $t_2(\cdot)$. The domain of $\mathbf{T}$ induced by $\mathcal{P}$ is then written as $\mathcal{T} = \cup_{\boldsymbol{\theta} \in \mathcal{P}} \mathcal{T}(\boldsymbol{\theta})$, and from Assumption (A3) it satisfies

$$\mathcal{T} = \mathcal{P}. \tag{2.8}$$

Using $t_1(\cdot)$ and $t_2(\cdot)$, we then define the mapping $a(\cdot)$, as

$$a(\mathbf{x}) = \frac{g^{-1}(\mathbf{x}) - t_1(\mathbf{x})}{t_2(\mathbf{x})}, \text{ for all } \mathbf{x} \in \mathcal{X}.$$

This yields the statistic $\mathbf{A} = a(\mathbf{X}) = \frac{g^{-1}(\mathbf{X}) - T_1}{T_2}$, which from Equations (2.3) and (2.5) can be expressed as

$$\mathbf{A} = \frac{\mathbf{U} - W_1}{W_2}. \tag{2.9}$$

Since W_1 and W_2 are functions of $\mathbf{U}$ only, they do not depend on $\boldsymbol{\theta}$, and hence nor does the distribution of $\mathbf{A}$, which by definition makes it being an ancillary statistic (see e.g. Basu [10] for some background).

Besides that, from $\mathbf{T}$ and Y we define the random variable $B = \frac{h^{-1}(Y) - T_1}{T_2}$, and according to Equations (2.4) and (2.5), the latter can be expressed as

$$B = \frac{V - W_1}{W_2}. \tag{2.10}$$

Therefore, as for $\mathbf{A}$, the distribution B does not depend on $\boldsymbol{\theta}$. However, B is calculated from both $\mathbf{X}$ (via $\mathbf{T}$) and Y, and it is thus not a statistic of $\mathbf{X}$, contrarily to $\mathbf{A}$. Also, note that Assumption (A1) makes $\mathbf{U} = g^{-1}(\mathbf{X}(\mathbf{e}))$ and $V = h^{-1}(Y(\mathbf{e}))$ being independent. Therefore, since V is a continuous random variable, so is B.

Finally, notice that by construction $\mathbf{X}$ and Y can be rewritten using $\mathbf{A}$, B and $\mathbf{T}$, as

$$\mathbf{X} = g(T_1 + T_2 \mathbf{A}), \tag{2.11}$$

and

$$Y = h(T_1 + T_2 B), \tag{2.12}$$

which will be helpful later.

2.4 Expected Solvency probability

In the setting of Section 2.3, assume that we need to determine an amount of capital to back the future loss Y, so as to reach a given target solvency probability of $p \in [0, 1]$, where p is typically smaller but close to 1. By definition, this amount of capital corresponds to the value-at-risk of Y at level p, and shall therefore be calculated as

$$\text{VaR}_p[Y(\boldsymbol{\theta})] = \inf\{x \in \mathbb{R} : F_Y(x; \boldsymbol{\theta}) \geq p\}.$$

However, we do not know the value of the parameter $\boldsymbol{\theta}$, so we cannot calculate the above quantity. Let $\mathbf{x}$ be the observed realization of $\mathbf{X}$, and remember that the distribution of the latter random vector also depends on $\boldsymbol{\theta}$. The data $\mathbf{x}$ thus contains information about that unknown parameter, and one aims to use it for estimating the above value-at-risk.

Let $\delta_p(\mathbf{x})$ be such an estimate, where $\delta_p(\cdot)$ is some estimator function. Of course, for one particular value of the observed sample $\mathbf{x}$, the estimated value-at-risk $\delta_p(\mathbf{x})$ is likely to differ from $\text{VaR}_p[Y(\boldsymbol{\theta})]$. Therefore, by holding $\delta_p(\mathbf{x})$ as capital, the resulting solvency probability $F_Y(\delta_p(\mathbf{x}); \boldsymbol{\theta})$ is likely to differ from the targeted one of p. Yet, in order to assess the reliability of the estimator $\delta_p(\cdot)$, we may consider the resulting expected solvency probability

$$\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \delta_p(\mathbf{X})) = \mathbb{E}_{\boldsymbol{\theta}}[F_Y(\delta_p(\mathbf{X}); \boldsymbol{\theta})] = \int_{\mathbf{x} \in \mathcal{X}(\boldsymbol{\theta})} F_Y(\delta_p(\mathbf{x}); \boldsymbol{\theta}) f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) \, \mathrm{d}\mathbf{x}, \tag{2.13}$$

which can be interpreted as follows. Assume that $\mathbf{X}$ has not yet realized, but we already decided to rely on some estimator $\delta_p(\cdot)$, for calculating the future capital that will back Y. At that point, the capital turns out to be the random variable $\delta_p(\mathbf{X})$, of which $\delta_p(\mathbf{x})$ is thus the future realization. Accordingly, the resulting solvency probability is the random variable $F_Y(\delta_p(\mathbf{X}); \boldsymbol{\theta})$, and probability (2.13) is simply its expectation.

In some situations, it is possible to find one or more estimators $\delta_p(\cdot)$ with the property to make the above expected solvency probability always equal the targeted one of p, i.e.

$$\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \delta_p(\mathbf{X})) = p, \text{ for all } \boldsymbol{\theta} \in \mathcal{P}. \tag{2.14}$$

This potential property of an estimator makes it a good candidate for capital calculation, since it ensures that the resulting solvency probability can be expected to equal the target solvency probability, despite parameter uncertainty. By contrast, an estimator that does not have property (2.14) can hardly be regarded as suitable for capital calculation under parameter uncertainty.

Remark 2.4.1. In the literature, the above property has been considered multiple times, by several authors and under different denominations. For instance, Francioni and Herzog [41] called an estimator $\delta_p(\cdot)$ satisfying Equation (2.14), a probability unbiased value-at-risk estimator, Bignozzi and Tsanakas [13] said it to lead to a residual risk of zero, Pitera and Schmidt [70] referred to it simply as an unbiased estimator, and Jarvis et al. [54] said it to pass some percentile test (note that in the latter contribution, the reference distribution

function of the maximum multiplier method yields infinite mean and variance, in contrast to what was calculated in Table 2, and in the subsequent illustrations). Besides that, in statistics (2.13) is viewed as a frequentist probability, and a value-at-risk (quantile) estimator with property (2.14) is therefore said to be valid from a frequentist point of view, see for instance Mukerjee and Dey [65].

2.5 Capital calculation under parameter uncertainty

2.5.1 Maximum likelihood estimation

In order to estimate the desired value-at-risk, a first option often consists to compute the maximum likelihood estimate of $\boldsymbol{\theta}$,

$$\hat{\boldsymbol{\theta}}(\mathbf{x}) = \underset{\boldsymbol{\theta}\in\mathcal{P}}{\arg\max}\, f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}), \tag{2.15}$$

then pick the corresponding distribution in $\mathcal{D}_Y$ for describing the risk Y, and finally come up with the resulting estimated value-at-risk

$$\hat{\delta}_p(\mathbf{x}) = \text{VaR}_p[Y(\hat{\boldsymbol{\theta}}(\mathbf{x}))] = \inf\{x\in\mathbb{R} : F_Y(x;\hat{\boldsymbol{\theta}}(\mathbf{x}))\geq p\}. \tag{2.16}$$

In their Proposition 1, Gerrard and Tsanakas [47] have shown that when each $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ are transformed location-scale families, and $X_1(\boldsymbol{\theta}),...,X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ are mutually independent given $\boldsymbol{\theta}$, the resulting expected solvency probability $\mathbb{P}_{\boldsymbol{\theta}}(Y\leq\hat{\delta}_p(\mathbf{X}))$ does not depend on $\boldsymbol{\theta}$ and can thus be calculated despite not knowing the value of that parameter. Also, and more importantly, it turns out that for finite sample size n, that probability does not equal the target solvency probability of p in general, and for relevant cases (essentially p close to 1) it is even strictly smaller than p. This is shown for various distribution families in [47], and we illustrate it for the Pareto family in Example 2.5.1 below.

So the estimator $\hat{\delta}_p(\cdot)$ generally leads to a capital whose solvency probability cannot be expected to equal the targeted one, and this can be understood as follows. When estimating $\text{VaR}_p[Y(\boldsymbol{\theta})]$ with the estimated distribution $F_Y(\cdot;\hat{\boldsymbol{\theta}}(\mathbf{x}))$, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}(\mathbf{x})$ is implicitly treated as the true parameter value, which boils down to ignore parameter uncertainty. Indeed, while – given $\hat{\boldsymbol{\theta}}(\mathbf{x})$ – the distribution $F_Y(\cdot;\hat{\boldsymbol{\theta}}(\mathbf{x}))$ is one the of most likely true descriptions of Y's aleatoric uncertainty, it is only one among all the other possible distributions in $\mathcal{D}_Y$. Depending on $\mathbf{x}$, the individual parameters in $\hat{\boldsymbol{\theta}}(\mathbf{x})$ can be either under- or overestimating their true value. Those two kinds of estimation errors tend to have non-symmetric effects on the resulting solvency probability $F_Y(\hat{\delta}_p(\mathbf{x});\boldsymbol{\theta})$, which usually do not compensate on average. Therefore, by ignoring parameter uncertainty, these eventual non-symmetric effects are not taken into account, and this generally makes the resulting expected solvency probability $\mathbb{P}_{\boldsymbol{\theta}}(Y\leq\hat{\delta}_p(\mathbf{X}))$ differ from the targeted one of p.

Example 2.5.1. Consider the setting of Example 2.3.2, so that $X_1(\boldsymbol{\theta}),...,X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ are iid Pareto$(e^{\theta_1},\theta_2^{-1})$ random variables. The distribution of $Y(\boldsymbol{\theta})$ is thus $F_Y(x;\boldsymbol{\theta}) =$

$1-(\frac{e^{\theta_1}}{x})^{\theta_2^{-1}}$, for all $x \geq e^{\theta_1}$ and $\boldsymbol{\theta} \in \mathcal{P}$, and hence the value-at-risk to be estimated may be expressed as $\mathrm{VaR}_p[Y(\boldsymbol{\theta})] = e^{\theta_1 - \theta_2 \ln(1-p)}$. Regarding the pdf of $\mathbf{X}(\boldsymbol{\theta})$, it can be written as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) &= \prod_{i=1}^{n} f_{X_i}(x_i;\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_i}(x_{i:n};\boldsymbol{\theta}) \\ &= \theta_2^{-n} e^{n\theta_1\theta_2^{-1}} \left(\prod_{i=1}^{n} x_{i:n}\right)^{-\theta_2^{-1}-1} \mathbb{1}\{x_{1:n} > e^{\theta_1}\}, \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \end{aligned}$$

which yields maximum likelihood estimates $\hat{\theta}_1(\mathbf{x}) = t_1(\mathbf{x})$ and $\hat{\theta}_2(\mathbf{x}) = t_2(\mathbf{x})n^{-1}$, where $t_1(\mathbf{x}) = \ln(x_{i:n})$ and $t_2(\mathbf{x}) = \sum_{i=2}^{n}(\ln(x_{i:n}) - \ln(x_{1:n}))$, as in Example 2.3.3. The resulting estimated value-at-risk is then $\hat{\delta}_p(\mathbf{x}) = e^{t_1(\mathbf{x}) - t_2(\mathbf{x})n^{-1}\ln(1-p)}$. Letting $T_1 = t_1(\mathbf{X})$ and $T_2 = t_2(\mathbf{X})$, we obtain the expected solvency probability $\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \hat{\delta}(\mathbf{X})) = \mathbb{P}(Y(\boldsymbol{\theta}) \leq e^{T_1(\boldsymbol{\theta}) - T_2(\boldsymbol{\theta})n^{-1}\ln(1-p)})$, for all $\boldsymbol{\theta} \in \mathcal{P}$. Since $T_1(\boldsymbol{\theta}) = \theta_1 + \theta_2 W_1$ and $T_2(\boldsymbol{\theta}) = \theta_2 W_2$, the latter simplifies to

$$\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \hat{\delta}_p(\mathbf{X})) = \mathbb{P}_{\boldsymbol{\theta}}(B \leq n^{-1}\ln(1-p)), \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.17}$$

where $B = \frac{V - W_1}{W_2}$, as defined in Section 2.3.3, and with $W_1 = U_{1:n}$ and $W_2 = \sum_{i=2}^{n}(U_{i:n} - U_{1:n})$. Since $\mathbf{U}$ and V are independent, so are V and (W_1, W_2). Then, on the one hand, V is an Exponential(1) random variable. On the other hand, since the setting here is the same as that of Example 2.3.4, from Equations (2.6) and (2.7) we have that W_1 and W_2 are jointly distributed as $W_1^* = U_1^*/n$ and $W_2^* = \sum_{i=2}^{n} U_i^*$, which are two independent random variables, respectively following an Exponential(n) and Erlang$(n-1, 1)$ distribution. Therefore, since V, W_1 and W_2 are mutually independent and do not depend on $\boldsymbol{\theta}$, after some calculation steps we obtain

$$\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \hat{\delta}_p(\mathbf{X})) = 1 - \frac{n}{n+1}\Big(1 - n^{-1}\ln(1-p)\Big)^{-n+1}, \text{ for all } \boldsymbol{\theta} \in \mathcal{P}. \tag{2.18}$$

Hence, while in the limit have

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}}(Y \leq \hat{\delta}_p(\mathbf{X})) = 1 - e^{\ln(1-p)} = p, \text{ for all } \boldsymbol{\theta} \in \mathcal{P},$$

for finite n the expected solvency probability differs from p in general, and it is strictly smaller than p, for p smaller but close to 1. This is illustrated in Figure 2.1, where we plotted the expected solvency probability (2.18) as a function of n, for $p \in \{0.99, 0.995, 0.999\}$.

2.5.2 Predictive estimation

In order to account for parameter uncertainty while estimating the value-at-risk $\mathrm{VaR}_p[Y(\boldsymbol{\theta})]$, an alternative option consists to rely on a predictive distribution, whose calculation can be described as follows. In a first step, the unknown parameter $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is replaced by an independent random vector $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)$, so that $\mathbf{X} = \mathbf{X}(\boldsymbol{\theta})$ and $Y = Y(\boldsymbol{\theta})$ become the mixtures $\mathbf{X} = \mathbf{X}(\boldsymbol{\theta})$ and $Y = Y(\boldsymbol{\Theta})$, and we may write $\mathbf{X}(\boldsymbol{\theta}) = (\mathbf{X}|\boldsymbol{\Theta} = \boldsymbol{\theta})$ and

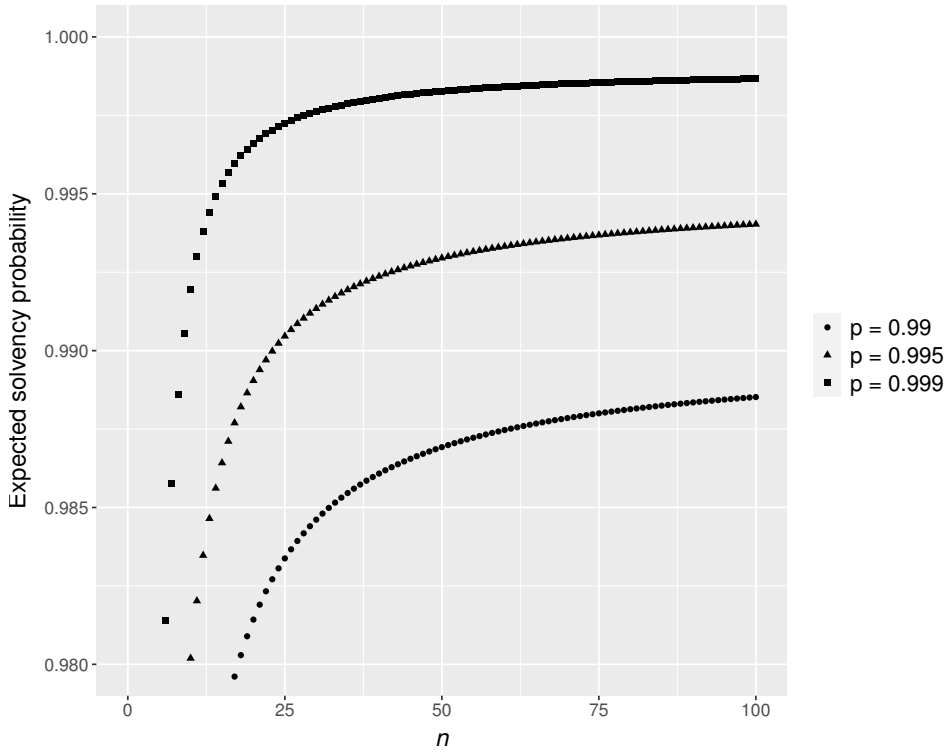

Figure 2.1: Expected solvency probability $\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \hat{\delta}_p(\mathbf{X}))$ as a function of the sample size n, for a Pareto random loss Y, and target solvency probabilities $p \in \{0.99, 0.995, 0.999\}$.

$Y(\boldsymbol{\theta}) = (Y|\boldsymbol{\Theta} = \boldsymbol{\theta})$. The random parameter $\boldsymbol{\Theta}$ is assumed to be distributed according to some prior pdf $f_{\boldsymbol{\Theta}}(\cdot)$, which is meant to represent our epistemic (parameter) uncertainty, before considering the data. On observing $\mathbf{x}$, that prior is then updated according to Bayes' formula, leading to the posterior pdf

$$f_{\boldsymbol{\Theta}|\mathbf{X}}(\cdot|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x};\cdot) f_{\boldsymbol{\Theta}}(\cdot)}{\int_{\boldsymbol{\theta}\in\mathcal{P}} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}. \tag{2.19}$$

The predictive distribution of Y can be subsequently obtained by averaging all the distributions in $\mathcal{D}_Y$ according to their posterior probabilities, as

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = \int_{\boldsymbol{\theta}\in\mathcal{P}} F_Y(\cdot;\boldsymbol{\theta}) f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x})\,\mathrm{d}\boldsymbol{\theta}. \tag{2.20}$$

That is, given prior $f_{\boldsymbol{\Theta}}(\cdot)$ and data $\mathbf{x}$, for all $\boldsymbol{\theta} \in \mathcal{P}$ the quantity $f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x})\,\mathrm{d}\boldsymbol{\theta}$ is the posterior probability attributed to $\boldsymbol{\theta}$ of being the true value of the unknown parameter, and hence to $F_Y(\cdot;\boldsymbol{\theta})$ of being the true description of Y's aleatoric uncertainty.

That predictive distribution can then be used to estimate $\mathrm{VaR}_p[Y(\boldsymbol{\theta})]$ with the resulting predictive value-at-risk

$$\tilde{\delta}_p(\mathbf{x}) = \mathrm{VaR}_p[Y|\mathbf{X} = \mathbf{x}] = \inf\{x \in \mathbb{R} : F_{Y|\mathbf{X}}(x|\mathbf{x}) \geq p\}. \tag{2.21}$$

By construction, the predictive distribution $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$ accounts for both the aleatoric and parameter uncertainty relative to the future loss Y, and hence so does the predictive value-at-risk $\tilde{\delta}_p(\mathbf{x})$. Yet, the reliability of the underlying estimator $\tilde{\delta}_p(\cdot)$ will depend on the particular prior it stems from, and the choice of the latter is therefore an important step.

In the capital calculation context that we are considering here, it is natural to choose $f_{\boldsymbol{\Theta}}(\cdot)$ such that the resulting estimator $\tilde{\delta}_p(\cdot)$ is endowed with property (2.14). As proven in Theorem 2.5.1 below, in the setting of this article, that is achieved by the prior

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \propto \theta_2^{-1}, \text{ for all } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.22}$$

which is the well-known Jeffreys' independence prior or right Haar prior for a location-scale parameter. This prior has several useful properties that makes it being viewed as some sort of objective or neutral state of epistemic uncertainty (see for instance Kass and Wasserman [57] for details).

Remark 2.5.1. Note that (2.22) defines an improper probability distribution, i.e. one cannot find a normalizing constant that makes it integrate to 1 on its domain. According to the axioms of probability theory, improper priors are thus not probability distributions. This being said, the posterior and predictive distribution resulting from an improper prior may still be proper probability distributions. Indeed, when computing a posterior according to Bayes' formula (e.g. Equation (2.19) above), the prior is involved in a way that makes its normalizing constant (when it exists) cancel out. Therefore, the normalizing constant of a prior virtually plays no role in that formula, and its eventual non-existence (because of impropriety of the prior) does not necessarily halt for obtaining a proper posterior and predictive distribution. In fact, an improper prior can usually be formulated as the limit of a sequence of proper priors (see Jaynes [55]). And while in the limit the prior becomes an improper probability distribution, the posterior and predictive distribution in which it is involved can remain proper.

Notice that, as for $\mathbf{X}$ and Y, replacing $\boldsymbol{\theta}$ by $\boldsymbol{\Theta}$ makes the statistic $\mathbf{T} = \mathbf{T}(\boldsymbol{\theta})$ become the mixture $\mathbf{T} = \mathbf{T}(\boldsymbol{\Theta})$, and from Equation (2.5), the latter can be written as

$$\mathbf{T} = (\Theta_1 + \Theta_2 W_1, \Theta_2 W_2). \tag{2.23}$$

Moreover, since prior (2.22) makes the domain of $\boldsymbol{\Theta}$ to be $\mathcal{P}$, the respective domain of mixtures $\mathbf{X}$, Y and $\mathbf{T}$ are now $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{T}$, as previously introduced.

Consider then the subsequent lemma.

Lemma 2.5.1. *Let each $\mathcal{D}_{X_i}$ be a transformed location-scale family. Then under Assumptions* (A1)-(A3)*, the prior* (2.22) *makes* $\mathbf{U}$ *being independent of* $\mathbf{T}$.

An interesting discussion arises from that lemma, on the connection between Bayesian and fiducial inference. But since that discussion is related to statistical inference in general rather than capital calculation, we delegate it to the appendix in Section 2.9. For our purpose here, Lemma 2.5.1 serves to prove the following theorem, which is the main result of this section (the proof will be given in Section 2.10).

Theorem 2.5.1. *Let each $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ be transformed location-scale families, and $p \in [0,1]$. Then under Assumptions* (A1)-(A3)*, the prior* (2.22) *leads to $\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\delta}_p(\mathbf{X})) = p$, for all $\boldsymbol{\theta} \in \mathcal{P}$.*

So under the applicability conditions of Theorem 2.5.1, the value-at-risk estimator $\tilde{\delta}_p(\cdot)$ that results from the prior (2.22) satisfies property (2.14), and hence using it for capital calculation ensures that the resulting expected solvency probability is always equal to the targeted one of p, despite parameter uncertainty.

Remark 2.5.2. We would like to emphasize that Theorem 2.5.1 extends Proposition 2 in Gerrard and Tsanakas [47], by relaxing three of their assumptions. Firstly, the applicability conditions of Theorem 2.5.1 do not include that $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ are mutually independent, but only that their dependence structure does not depend on $\boldsymbol{\theta}$, as stated in Assumption (A2). This will be useful in Section 2.6, where $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ are n consecutive order statistics of a larger random sample, and hence they are not mutually independent. Secondly, Proposition 2 in [47] relies on Proposition 1 of Severini et al. [74], which requires that the domain of $\mathbf{X}(\boldsymbol{\theta})$, i.e. $\mathcal{X}(\boldsymbol{\theta})$, does not depend on $\boldsymbol{\theta}$. Regarding Theorem 2.5.1, it also applies to cases where $\mathcal{X}(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$, such as $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ being Pareto or Uniform random variables. Thirdly, in [47] the predictive distribution is built from the maximum likelihood estimate $\hat{\boldsymbol{\theta}}(\mathbf{x})$. Here the predictive distribution (2.20) is built from $\mathbf{x}$ directly. Since in both cases the resulting capital yields the targeted solvency probability, Theorem 2.5.1 shows that for achieving the latter result, the predictive distribution needs not to be built from $\hat{\boldsymbol{\theta}}(\mathbf{x})$. In fact, as proven in Proposition 2.5.2 below, if the random vector $\hat{\boldsymbol{\theta}}(\mathbf{X})$ turns out to be a sufficient statistic, then using $\hat{\boldsymbol{\theta}}(\mathbf{x})$ rather than $\mathbf{x}$ does not affect the predictive distribution, and hence it does not affect the resulting capital either. On the contrary, if $\hat{\boldsymbol{\theta}}(\mathbf{X})$ is not a sufficient statistic (e.g. in the Weibull case, see Example 2.5.3), then using $\hat{\boldsymbol{\theta}}(\mathbf{x})$ instead of $\mathbf{x}$ boils down to ignore some information contained in the data, and the capital ends up being affected. That point is further discussed in the following section.

Remark 2.5.3. Among the various assumptions on which Theorem 2.5.1 relies, there is Assumption (A1), which establishes the independences of $\mathbf{X}$ and Y. This assumption is classical in predictive inference problems, but it may be relevant to consider relaxing it in some contexts, e.g. when dealing with time series. In fact, it turns out that there are ways to allow for the dependence of Y on $\mathbf{X}$ and still obtain the result of Theorem 2.5.1, for instance by letting $\mathbf{X}$ enter as a parameter in the distribution of Y, as long as the transformed location-scale structure defined by Equation (2.4) is preserved. They may be interesting applications to consider in that direction, but we leave this for future research.

2.5.3 Calculation issues

As we just proved in Theorem 2.5.1, the value-at-risk estimator $\tilde{\delta}_p(\cdot)$ resulting from the prior (2.22) satisfies property (2.14). If one thus wishes to use it and hold $\tilde{\delta}_p(\mathbf{x})$ as capital, one needs to calculate the latter quantity. This can of course be done simply by plugging

the prior (2.22) into Formula (2.19), so as to obtain the posterior

$$f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) = \frac{f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})\theta_2^{-1}}{\int_{\boldsymbol{\theta}'\in\mathcal{P}} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}')\theta_2'^{-1}\,\mathrm{d}\boldsymbol{\theta}'}, \text{ for all } \boldsymbol{\theta}\in\mathcal{P}, \tag{2.24}$$

then calculate the resulting predictive distribution according to Formula (2.20), and finally invert it for obtaining $\tilde{\delta}_p(\mathbf{x})$, as expressed in Formula (2.21). As proven in Proposition 2.5.1 below, it turns out that when $\mathbf{T}$ is a sufficient statistic, we may proceed differently. But beforehand, we need to recall the following well-known result. The proof is provided in the appendix.

Lemma 2.5.2. *If $\mathbf{S} = s(\mathbf{X})$ is a sufficient statistic of $\mathbf{X}$, then we have $F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{Y|\mathbf{S}}(\cdot|s(\mathbf{x}))$, for all $\mathbf{x}\in\mathcal{X}$.*

With this, we are ready to prove the following proposition.

Proposition 2.5.1. *Let each $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ be transformed location-scale families, and write $t_1(\mathbf{x}) = t_1$ and $t_2(\mathbf{x}) = t_2$. If $\mathbf{T} = t(\mathbf{X})$ is a sufficient statistic, then under Assumptions* (A1)-(A3)*, prior* (2.22) *leads to*

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_B\left(\frac{h^{-1}(\cdot) - t_1}{t_2}\right),$$

and hence

$$\tilde{\delta}_p(\mathbf{x}) = \mathit{VaR}_p[Y|\mathbf{X}=\mathbf{x}] = h(t_1 + t_2\mathit{VaR}_p[B]),$$

for any $p\in[0,1]$.

So concretely, if one is able to find a pair of functions $t_1(\cdot)$ and $t_2(\cdot)$ as described in Section 2.3.3, which make $\mathbf{T} = t(\mathbf{X})$ a sufficient statistic of $\mathbf{X}$, then Proposition 2.5.1 provides a useful simplification: Indeed, if we can then find an analytical expression for the distribution of B, so we will for $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$ and $\tilde{\delta}_p(\mathbf{x})$. That case is illustrated in Example 2.5.2 and Section 2.6. If not, then we can easily approximate $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$ and $\tilde{\delta}_p(\mathbf{x})$, by simulating random variable B, from $\mathbf{U}$ and V (again, see Section 2.6).

In some situations, there may also simply be no such sufficient statistic (see Example 2.5.3). Obviously, then Proposition 2.5.1 cannot be applied, and hence for calculating $\tilde{\delta}_p(\mathbf{x})$ one must proceed as previously described, using posterior (2.24). An alternative consists of replacing the sample $\mathbf{X}$ by a suitable equivariant statistic $\mathbf{T}$, and hence consider the predictive distribution $F_{Y|\mathbf{T}}(\cdot|\mathbf{t})$ instead of $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$. It is easy to see from Lemma 2.5.2 that when $\mathbf{T}$ is not a sufficient statistic of $\mathbf{X}$, the two latter predictive distributions differ (and hence so will the resulting calculated capital amounts), meaning that considering $F_{Y|\mathbf{T}}(\cdot|\mathbf{t})$ instead of $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$ boils down to ignore some part of the information contained in the data $\mathbf{x}$. Yet, by doing so the predictive distribution $F_{Y|\mathbf{T}}(\cdot|\mathbf{t})$ may be easier to calculate, and the resulting capital estimator still satisfies property (2.14), as hereafter proven.

Proposition 2.5.2. *Let each $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ be transformed location-scale families, and write $\mathbf{t} = (t_1, t_2)$, where $t_1(\mathbf{x}) = t_1$ and $t_2(\mathbf{x}) = t_2$. Then under Assumptions* (A1)-(A3)*, the prior* (2.22) *leads to*

$$F_{Y|\mathbf{T}}(\cdot|\mathbf{t}) = F_B\left(\frac{h^{-1}(\cdot) - t_1}{t_2}\right),$$

and hence

$$\tilde{\eta}_p(\mathbf{t}) = \textit{VaR}_p[Y|\mathbf{T} = \mathbf{t}] = h(t_1 + t_2 \textit{VaR}_p[B]),$$

for any $p \in [0, 1]$. Moreover, the value-at-risk estimator $\tilde{\eta}_p(\cdot)$ satisfies $\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\eta}_p(\mathbf{T})) = p$, for all $\boldsymbol{\theta} \in \mathcal{P}$.

Remark 2.5.4. We would like to underline that according to Proposition 2.5.2, any equivariant statistic $\mathbf{T}$ leads to a value-at-risk estimator $\tilde{\eta}_p(\cdot)$ which satisfies property (2.5.2), as long as Assumption (A3) holds. Since different equivariant statistics in general lead to different value-at-risk estimators, Proposition 2.5.2 in fact shows the potential multiplicity of estimators satisfying property (2.14). However, as we previously commented, while it may be easier to calculate, considering an estimator $\tilde{\eta}_p(\cdot)$ built from an equivariant statistic which is not also sufficient boils down to throw out some part of the information contained in the data.

Example 2.5.2. Let $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ be iid Log-Normal(θ_1, θ_2) random variables. As described in Table (2.1), the Log-Normal distribution defines a transformed location-scale family. Also, since $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ are mutually independent, Assumptions (A1) and (A2) are both satisfied. The maximum likelihood estimates of θ_1 and θ_2 are then $\hat{\theta}_1(\mathbf{x}) = n^{-1}\sum_{i=1}^n \ln(x_i)$ and $\hat{\theta}_2(\mathbf{x}) = (n^{-1}\sum_{i=1}^n (x_i - \hat{\theta}_1(\mathbf{x}))^2)^{1/2}$. It can be checked that the functions $t_1(\cdot) = \hat{\theta}_1(\cdot)$ and $t_2(\cdot) = \hat{\theta}_2(\cdot)$ satisfy Properties (P1)-(P2) (see e.g. Lemma 3 in Gerrard and Tsanakas [47]), which endows the resulting statistic $\mathbf{T} = (t_1(\mathbf{X}), t_2(\mathbf{X}))$ with the equivariance property

$$\mathbf{T}(\boldsymbol{\theta}) = (\theta_1 + \theta_2 W_1, \theta_2 W_2), \text{ for all } \boldsymbol{\theta} \in \mathcal{P},$$

where $\mathbf{W} = (W_1, W_2)$ has domain $\mathcal{P}$. Assumption (A3) is thus satisfied too, and the applicability conditions of Proposition (2.5.2) therefore hold. It is then known that under prior (2.22) (see Example 6 in Bignozzi and Tsanakas [13]), the predictive distribution of Y given $\mathbf{T}$ is

$$F_{Y|\mathbf{T}}(\cdot|\mathbf{t}) = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{\ln(\cdot) - t_1}{t_2}\right),$$

where $t_{n-1}(\cdot)$ denotes the distribution of a standard Student-t random variable with $n - 1$ degrees of freedom. From Proposition 2.5.2 the resulting estimator $\tilde{\eta}_p(\cdot)$ satisfies property (2.14), and hence using it for calculating the capital always leads to an expected solvency probability of exactly p, i.e. $\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\eta}_p(\mathbf{X})) = p$, for all $\boldsymbol{\theta} \in \mathcal{P}$. Finally, it can be shown that here $\mathbf{T}$ is a sufficient statistic of $\mathbf{X}$, which with Proposition 2.5.1 leads to

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{Y|\mathbf{T}}(\cdot|\mathbf{t}),$$

and hence $\tilde{\delta}_p(\cdot) = \tilde{\eta}_p(\cdot)$.

Example 2.5.3. Let $X_1(\boldsymbol{\theta}), ..., X_n(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ be iid Weibull random variables, with scale parameter $e^{\theta_1} > 0$ and shape parameter θ_2^{-1}. As described in Table 2.1, the Weibull distribution defines a transformed location-scale family, and the pdf of $\mathbf{X}(\boldsymbol{\theta})$ can be written as

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = \theta_2^{-n} e^{-n\theta_1 \theta_2^{-1}} \exp\left(- e^{-\theta_1} \sum_{i=1}^{n} x_i^{\theta_2^{-1}} \right) \prod_{i=1}^{n} x_i^{\theta_2^{-1}-1}.$$

For $n = 2$, we may take for instance $t_1(\cdot)$ and $t_2(\cdot)$ as in Example 2.3.3, and verify that then $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ factors as

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta}) = \gamma_1(t(\mathbf{x}); \boldsymbol{\theta}) \gamma_2(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.25}$$

meaning that the resulting $\mathbf{T} = t(\mathbf{X})$ is, in that particular case, a sufficient statistic of $\mathbf{X}$. But for $n > 2$, the pair $t_1(\cdot)$ and $t_2(\cdot)$ does not allow to factor $f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})$ as in Equation (2.25), and $\mathbf{T}$ is thus no more a sufficient statistic of $\mathbf{X}$. In fact, for $n > 2$, the sum in Equation (2.25) makes it impossible to find a two-dimensional sufficient statistic for $\mathbf{X}$.

2.6 The Pareto case

In this section, we illustrate Theorem 2.5.1 and Proposition 2.5.1, for the special case where the data is given by n consecutive order statistics of a larger Pareto random sample. We provide here more details than in Example 2.5.2, and hence we dedicated an entire section to the present illustration.

For $k \geq 2$, let $\tilde{X}_1(\boldsymbol{\theta}), ..., \tilde{X}_k(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ be iid Pareto random variables, again with scale parameter $e^{\theta_1} > 0$ and shape parameter $\theta_2^{-1} > 0$. The setting is thus similar to that of Example 2.3.2, but here we observe only

$$\mathbf{X} = (X_1, ..., X_n) = (\tilde{X}_{m+1:k}, ..., \tilde{X}_{m+n:k}),$$

i.e. the $(m+1)^{th}$ to the $(m+n)^{th}$ order statistics, with $0 \leq m \leq k - n$ and $n \geq 2$ (note that for $m = 0$ and $n = k$ we recover the standard case where the whole sample is observed).

We now check the applicability conditions of Theorem 2.5.1 and Proposition 2.5.1: Firstly, since $\tilde{X}_1(\boldsymbol{\theta}), ..., \tilde{X}_k(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ are iid, $\mathbf{X}(\boldsymbol{\theta})$ and $Y(\boldsymbol{\theta})$ are independent, and hence Assumption (A1) is satisfied. Let $\tilde{U}_1, ..., \tilde{U}_k$ and V, be mutually independent Exponential(1), and define

$$\mathbf{U} = (U_1, ..., U_n) = (\tilde{U}_{m+1:k}, ..., \tilde{U}_{m+n:k}).$$

We can then write $X_i(\boldsymbol{\theta}) = e^{\theta_1 + \theta_2 U_i}$, for each i, and $Y(\boldsymbol{\theta}) = e^{\theta_1 + \theta_2 V}$, for all $\boldsymbol{\theta} \in \mathcal{P}$, meaning that each $\mathcal{D}_{X_i}$ and $\mathcal{D}_Y$ are transformed location-scale families. Also, the copula of $\mathbf{U}$ does not depend on $\boldsymbol{\theta}$, and hence nor does that of $\mathbf{X}(\boldsymbol{\theta})$, so Assumption (A2) is satisfied. Regarding the functions $t_1(\cdot)$ and $t_2(\cdot)$, we consider

$$t_1(\mathbf{x}) = \ln(x_{1:n}), \text{ for all } \mathbf{x} \in \mathcal{X}, \tag{2.26}$$

and

$$t_2(\mathbf{x}) = \sum_{i=2}^{n}(\ln(x_{i:n})-\ln(x_{1:n}))+(k-m-n)(\ln(x_{n:n})-\ln(x_{1:n})), \text{ for all } \mathbf{x} \in \mathcal{X}. \quad (2.27)$$

Here $t_1(\cdot)$ is as in Example 2.3.3, so it satisfies Property (P1). Then, since

$$\begin{aligned} t_2(g(\theta_1 + \theta_2 g^{-1}(\mathbf{x}))) &= \sum_{i=2}^{n}(\theta_1 \ln(x_{i:n}) - \theta_2 \ln(x_{1:n})) + (k-m-n)(\theta_2 \ln(x_{n:n}) - \theta_2 \ln(x_{1:n})) \\ &= \theta_2 t_2(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \end{aligned}$$

the function $t_2(\cdot)$ satisfies Property (P2). The resulting statistic $\mathbf{T}$ therefore satisfies the equivariance property $\mathbf{T}(\boldsymbol{\theta}) = (\theta_1 + \theta_2 W_1, \theta_2 W_2)$, for all $\boldsymbol{\theta} \in \mathcal{P}$, where

$$W_1 = U_{1:n} = \tilde{U}_{m+1:k},$$

and

$$\begin{aligned} W_2 &= \sum_{i=2}^{n}(U_{i:n} - U_{1:n}) + (k-m-n)(U_{n:n} - U_{1:n}) \\ &= \sum_{i=2}^{n}(\tilde{U}_{m+i:k} - \tilde{U}_{m+1:k}) + (k-m-n)(\tilde{U}_{m+n:k} - \tilde{U}_{m+1:k}). \end{aligned}$$

Since from the Rényi representation we have

$$(\tilde{U}_{1:k}, ..., \tilde{U}_{k:k}) \stackrel{d}{=} \left(\sum_{j=1}^{1} \frac{\tilde{U}_j^*}{k-j+1}, ..., \sum_{j=1}^{k} \frac{\tilde{U}_j^*}{k-j+1} \right), \quad (2.28)$$

where $\tilde{U}_1^*, ..., \tilde{U}_k^*$ are iid Exponential(1) random variables, it turns out that W_1 and W_2 are jointly distributed as

$$W_1^* = \tilde{U}_{m+1:k}^* = \sum_{j=1}^{m+1} \frac{\tilde{U}_j^*}{k-j+1}$$

and

$$\begin{aligned} W_2^* &= \sum_{i=2}^{n} \sum_{j=m+2}^{m+i} \frac{\tilde{U}_j^*}{k-j+1} + (k-m-n) \sum_{j=m+2}^{m+n} \frac{\tilde{U}_j^*}{k-j+1} \\ &= \sum_{i=2}^{n} \sum_{j=2}^{i} \frac{\tilde{U}_{j+m}^*}{k-j-m+1} + (k-m-n) \sum_{j=2}^{n} \frac{\tilde{U}_{j+m}^*}{k-j-m+1} \\ &= \sum_{j=2}^{n} \sum_{i=j}^{n} \frac{\tilde{U}_{j+m}^*}{k-j-m+1} + (k-m-n) \sum_{j=2}^{n} \frac{\tilde{U}_{j+m}^*}{k-j-m+1} \\ &= \sum_{j=2}^{n} \frac{n-j+1}{k-j-m+1} \tilde{U}_{j+m}^* + (k-m-n) \sum_{j=2}^{n} \frac{\tilde{U}_{j+m}^*}{k-j-m+1} \\ &= \sum_{j=2}^{n} \tilde{U}_{j+m}^* = \sum_{j=m+2}^{m+n} \tilde{U}_j^*, \end{aligned}$$

so W_1 and W_2 are independent, with W_1 being the $(m+1)^{th}$ order statistic of k iid Exponential(1) random variables and W_2 follows an Erlang$(n-1,1)$ distribution. Accordingly, the domain of $\mathbf{W}$ is $\mathcal{W} = \mathbb{R}_+ \subset \mathcal{P}$, and Assumption (A3) is thus satisfied. Finally, it turns out that $\mathbf{T}$ is a sufficient statistic. Indeed, since $X_1, ..., X_n$ are n consecutive order statistics of iid random variables $\tilde{X}_1, ..., \tilde{X}_k$, we have (see e.g. [3])

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) =& \frac{k!}{m!(k-m-n)!} F_{\tilde{X}}(x_1;\boldsymbol{\theta})^m (1 - F_{\tilde{X}}(x_n;\boldsymbol{\theta}))^{k-m-n} \\ & \cdot \prod_{i=1}^{n} f_{\tilde{X}}(x_i;\boldsymbol{\theta}) \mathbb{1}\{x_1 \leq ... \leq x_n\}, \text{ for all } \mathbf{x} \in \mathbb{R}^n \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \end{aligned}$$

where $F_{\tilde{X}}(x_i;\boldsymbol{\theta}) = F_{\tilde{X}_{m+i:k}}(x_i;\boldsymbol{\theta})$ and $f_{\tilde{X}}(x_i;\boldsymbol{\theta}) = f_{\tilde{X}_{m+i:k}}(x_i;\boldsymbol{\theta})$, for each i. With this, we may define $\mathcal{X}(\boldsymbol{\theta}) = \{\mathbf{x} \in \mathbb{R}^n : f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) > 0\}$, for all $\boldsymbol{\theta} \in \mathcal{P}$, and $\mathcal{X} = \cup_{\boldsymbol{\theta}\in\mathcal{P}} \mathcal{X}(\boldsymbol{\theta})$, so that from the term $\mathbb{1}\{x_1 \leq ... \leq x_n\}$, if $\mathbf{x} = (x_1, ..., x_n) \in \mathcal{X}$, then $\mathbf{x} = (x_{1:n}, ..., x_{n:n})$ necessarily, leading to

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) =& \frac{k!}{m!(k-m-n)!} F_{\tilde{X}}(x_{1:n};\boldsymbol{\theta})^m (1 - F_{\tilde{X}}(x_{n:n};\boldsymbol{\theta}))^{k-m-n} \\ & \cdot \prod_{i=1}^{n} f_{\tilde{X}}(x_{i:n};\boldsymbol{\theta}), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}. \end{aligned}$$

Since we are dealing with the Pareto distribution, we thus obtain

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) = & \frac{k!}{m!(k-m-n)!} F_{\tilde{X}}(x_{1:n};\boldsymbol{\theta})^m \left(\frac{e^{\theta_1}}{x_{n:n}}\right)^{(k-m-n)\theta_2^{-1}} \\ & \cdot \theta_2^{-n} e^{n\theta_1\theta_2^{-1}} \left(\prod_{i=1}^{n} x_{i:n}\right)^{-\theta_2^{-1}-1} \mathbb{1}\{x_{1:n} > e^{\theta_1}\}, \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}. \end{aligned}$$

Letting $\varsigma(\boldsymbol{\theta}) = \theta_2^{-n} e^{(k-m)\theta_1\theta_2^{-1}}$ and $\gamma_2(\mathbf{x}) = \frac{k!}{m!(k-m-n)!}(\prod_{i=1}^{n} x_{i:n})^{-1}$, this results in

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) &= \varsigma(\boldsymbol{\theta}) F_{\tilde{X}}(x_{1:n};\boldsymbol{\theta})^m x_{n:n}^{-(k-m-n)\theta_2^{-1}} \left(\prod_{i=1}^{n} x_{i:n}\right)^{-\theta_2^{-1}} \mathbb{1}\{x_{1:n} > e^{\theta_1}\}\gamma_2(\mathbf{x}) \\ &= \varsigma(\boldsymbol{\theta}) F_{\tilde{X}}(x_{1:n};\boldsymbol{\theta})^m \left(x_{n:n}^{(k-m-n)} x_{1:n}^{n} \prod_{i=1}^{n} \frac{x_{i:n}}{x_{1:n}}\right)^{-\theta_2^{-1}} \mathbb{1}\{x_{1:n} > e^{\theta_1}\}\gamma_2(\mathbf{x}) \\ &= \varsigma(\boldsymbol{\theta}) F_{\tilde{X}}(x_{1:n};\boldsymbol{\theta})^m \left(x_{1:n}^{k-m} \left(\frac{x_{n:n}}{x_{1:n}}\right)^{(k-m-n)} \prod_{i=1}^{n} \frac{x_{i:n}}{x_{1:n}}\right)^{-\theta_2^{-1}} \mathbb{1}\{x_{1:n} > e^{\theta_1}\}\gamma_2(\mathbf{x}) \\ &= \varsigma(\boldsymbol{\theta}) F_{\tilde{X}}(e^{t_1(\mathbf{x})};\boldsymbol{\theta})^m \left(e^{(k-m)t_1(\mathbf{x})+t_2(\mathbf{x})}\right)^{-\theta_2^{-1}} \mathbb{1}\{t_1(\mathbf{x}) > \theta_1\}\gamma_2(\mathbf{x}) \\ &= \gamma_1(t(\mathbf{x});\boldsymbol{\theta})\gamma_2(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \end{aligned}$$

where $\gamma_1(t(\mathbf{x});\boldsymbol{\theta}) = \varsigma(\boldsymbol{\theta}) F_{\tilde{X}}(e^{t_1(\mathbf{x})};\boldsymbol{\theta})^m \left(e^{(k-m)t_1(\mathbf{x})+t_2(\mathbf{x})}\right)^{-\theta_2^{-1}} \mathbb{1}\{t_1(\mathbf{x}) > \theta_1\}$, and by definition $\mathbf{T} = t(\mathbf{X})$ is a sufficient statistic of $\mathbf{X}$.

The applicability conditions of Theorem 2.5.1 and Proposition 2.5.1 are thus satisfied, and the prior (2.22) leads to

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_B\left(\frac{\ln(\cdot) - t_1}{t_2}\right)$$

where $t_1 = t_1(\mathbf{x})$ and $t_2 = t_2(\mathbf{x})$. Here in $B = \frac{V-W_1}{W_2}$, the random variable V is an Exponential(1), and the respective distributions of W_1 and W_2 are as stated above. Moreover, $\mathbf{U}$ and V are independent and $\mathbf{W}$ is a function of $\mathbf{U}$ only, so V and $\mathbf{W}$ are independent. Therefore, since W_1 and W_2 are also independent, the triplet V, W_1 and W_2 are mutually independent, which after some calculation steps leads to

$$F_{Y|\mathbf{X}}(x|\mathbf{x}) = 1 - \frac{k-m}{k+1}\left(1 + \frac{\ln(x) - t_1}{t_2}\right)^{-n+1}, \text{ for } x \geq e^{t_1},$$

and hence

$$\tilde{\delta}_p(\mathbf{x}) = e^{t_1 + t_2 \rho(k,m,n,p)}, \text{ for } p \geq \frac{m+1}{k+1}, \tag{2.29}$$

where $\rho(k,m,n,p) = ((\frac{k+1}{k-m})(1-p))^{\frac{-1}{n-1}}$. For $x < e^{t_1}$, we may distinguish two cases:
If $m = 0$ (and thus $n = k$, i.e. the whole sample is observed), then the predictive distribution is

$$F_{Y|\mathbf{X}}(x|\mathbf{x}) = \frac{1}{n+1}\left(1 - n\frac{\ln(x) - t_1}{t_2}\right)^{-n+1}, \text{ for } x < e^{t_1},$$

and hence

$$\tilde{\delta}_p(\mathbf{x}) = e^{t_1 + t_2 \varrho(n,p)}, \text{ for } p < \frac{1}{n+1},$$

where $\varrho(n,p) = \frac{1}{n}(1 - ((n+1)p)^{\frac{-1}{n-1}})$.
Alternatively, if $m > 0$, then we did not find an analytical expression for $F_{Y|\mathbf{X}}(\cdot|\mathbf{x})$, and hence neither for $\tilde{\delta}_p(\mathbf{x})$. So if $m > 0$ and we wish to calculate $F_{Y|\mathbf{X}}(x|\mathbf{x})$ or $\tilde{\delta}_p(\mathbf{x})$ for some $x < e^{t_1}$ or $p < \frac{m+1}{k+1}$, then we can either perform a numerical integration, using the posterior (2.24) in Formula (2.20), or we may simply approximate the distribution of $B = \frac{V-W_1}{W_2}$ by simulating V, W_1 and W_2, and then rely on the formulas provided by Proposition 2.5.1 to obtain the quantity of interest.

2.7 Capital calculation under parameter and uncertainty

In the previous section we considered the problem of capital calculation under parameter uncertainty, and hence the model was assumed to be known. We now consider the following variant, which includes a certain type of model uncertainty: Assume that we do not know the overall distribution of Y, but we know that it has a Pareto-type tail. That distribution can thus be written as

$$F(x) = 1 - \ell(x)x^{-\alpha}, \text{ for } x \geq 0,$$

for some unknown scale parameter $\alpha > 0$ and function $\ell(\cdot)$ satisfying $\lim_{x\to\infty} \ell(x) = \tau \in (0, +\infty)$. Besides, we observe the realization $\tilde{\mathbf{x}} = (\tilde{x}_1, ..., \tilde{x}_k)$ of sample $\tilde{\mathbf{X}} = (\tilde{X}_1, ..., \tilde{X}_k)$, where $\tilde{X}_1, ..., \tilde{X}_k$ are $k > 1$ mutually independent past occurrences of Y. The desired capital is expected to be high enough, and hence to lie in the region where the function $\ell(\cdot)$ is close to have converged towards its constant value τ, where the behaviour of Y is close to that of a Pareto(τ, α) random variable.

In that setting, we consider calculating the capital for backing Y, using the value-at-risk estimator of the previous section (i.e. the one for the strict Pareto distribution, given by Formula (2.29) for $p \geq \frac{m+1}{k+1}$), but to evaluate it with only the n greatest order statistics of $\tilde{\mathbf{x}}$, so with $m = k - n$ and $\mathbf{x} = (x_1, ..., x_n) = (\tilde{x}_{k-n+1:k}, ..., \tilde{x}_{k:k})$, where n is typically not too large with respect to k. Hereafter, we assess the potential performance of this approach by means of the following example.

We consider four scenarios for the true distribution of Y:

- The Pareto distribution $F(x) = 1 - (\frac{\tau}{x})^\alpha$, for $x > \tau$, with $\tau = 1$ and $\alpha = 2$.
- The Burr distribution $F(x) = 1 - \frac{\upsilon}{\upsilon + x^\alpha}$, for $x > 0$, with $\upsilon = 1$ and $\alpha = 2$.
- The Fréchet distribution $F(x) = e^{-x^{-\alpha}}$, for $x > 0$, with $\alpha = 2$.
- The one-sided Student-t distribution $F(x) = 2\int_{-\infty}^{x} \sqrt{\nu\pi}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}(\frac{\nu}{\nu+y^2})^{\frac{\nu+1}{2}}\,\mathrm{d}y - 1$, for $x > 0$, with $\nu = 2$.

If the true distribution of Y is Pareto, then the model coincides with the one under which the estimator (2.29) has been obtained, and the expected solvency probability will then be exactly p, in accordance with Theorem 2.5.1. If not, then a model misspecification is at play, and the expected solvency probability will differ from p. In the latter case, we will be able to assess the robustness of estimator (2.29), by comparing the resulting expected solvency probability with the targeted one of p.

The parameters τ, α, υ and ν were chosen so that the resulting Burr, Fréchet and one-sided Student-t distributions can all be written as $F(x) = 1 - \ell(x)x^{-2}$, where each time $\ell(x) = (1 - F(x))x^2$ and $\lim_{x\to\infty} \ell(x) = 1$. For large x, these three distributions thus behave like the Pareto distribution with scale parameter 1 and shape parameter 2. The four distributions are plotted in Figure 2.2, and the corresponding log-log plots can be found in Figure 2.3. As one can see, the Burr, Fréchet and one-sided Student-t distributions have essentially converged to the Pareto distribution already at $x \approx 5$.

For target solvency probability $p = 0.995$ and $k \in \{20, 50, 100, 200\}$, we plotted in Figure 2.4 the expected solvency probability $\mathbb{P}(Y \leq \tilde{\delta}_{0.995}(\mathbf{X}))$ resulting from those four distributions, as a function of the number of greatest order statistics n, for n ranging from 2 to 20. Notice that as the original sample size k increases, a fixed n represents a smaller fraction of observations in the sample. So for instance, $n = 20$ represents 100% of the

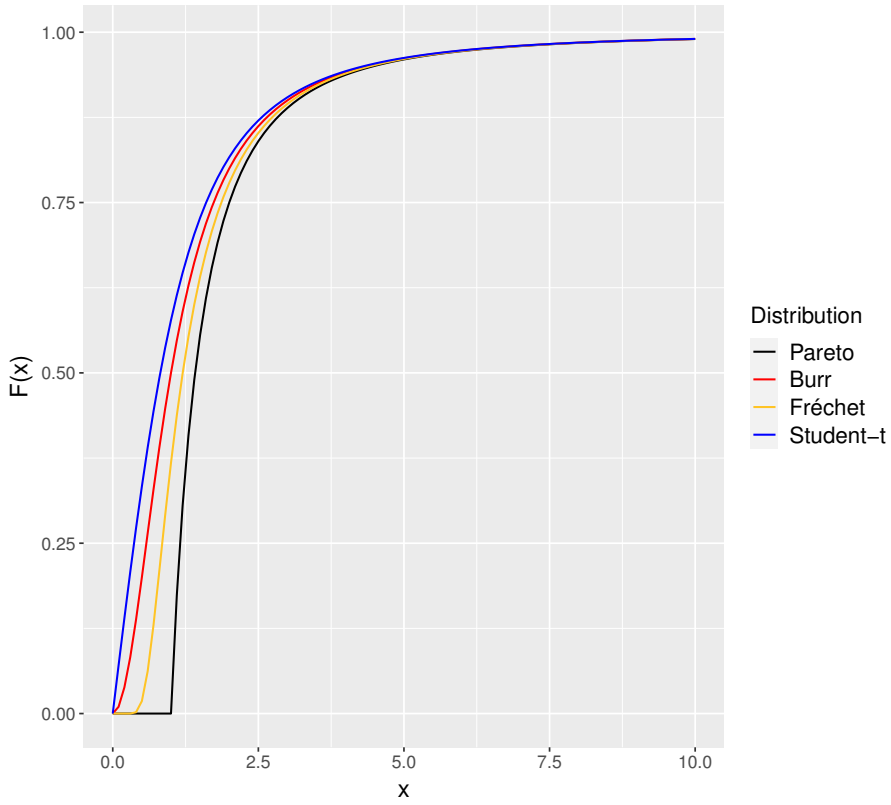

Figure 2.2: Comparison of distributions $F(x) = 1 - x^{-2}$ (Pareto), $F(x) = 1 - (1 + x^2)^{-1}$ (Burr), $F(x) = e^{-x^{-2}}$ (Fréchet), and $F(x) = 2 \int_{-\infty}^{x} \sqrt{2\pi}\Gamma(1.5)(\frac{2}{2+y^2})^{1.5} \, \mathrm{d}y - 1$ (one-sided Student-t).

observations in the sample of size $k = 20$, but only 10% of them in the sample of size $k = 200$. Regarding those plots, we may note that for the Pareto distribution, the expected solvency probability is always equal to $p = 0.995$, which is an application of Theorem 2.5.1, as stated above. Regarding the Burr, Fréchet and Student-t distributions, we would like to make the following comments:

- For a fixed k, the expected solvency probability increases in n: Note that when an element of $\mathbf{X}$ is small (e.g. smaller than 3), it comes from a part of the distribution $F(x)$ that does not behave like the corresponding Pareto distribution, and this biases the capital estimator. For a fixed k, as n increases, more of the greatest order statistics of $\tilde{\mathbf{X}}$ are considered, which are thus more likely to take small values. That results in a larger bias, and hence in a larger deviation of the expected solvency probability from p.

- For a fixed n, the expected solvency probability decreases in k: The explanation here is similar to that of the previous point. That is, as k increases, $\mathbf{X}$ turns out to be the n greatest order statistics of a larger sample. Therefore as k increases, the values in $\mathbf{X}$ are more likely to be large, and hence to come from the part of $F(x)$ that behaves like the Pareto distribution. This leads to a smaller bias of the capital

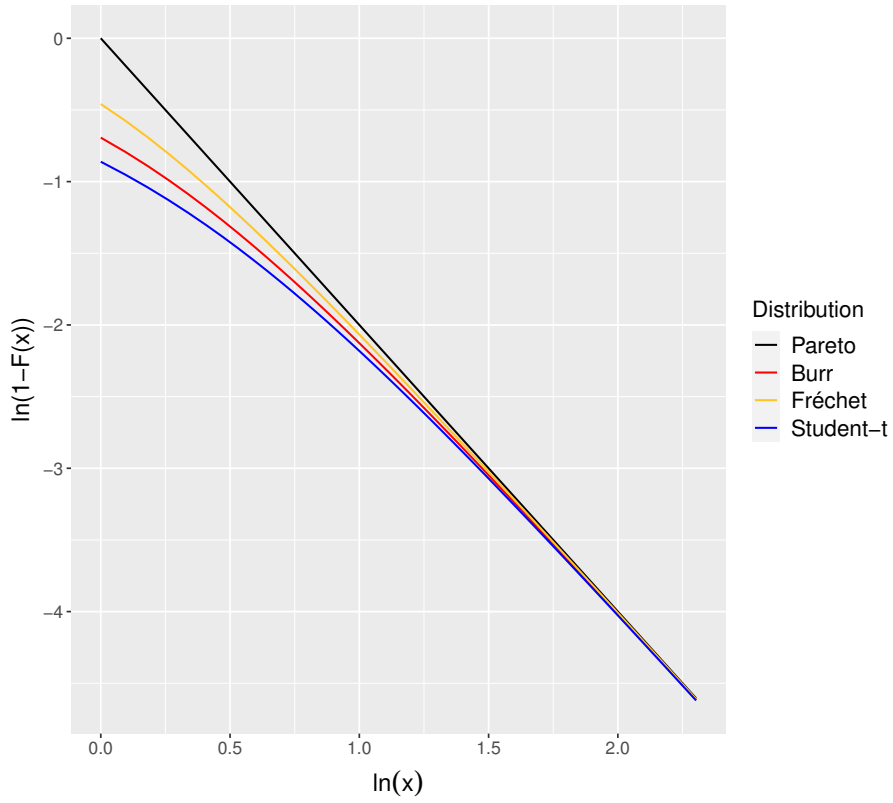

Figure 2.3: Comparison of the log-log plots of distributions $F(x) = 1 - x^{-2}$ (Pareto), $F(x) = 1 - (1 + x^2)^{-1}$ (Burr), $F(x) = e^{-x^{-2}}$ (Fréchet), and $F(x) = 2\int_{-\infty}^{x} \sqrt{2\pi}\Gamma(1.5)(\frac{2}{2+y^2})^{1.5}\,\mathrm{d}y - 1$ (one-sided Student-t).

estimator, and hence to a smaller deviation of the expected solvency probability from p.

- In each case, the expected solvency probability for small n (e.g. $n = 2$) almost equals p, even for the very small sample size $k = 20$. This is due to the fact that, for small n, the elements in $\mathbf{X}$ are likely taking large enough values, where $F(x)$ is already very close to the Pareto distribution. As already commented, this results in a small bias of the capital estimator, and hence the resulting expected solvency probability is close to p. Considering this, it turns out that the trade-off between bias and variance which frequently occurs in statistics, is here once again present: On the one hand, by selecting a small n one keeps the bias small, which ensures the expected solvency probability to be close to p. But on the other hand, the smaller n, the larger the sensitivity of $\tilde{\delta}_p(\mathbf{x})$ with respect to the data $\mathbf{x}$, and hence the greater variability of $\tilde{\delta}_p(\mathbf{X})$.

- In each case, the expected solvency probability is greater than p, which can be understood as follows. In Figure 2.2, we can see that the Burr, Fréchet and one-sided Student-t distributions are each above the Pareto distribution, and the difference is important for small x. Those distributions then converge rather quickly to the Pareto

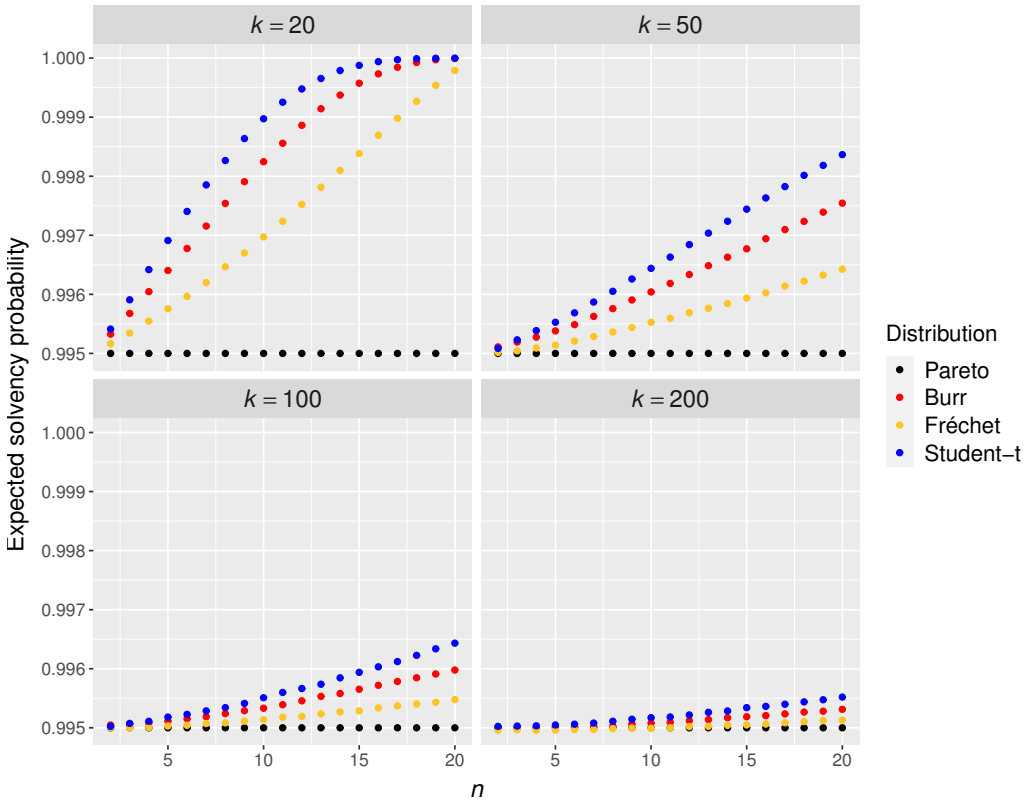

Figure 2.4: Expected solvency probability $\mathbb{P}(Y \leq \tilde{\delta}_p(\mathbf{X}))$ under capital estimator (2.29) of the Pareto distribution, as a function of n, for original sample sizes $k \in \{20, 50, 100, 200\}$, target solvency probability $p = 0.995$, and under potential model misspecification.

distribution, which makes them being flatter than the Pareto distribution, and this is reflected in the behaviour of each X_i in $\mathbf{X}$. For the Pareto case, a flatter distribution means a heavier tail, and hence since the estimator $\tilde{\delta}_p(\cdot)$ considered here is designed for the Pareto distribution, it interprets the behaviour of each X_i as the signal of a heavier tail. Therefore, the resulting capital ends up being larger than it should be, resulting in an expected solvency probability greater than p.

In conclusion, this example illustrates the relatively good robustness potential of the capital estimator (2.29) against model misspecification (in particular when the sample size is not too small), when the model is unknown but has a Pareto-type tail.

2.8 Conclusion

In this paper, we considered the problem of capital calculation under parameter and model uncertainty, when the risk under consideration can be described by a transformed location-scale family. After introducing the notion of the expected solvency probability, we proved that, when only parameter uncertainty is involved, a special type of predictive distribution

can be built, which allows to calculate a capital that yields a target expected solvency probability, despite parameter uncertainty. Also, we provided useful tools that can help the calculation of such predictive distributions. Afterwards, for the case where the random loss follows a Pareto distribution, we used those tools to calculate the explicit formula for the predictive distribution and the resulting capital estimator, which can still serve in case the data consists of consecutive order statistics, rather than a complete random sample. Finally, we tested and showed the robustness of the latter capital estimator against potential model misspecification, for distributions with a Pareto-type tail.
Interesting extensions of this work would include to consider other risk measures, such as the expected-shortfall, and other types of distribution families, typically with more parameters. Also, note that in this chapter we considered only the objective prior (2.22). It can be worthwhile to consider other types of priors, with resulting capital estimators $\widetilde{\delta}_p(\cdot)$ that do not yield exactly the targeted expected solvency probability, but which could e.g. be more robust against model misspecification in some relevant situations. Finally, it may be interesting to consider a variant of the approach of Section 2.7, in which the number n of greatest order statistics entering into the capital estimator may vary within each simulated scenario, to avoid considering observations falling too far from the tail. However, note that under such an approach n becomes a random variable, and this requires a careful treatment.

Acknowledgements. I would like to thank Hansjörg Albrecher for his valuable inputs, which have helped me substantially improve this chapter. I would like to also thank Andreas Tsanakas for his valuable and insightful comments.

2.9 Appendix I: Discussion about Lemma 2.5.1

In Section 2.5.2, Lemma 2.5.1 proves and outlines a key property of the prior (2.22), which then serves in the proof of Theorem 2.5.1. This lemma turns out to also provide an interesting explanation of the relationship between fiducial inference and Bayesian inference, which we will discuss now.

In short, fiducial inference could be described as a set of methods, allowing to perform statistical inference about an unknown parameter (or a random variable, or a vector depending on it), but which (apparently) does not require to formulate a prior, contrarily to the Bayesian approach. Fiducial inference was coined by Fisher, as an attempt to provide a prior-free alternative to Bayesian inference. Fisher developed his method through a series of examples, see for instance [36], [37], [38], and [39]. At that time, most statisticians rejected fiducial inference, because it was viewed as suffering from several inconsistencies, to which Fisher did not provide satisfying answers (see Zabell [82] for a detailed and interesting description of that history). After Fisher, several authors proposed and developed a variety of extensions (for a recent overview, we refer to Dawid [27]), and recent and well noted contributions include e.g. Hannig [51] and Hannig et al. [52].

In the setting of this article, a fiducial inference approach would be as follows (in order

to distinguish it from the notation of the previous sections, we write here the true but unknown value of the parameter as $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*)$): Consider a statistic $\mathbf{T} = t(\mathbf{X}) = (T_1, T_2)$, as described in Section 2.3.3, so that it satisfies

$$\mathbf{T}(\boldsymbol{\theta}^*) = (\theta_1^* + \theta_2^* W_1, \theta_2^* W_2), \tag{2.30}$$

On observing $\mathbf{x}$, we learn that $\mathbf{T}$ realized to $\mathbf{t} = t(\mathbf{x}) = (t_1, t_2)$. Therefore, letting $\mathbf{w} = (w_1, w_2)$ be the realization of $\mathbf{W} = (W_1, W_2)$, we can update Equation (2.30) to

$$\mathbf{t} = (\theta_1^* + \theta_2^* w_1, \theta_2^* w_2).$$

Yet, we observe $\mathbf{x}$ (and hence t_1 and t_2), not w_1 and w_2. If we knew $\boldsymbol{\theta}^*$, we could deduce the two latter quantities, as $\mathbf{w} = ((t_1 - \theta_1^*)/\theta_2^*, t_2/\theta_2^*)$. But since we do not know $\boldsymbol{\theta}^*$, we may assume that, in consequence, we did not learn anything about $\mathbf{W}$ from $\mathbf{t}$. This assumption is the key point on which most fiducial inference methods rely, and it is usually referred to as the fiducial argument. Correspondingly, when updating $\mathbf{T}$ to $\mathbf{t}$ in Equation (2.30), according to the fiducial argument, we may keep $\mathbf{W}$ random (and let its distribution remain unchanged), which results in $\mathbf{t} = (\theta_1^* + \theta_2^* W_1, \theta_2^* W_2)$, and hence

$$\boldsymbol{\theta}^* = (t_1 - t_2 W_1/W_2, t_2/W_2). \tag{2.31}$$

The fiducial argument thus leads to viewing $\boldsymbol{\theta}^*$ as a random vector, and a distribution for $\boldsymbol{\theta}^*$ can then be obtained from $\mathbf{t}$ and $\mathbf{W}$. That distribution is referred to as a fiducial distribution for $\boldsymbol{\theta}^*$. The random vector (2.31) can then be plugged into $Y(\boldsymbol{\theta}^*)$, leading to

$$Y(\boldsymbol{\theta}^*) = h(t_1 + t_2 B), \tag{2.32}$$

where $B = \frac{V - W_1}{W_2}$ is defined as in the previous sections. As for $\boldsymbol{\theta}^*$, a distribution can then be calculated for $Y(\boldsymbol{\theta}^*)$, but from $\mathbf{t}$ and B. This distribution is called a fiducial distribution for $Y(\boldsymbol{\theta}^*)$.

In this example, it seems that indeed, the fiducial distributions of both $\boldsymbol{\theta}^*$ and $Y(\boldsymbol{\theta}^*)$ have been obtained without formulating a prior, as claimed by advocates of fiducial inference. Moreover, it can be verified that the fiducial distribution for $\boldsymbol{\theta}^*$ induced by Equation (2.31) corresponds to a Bayesian posterior $f_{\boldsymbol{\Theta}|\mathbf{T}}(\cdot|\mathbf{t})$ calculated under the objective prior (2.22). Likewise, it is easy to see from Proposition 2.5.2, that the fiducial distribution for $Y(\boldsymbol{\theta}^*)$ induced by Equation (2.32) corresponds to a predictive distribution stemming from the prior (2.22). While this equivalence between the fiducial and objective Bayesian approaches is well-known (see for instance Section 7 in Fraser [42]), to the best of our knowledge, it has only been acknowledged, and never explained.

It turns out that Lemma 2.5.1 provides the explanation, which is as follows: According to Lemma 2.5.1, the prior (2.22) has the property to make $\mathbf{U}$ and $\mathbf{T}$ being independent, and since W_1 and W_2 are both functions of $\mathbf{U}$ only, it also makes $\mathbf{W}$ and $\mathbf{T}$ being independent. In fact, an inspection of the proof of Lemma 2.5.1, in particular Equation (2.33), reveals that for $\mathbf{U}$ (or $\mathbf{W}$) and $\mathbf{T}$ being independent in the Bayesian setting, the prior must satisfy Equation (2.22). As a result, raising the fiducial argument, i.e. assuming the independence of $\mathbf{W}$ from $\mathbf{T}$, simply appears to be an alternative way to uniquely selecting the

prior (2.33).

We think that this connection can be used to purge fiducial inference from its apparent inconsistencies and paradoxes. Also, and most importantly, we believe that the fiducial argument, interpreted in a Bayesian framework, provides a very intuitive and promising new application of the indifference principle. But this will be the topic of further research.

2.10 Appendix II: Proofs

2.10.1 Proof of Lemma 2.5.1

By the definition of independence, for proving Lemma 2.5.1 we may simply show that

$$\frac{f_{\mathbf{T}|\mathbf{U}}(\mathbf{t}|\mathbf{u})}{f_{\mathbf{T}}(\mathbf{t})} = 1, \text{ for all } \mathbf{t} \in \mathcal{T} \text{ and } \mathbf{u} \in \mathcal{U},$$

which we do hereafter.

Letting $w_1 = t_1(g(\mathbf{u}))$ and $w_2 = t_2(g(\mathbf{u}))$, according to Equation (2.23) we can write

$$(\mathbf{T}|\mathbf{U} = \mathbf{u}) = (\Theta_1 + \Theta_2 w_1, \Theta_2 w_2 | \mathbf{U} = \mathbf{u}), \text{ for all } \mathbf{u} \in \mathcal{U},$$

which, since $\mathbf{\Theta}$ is an independent random vector, simplifies to

$$(\mathbf{T}|\mathbf{U} = \mathbf{u}) = (\Theta_1 + \Theta_2 w_1, \Theta_2 w_2), \text{ for all } \mathbf{u} \in \mathcal{U}.$$

This yields

$$f_{\mathbf{T}|\mathbf{U}}(\mathbf{t}|\mathbf{u}) = \frac{\mathrm{d}}{\mathrm{d}t_1}\frac{\mathrm{d}}{\mathrm{d}t_2}\mathbb{P}(\Theta_1 + \Theta_2 w_1 \leq t_1, \Theta_2 w_2 \leq t_2), \text{ for all } \mathbf{t} \in \mathcal{T} \text{ and } \mathbf{u} \in \mathcal{U}.$$

From Assumption (A3), we have $w_1 = t_1(g(\mathbf{u})) \in \mathbb{R}$ and, more importantly, $w_2 = t_2(g(\mathbf{u})) > 0$, for all $\mathbf{u} \in \mathcal{U}$, which in the above equation leads to

$$\begin{aligned} f_{\mathbf{T}|\mathbf{U}}(\mathbf{t}|\mathbf{u}) &= \frac{\mathrm{d}}{\mathrm{d}t_1}\frac{\mathrm{d}}{\mathrm{d}t_2}\int_0^{\frac{t_2}{w_2}}\int_{-\infty}^{t_1-\theta_2 w_1} f_{\mathbf{\Theta}}(\theta_1,\theta_2)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2 \\ &= \frac{1}{w_2} f_{\mathbf{\Theta}}\left(t_1 - t_2\frac{w_1}{w_2}, \frac{t_2}{w_2}\right), \text{ for all } \mathbf{t} \in \mathcal{T} \text{ and } \mathbf{u} \in \mathcal{U}, \end{aligned} \tag{2.33}$$

and hence

$$\frac{f_{\mathbf{T}|\mathbf{U}}(\mathbf{t}|\mathbf{u})}{f_{\mathbf{T}}(\mathbf{t})} = \frac{\frac{1}{w_2} f_{\mathbf{\Theta}}(t_1 - t_2\frac{w_1}{w_2}, \frac{t_2}{w_2})}{\int_{\mathbf{u}'\in\mathcal{U}} \frac{1}{w_2'} f_{\mathbf{\Theta}}(t_1 - t_2\frac{w_1'}{w_2'}, \frac{t_2}{w_2'}) f_{\mathbf{U}}(\mathbf{u}')\,\mathrm{d}\mathbf{u}'}, \text{ for all } \mathbf{t} \in \mathcal{T} \text{ and } \mathbf{u} \in \mathcal{U},$$

where $w_1' = t_1(g(\mathbf{u}')) \in \mathbb{R}$ and $w_2' = t_2(g(\mathbf{u}')) > 0$. From Equation (2.8), we have $t_1 \in \mathbb{R}$ and $t_2 > 0$, for all $\mathbf{t} \in \mathcal{T}$, and hence since $w_1, w_1' \in \mathbb{R}$ and $w_2, w_2' > 0$, for all $\mathbf{u}, \mathbf{u}' \in \mathcal{U}$, this yields $(t_1 - t_2\frac{w_1}{w_2}, \frac{t_2}{w_2}) \in \mathcal{P}$ and $(t_1 - t_2\frac{w_1'}{w_2'}, \frac{t_2}{w_2'}) \in \mathcal{P}$, so that under the prior (2.22), the latter equation becomes

$$\frac{f_{\mathbf{T}|\mathbf{U}}(\mathbf{t}|\mathbf{u})}{f_{\mathbf{T}}(\mathbf{t})} = \frac{1}{\int_{\mathbf{u}'\in\mathcal{U}} f_{\mathbf{U}}(\mathbf{u}')\,\mathrm{d}\mathbf{u}'} = 1, \text{ for all } \mathbf{t} \in \mathcal{T} \text{ and } \mathbf{u} \in \mathcal{U},$$

which completes the proof.

2.10.2 Proof of Theorem 2.5.1

Let $\mathbf{x} \in \mathcal{X}$, $\mathbf{a} = a(\mathbf{x})$ and $\mathbf{t} = t(\mathbf{x}) = (t_1, t_2)$, where $t_1 = t_1(\mathbf{x})$ and $t_2 = t_2(\mathbf{x})$. Moreover, note that from Equation (2.8) we have $t_2 > 0$ necessarily, and by definition the function $h(\cdot)$ is strictly increasing. As a result, Equations (2.11) and (2.12) lead to

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{Y|\mathbf{A},\mathbf{T}}(\cdot|\mathbf{a},\mathbf{t}) = F_{B|\mathbf{A},\mathbf{T}}\bigg(\frac{h^{-1}(\cdot) - t_1}{t_2} \;\Big|\; \mathbf{a},\mathbf{t}\bigg).$$

Then, on the one hand, Assumption (A1) makes $\mathbf{U} = g^{-1}(\mathbf{X}(\mathbf{e}))$ and $V = h^{-1}(Y(\mathbf{e}))$ independent, so from Equation (2.23), and since $\boldsymbol{\Theta}$ is an independent random vector, we have that V is independent of $\mathbf{T}$. On the other hand, from Lemma 2.5.1 we know that under the prior (2.22), $\mathbf{U}$ is independent of $\mathbf{T}$. Therefore, since according to Equations (2.9) and (2.10), both $\mathbf{A}$ and B can be expressed as functions of $\mathbf{U}$ and V only, we deduce that prior (2.22) makes $\mathbf{A}$ and B being both independent of $\mathbf{T}$. In the above equation, this leads to

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{B|\mathbf{A}}\bigg(\frac{h^{-1}(\cdot) - t_1}{t_2} \;\Big|\; \mathbf{a}\bigg),$$

and hence

$$\tilde{\delta}_p(\mathbf{x}) = \text{VaR}_p[Y|\mathbf{X} = \mathbf{x}] = h(t_1 + t_2\text{VaR}_p[B|\mathbf{A} = \mathbf{a}]),$$

for any $p \in [0,1]$. Using this, together with Equation (2.12), the resulting expected solvency probability can be expressed as

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\delta}_p(\mathbf{X})) &= \mathbb{P}_{\boldsymbol{\theta}}(h(T_1 + T_2 B) \leq h(T_1 + T_2\text{VaR}_p[B|\mathbf{A}])) \\
&= \mathbb{P}_{\boldsymbol{\theta}}(B \leq \text{VaR}_p[B|\mathbf{A}]) \\
&= \int_{\mathbf{a}\in\mathbb{R}^n} \mathbb{P}_{\boldsymbol{\theta}}(B \leq \text{VaR}_p[B|\mathbf{A}]|\mathbf{A} = \mathbf{a}) f_{\mathbf{A}}(\mathbf{a})\,\mathrm{d}\mathbf{a} \\
&= \int_{\mathbf{a}\in\mathbb{R}^n} \mathbb{P}_{\boldsymbol{\theta}}(B \leq \text{VaR}_p[B|\mathbf{A} = \mathbf{a}]|\mathbf{A} = \mathbf{a}) f_{\mathbf{A}}(\mathbf{a})\,\mathrm{d}\mathbf{a}, \text{ for all } \boldsymbol{\theta} \in \mathcal{P}.
\end{aligned}$$

By the definition of the value-at-risk, and since B is a continuous random variable, we have $\mathbb{P}_{\boldsymbol{\theta}}(B \leq \text{VaR}_p[B|\mathbf{A} = \mathbf{a}]|\mathbf{A} = \mathbf{a}) = p$, for all $\boldsymbol{\theta} \in \mathcal{P}$, which results in

$$\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\delta}_p(\mathbf{X})) = \int_{\mathbf{a}\in\mathbb{R}^n} p f_{\mathbf{A}}(\mathbf{a})\,\mathrm{d}\mathbf{a} = p, \text{ for all } \boldsymbol{\theta} \in \mathcal{P},$$

and hence Theorem 2.5.1 indeed holds.

2.10.3 Proof of Lemma 2.5.2

By Definition 2.2.1, if $\mathbf{S} = s(\mathbf{X})$ is a sufficient statistic of $\mathbf{X}$, then we have

$$f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta}) = \gamma_1(s(\mathbf{x});\boldsymbol{\theta})\gamma_2(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P},$$

where $\gamma_2(\mathbf{x})$ does not depend on $\boldsymbol{\theta}$. In Formula (2.19), this yields

$$f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) = \frac{\gamma_1(s(\mathbf{x});\boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}'\in\mathcal{P}} \gamma_1(s(\mathbf{x});\boldsymbol{\theta}') f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'}, \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}.$$

Also, note that since $\mathbf{S}$ is a function of $\mathbf{X}$ only, we can write

$$f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) = f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \tag{2.34}$$

which results in

$$f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})) = \frac{\gamma_1(s(\mathbf{x}); \boldsymbol{\theta}) f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}' \in \mathcal{P}} \gamma_1(s(\mathbf{x}); \boldsymbol{\theta}') f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}') \, \mathrm{d}\boldsymbol{\theta}'}, \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}. \tag{2.35}$$

Now consider the posterior $f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{x}))$. Using the law of total probability, the latter can be written as

$$f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{x})) = \int_{\mathbf{x}' \in \mathcal{X}} f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}', s(\mathbf{x})) \, \mathrm{d}F_{\mathbf{X}|\mathbf{S}}(\mathbf{x}'|s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}.$$

On the one hand, since $\mathbf{S} = s(\mathbf{X})$, the random vector $(\mathbf{X}|\mathbf{S} = s(\mathbf{x}))$ obviously cannot take a value $\mathbf{x}'$ outside the set $\mathcal{X}_{s(\mathbf{x})} = \{\mathbf{x}' \in \mathcal{X} : s(\mathbf{x}') = s(\mathbf{x})\}$, meaning that

$$\mathrm{d}F_{\mathbf{X}|\mathbf{S}}(\mathbf{x}'|s(\mathbf{x})) = 0, \text{ for all } \mathbf{x}' \notin \mathcal{X}_{s(\mathbf{x})}, \tag{2.36}$$

and hence

$$f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{x})) = \int_{\mathbf{x}' \in \mathcal{X}_{s(\mathbf{x})}} f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}', s(\mathbf{x})) \, \mathrm{d}F_{\mathbf{X}|\mathbf{S}}(\mathbf{x}'|s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}. \tag{2.37}$$

On the other hand, according to Equation (2.35), for any $\mathbf{x}' \in \mathcal{X}_{s(\mathbf{x})}$, we have

$$f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}', s(\mathbf{x})) = f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P},$$

which in Equation (2.37) leads to

$$\begin{aligned} f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{s})) &= \int_{\mathbf{x}' \in \mathcal{X}_{s(\mathbf{x})}} f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})) \, \mathrm{d}F_{\mathbf{X}|\mathbf{S}}(\mathbf{x}'|s(\mathbf{x})) \\ &= f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})) \int_{\mathbf{x}' \in \mathcal{X}_{s(\mathbf{x})}} \mathrm{d}F_{\mathbf{X}|\mathbf{S}}(\mathbf{x}'|s(\mathbf{x})) \\ &= f_{\boldsymbol{\Theta}|\mathbf{X},\mathbf{S}}(\boldsymbol{\theta}|\mathbf{x}, s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}, \end{aligned} \tag{2.38}$$

where the last line stems from Equation (2.36). By assembling Equations (2.34) and (2.38), we obtain

$$f_{\boldsymbol{\Theta}|\mathbf{X}}(\boldsymbol{\theta}|\mathbf{x}) = f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X} \text{ and } \boldsymbol{\theta} \in \mathcal{P}.$$

Finally, using the above result in Formula (2.20) leads to

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = \int_{\boldsymbol{\theta} \in \mathcal{P}} F_Y(\cdot; \boldsymbol{\theta}) f_{\boldsymbol{\Theta}|\mathbf{S}}(\boldsymbol{\theta}|s(\mathbf{x})) \, \mathrm{d}\boldsymbol{\theta} = F_{Y|\mathbf{S}}(\cdot|s(\mathbf{x})), \text{ for all } \mathbf{x} \in \mathcal{X},$$

and the proof is complete.

2.10.4 Proof of Proposition 2.5.1

Since $\mathbf{T}$ is a sufficient statistic, from Lemma 2.5.2 we have

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{Y|\mathbf{T}}(\cdot|\mathbf{t}),$$

which from Proposition 2.5.2 results in

$$F_{Y|\mathbf{X}}(\cdot|\mathbf{x}) = F_{Y|\mathbf{T}}(\cdot|\mathbf{t}) = F_B\left(\frac{h^{-1}(\cdot) - t_1}{t_2}\right),$$

and hence

$$\tilde{\delta}_p(\mathbf{x}) = \text{VaR}_p[Y|\mathbf{X} = \mathbf{x}] = h(t_1 + t_2\text{VaR}_p[B]),$$

for any $p \in [0, 1]$, so the proof is complete.

2.10.5 Proof of Proposition 2.5.2

From Equation (2.12), we have

$$F_{Y|\mathbf{T}}(\cdot|\mathbf{t}) = F_{B|\mathbf{T}}\left(\frac{h^{-1}(\cdot) - t_1}{t_2} \;\middle|\; \mathbf{t}\right).$$

Using the same arguments as in the proof of Theorem 2.5.1, Lemma 2.5.1 leads to B being independent of $\mathbf{T}$, hence

$$F_{Y|\mathbf{T}}(\cdot|\mathbf{t}) = F_B\left(\frac{h^{-1}(\cdot) - t_1}{t_2}\right),$$

and

$$\tilde{\eta}_p(\mathbf{t}) = \text{VaR}_p[Y|\mathbf{T} = \mathbf{t}] = h(t_1 + t_2\text{VaR}_p[B]),$$

for any $p \in [0, 1]$. Then again, proceeding as in the proof of Theorem 2.5.1, we obtain

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\theta}}(Y \leq \tilde{\eta}_p(\mathbf{T})) &= \mathbb{P}_{\boldsymbol{\theta}}(h(T_1 + T_2 B) \leq h(T_1 + T_2\text{VaR}_p[B])) \\ &= \mathbb{P}_{\boldsymbol{\theta}}(B \leq \text{VaR}_p[B]) \\ &= p, \text{ for all } \boldsymbol{\theta} \in \mathcal{P},\end{aligned}$$

which completes the proof.

Chapter 3

A note on risk assessment under parameter uncertainty

This chapter is based on L. Vincent (2022): *A note on risk assessment under parameter uncertainty*. Preprint, University of Lausanne [78].

Abstract. In the risk management literature, it has been shown several times that (Bayesian) predictive distributions can be constructed so as to account for parameter uncertainty in a relevant way. Those contributions however focused on risk assessment problems involving a single risk. In this note, we show that predictive distributions may be relevant for resource allocation problems also, when several risks are involved and parameter uncertainty is present for each of them.

3.1 Introduction

3.1.1 Estimated and predictive distributions

This note considers the problem of risk assessment under parameter uncertainty. The setting is as follows. Let X be a continuous random variable representing a risk. The distribution of X depends on some parameter θ (possibly a vector) and is thus denoted by $F(x;\theta) = \mathbb{P}_\theta(X \leq x)$. The function $F(x;\theta)$ is known, but the value of θ is not. Rather, the realization $\mathbf{x} = (x_1, ..., x_n)$ of a random vector $\mathbf{X} = (X_1, ..., X_n)$ is observed, where $X_1, ..., X_n$ denote $n > 1$ mutually independent occurrences of X. The resulting likelihood function is $\ell(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i;\theta)$, where $f(x;\theta) = \frac{\mathrm{d}}{\mathrm{d}x}F(x;\theta)$ denotes the probability density function of X.

In this setting, assume that one needs to come up with a distribution for X, so as to inform some risk management decision process. The first option would often be to calculate the maximum likelihood estimate $\hat{\theta}(\mathbf{x}) = \arg\max_\theta \ \ell(\theta|\mathbf{x})$ of the unknown parameter, and then use it to estimate the distribution of X simply as

$$\hat{F}(x|\mathbf{x}) = F(x;\hat{\theta}(\mathbf{x})).$$

By definition, $\hat{F}(x|\mathbf{x})$ is the distribution which is most likely to have generated the data $\mathbf{x}$, and one may reasonably rely on it for describing the risk X. However, by doing so one implicitly treats $\hat{\theta}(\mathbf{x})$ as the true parameter value, which boils down to ignore parameter uncertainty. As shown for instance by Gerrard and Tsanakas [47], Fröhlich and Weng [43] and Pitera and Schmidt [70], this generally leads to an underestimation of risk, in particular when the sample size n is small.

In order to account for parameter uncertainty, an alternative option is to calculate a predictive distribution. To do so, one must first select a prior $\pi(\theta)$, which is meant to represent one's uncertainty about the true parameter value before considering the data $\mathbf{x}$. The data can then be used to update the prior according to Bayes' formula, leading to the posterior $\pi(\theta|\mathbf{x}) \propto \ell(\theta|\mathbf{x})\pi(\theta)$. The predictive distribution is finally obtained by averaging all possible distributions with that posterior $\pi(\theta|\mathbf{x})$, as

$$\tilde{F}(x|\mathbf{x}) = \int F(x;\theta)\pi(\theta|\mathbf{x})\,\mathrm{d}\theta.$$

The predictive distribution $\tilde{F}(x|\mathbf{x})$ is thus an average distribution, which by construction accounts for parameter uncertainty, through $\pi(\theta|\mathbf{x})$. Also, it generally differs from the above estimated distribution $\hat{F}(x|\mathbf{x})$, just as the mean of a random variable typically differ from its mode.

Of course, the relevance of a predictive distribution depends on the particular prior it stems from, and the selection of the latter is thus a question of interest. Once a prior is selected, it is natural to compare the resulting predictive distribution with the above estimated distribution, so as to determine the concrete consequences of taking into account parameter uncertainty. For example, say that one wishes to determine the amount of resources (capital) required for backing X, in order to have a survival (solvency) probability of $p \in (0,1)$. This quantity is by definition the value-at-risk (or quantile) at level p, which in principle can be calculated as $\delta_p = \inf\{x \in \mathbb{R} : F(x;\theta) \geq p\}$. Yet, since θ is unknown, so is δ_p, and hence the value-at-risk must be estimated. Consider the estimates $\hat{\delta}_p(\mathbf{x}) = \inf\{x \in \mathbb{R} : \hat{F}(x|\mathbf{x}) \geq p\}$ and $\tilde{\delta}_p(\mathbf{x}) = \inf\{x \in \mathbb{R} : \tilde{F}(x|\mathbf{x}) \geq p\}$. In order to assess their respective reliability, one may study the properties of the related estimators $\hat{\delta}_p(\mathbf{X})$ and $\tilde{\delta}_p(\mathbf{X})$. For instance, these estimators yield the expected survival probabilities $\mathbb{E}[F(\hat{\delta}_p(\mathbf{X});\theta)]$ and $\mathbb{E}[F(\tilde{\delta}_p(\mathbf{X});\theta)]$, and Gerrard and Tsanakas [47] showed that for a large class of distributions $F(x;\theta)$ that are frequently used in practice (those who belong to a so-called transformed location-scale distribution family, which for instance include the log-Normal, Pareto and Weibull distributions):

- On the one hand, if $n < \infty$, then except in very rare cases, we have $\mathbb{E}[F(\hat{\delta}_p(\mathbf{X});\theta)] \neq p$, for all θ, and with $\mathbb{E}[F(\hat{\delta}_p(\mathbf{X});\theta)] < p$ in general when p is close to 1. That is, in relevant cases the value-at-risk calculated from the estimated distribution $\hat{F}(x|\mathbf{x})$ generally leads to an expected survival probability that is smaller than the targeted one.

- On the other hand, a non-informative prior can be selected such that $\mathbb{E}[F(\tilde{\delta}_p(\mathbf{X});\theta)]$

$= p$, for all θ and n, so the predictive distribution $\hat{F}(x|\mathbf{x})$ allows to reach the target solvency probability on average, despite not knowing the parameter θ, and regardless of the sample size.

This example illustrates the underestimation of risk that may arise from ignoring parameter uncertainty, and the ability of predictive distributions to take it into account in a relevant way. Other studies of this kind include for instance Bignozzi and Tsanakas [13], where similar results were obtained, but for other risk measures, and Bignozzi and Tsanakas [12] also considered model uncertainty. Before that, Cairns [19] also showed the relevance of Bayesian approaches to account for parameter and model uncertainty, through several examples related to insurance.

However, note that by accounting for parameter uncertainty, predictive distributions are often leading to an assessment of risk that is fairly conservative, in particular when built from a non-informative prior. For instance, in the above example, if the estimators $\tilde{\delta}_p(\mathbf{X})$ and $\hat{\delta}_p(\mathbf{X})$ indeed yield $\mathbb{E}[F(\tilde{\delta}_p(\mathbf{X}); \theta)] = p$, for all θ, and $\mathbb{E}[F(\hat{\delta}_p(\mathbf{X}); \theta)] < p$, then $\tilde{\delta}_p(\mathbf{X})$ will take larger values than $\hat{\delta}_p(\mathbf{X})$ in general, and the difference can be quite substantial. This conservative aspect of predictive distributions may be viewed as a drawback: At the end of the risk assessment process, limited resources (e.g. capital) are available, and by being too conservative one may not be effective. Also, we mention that while predictive distributions allow to control survival (solvency) probabilities in a relevant way, they may induce some potentially undesirable bias, e.g. when used for pricing risks, as pointed out by Landsman and Tsanakas [60].

3.1.2 The content of this note

As we just commented, the ability of predictive distributions to address the problem of parameter uncertainty in a relevant way has been demonstrated several times in the risk management literature. Yet, in those contributions, a single risk was considered, with no constraint on the available resources, which may differ from several practical situations. For instance, when endowed with a limited budget, a government department responsible for the maintenance of public infrastructures has to compare their respective failure risk, for prioritizing the most threatening ones. An insurance or banking company that has to determine some capital amount for each of its business lines will do so by considering their relative risk, for then allocating them a fixed overall capital.

Here, we would like to illustrate that predictive distributions can also be relevant in the above type of situations, i.e. when limited resources must be allocated to several competing risks, and parameter uncertainty is involved for each of them. The motivation for this is two-fold: Firstly, since the magnitude of the parameter uncertainty relative to each risk may be unequal (e.g. as the result of different sample sizes), considering it explicitly through a predictive distribution allows for a better assessment of each individual risk in principle, and hence it leads to potentially better resource allocation decisions. Secondly, when limited resources are available, if each risk is modelled with a predictive distribution, then the conservative aspect of the latter type of distributions likely ceases to be an

issue.

Hereafter we do so by means of an example, which illustrates the above ideas in a simple way.

3.2 Illustration

Assume that we face two risks, say X and Y, which are known to be independent. Also, we know that both X and Y follow a Pareto distribution, with scale parameter 1 and unknown respective shape parameters $\theta > 0$ and $\lambda > 0$, so

$$F_X(x;\theta) = 1 - x^{-\theta},\ F_Y(x;\lambda) = 1 - x^{-\lambda},\ \text{for } x > 1.$$

We need to allocate a limited amount of resources $k > 0$ separately to each of these two risks, and we wish to do so in a way that maximizes our joint survival probability. For instance, X and Y could represent the spendings required for making two different nuclear plants resilient against the maximal earthquake magnitude that will occur in their respective geographical regions, during a given future time period.

Denote by $s \geq 0$ and $k - s \geq 0$ the respective shares of k that we allocate to X and Y. Then our joint survival probability is defined as

$$\psi_k(s|\theta,\lambda) = \mathbb{P}_{\theta,\lambda}(X \leq s, Y \leq k - s) = F_X(s;\theta)F_Y(k - s;\lambda). \tag{3.1}$$

Ideally, we could simply calculate the optimal allocation

$$s_k^*(\theta,\lambda) = \underset{s\in[0,k]}{\arg\max}\ \psi_k(s;\theta,\lambda),$$

and end up with maximal joint survival probability

$$\psi_k^*(\theta,\lambda) = \psi_k(s_k^*(\theta,\lambda);\theta,\lambda).$$

However, we do not know the values of θ and λ. Instead, we observed the realizations $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_m)$ of $\mathbf{X} = (X_1, ..., X_n)$ and $\mathbf{Y} = (Y_1, ..., Y_m)$, where $X_1, ..., X_n$ and $Y_1, ..., Y_m$ are $n > 1$ and $m > 1$ mutually independent occurrences of X and Y. Accordingly, we consider performing the above maximization using either the estimated distributions of X and Y, or their posterior predictive distributions.

The maximum likelihood estimates of θ and λ are easily found to be $\hat{\theta}(\mathbf{x}) = n/\sum_{i=1}^n \ln(x_i)$ and $\hat{\lambda}(\mathbf{y}) = m/\sum_{i=1}^m \ln(y_i)$. This yields estimated distributions

$$\hat{F}_X(x|\mathbf{x}) = 1 - x^{-\hat{\theta}(\mathbf{x})}, \hat{F}_Y(x|\mathbf{y}) = 1 - x^{-\hat{\lambda}(\mathbf{y})},\ \text{for } x > 1, \tag{3.2}$$

and the objective function (3.1) then becomes $\hat{F}_X(s|\mathbf{x})\hat{F}_Y(k - s|\mathbf{y})$. Under the estimated distributions, the allocation is then

$$\hat{s}_k^*(\mathbf{x},\mathbf{y}) = \underset{s\in[0,k]}{\arg\max}\ \hat{F}_X(s|\mathbf{x})\hat{F}_Y(k - s|\mathbf{y}).$$

Regarding the predictive distributions of X and Y, we consider the non-informative prior $\pi(\theta) \propto \mathbb{1}\{\theta > 0\}/\theta$ and $\pi(\lambda) \propto \mathbb{1}\{\lambda > 0\}/\lambda$. After few calculation steps this results in

$$\tilde{F}_X(x|\mathbf{x}) = 1 - \left(\frac{n/\hat{\theta}(\mathbf{x})}{n/\hat{\theta}(\mathbf{x}) + \ln(x)}\right)^n, \tilde{F}_Y(x|\mathbf{y}) = 1 - \left(\frac{m/\hat{\lambda}(\mathbf{y})}{m/\hat{\lambda}(\mathbf{y}) + \ln(x)}\right)^m, \text{ for } x > 1. \tag{3.3}$$

With these distributions the objective function (3.1) becomes $\tilde{F}_X(s|\mathbf{x})\tilde{F}_Y(k-s|\mathbf{y})$, so the allocation resulting from the predictive distributions is

$$\tilde{s}_k^*(\mathbf{x}, \mathbf{y}) = \underset{s\in[0,k]}{\arg\max}\ \tilde{F}_X(s|\mathbf{x})\tilde{F}_Y(k-s|\mathbf{y}).$$

Since the above estimated distributions and posterior predictive distributions differ, the allocations $\hat{s}_k^*(\mathbf{x}, \mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x}, \mathbf{y})$ are likely to differ too. We are thus naturally led to compare the relevance of those two different approaches, and we will do so by considering the resulting joint survival probabilities $\psi_k(\hat{s}_k^*(\mathbf{x}, \mathbf{y}); \theta, \lambda)$ and $\psi_k(\tilde{s}(\mathbf{x}, \mathbf{y}); \theta, \lambda)$, for all possible realizations of $\mathbf{x}$ and $\mathbf{y}$ (note that by letting $\mathbf{x}$ and $\mathbf{y}$ vary, we are changing the distributions $\hat{F}_X(x|\mathbf{x})$ and $\hat{F}_Y(x|\mathbf{y})$ on the one hand, and $\tilde{F}_X(x|\mathbf{x})$ and $\tilde{F}_Y(x|\mathbf{y})$ on the other hand, which can thus be seen as a multivariate stress testing, in the sense of Millossovich et al. [63]). We are thus interested in the behaviour of random variables

$$\hat{\Psi}_k(\theta, \lambda) = \psi_k(\hat{s}_k^*(\mathbf{X}, \mathbf{Y}); \theta, \lambda) \text{ and } \tilde{\Psi}_k(\theta, \lambda) = \psi_k(\tilde{s}_k^*(\mathbf{X}, \mathbf{Y}); \theta, \lambda),$$

described by their respective distributions

$$\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x) \text{ and } \mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x).$$

To that end, we assume the amount of resources to be $k = 20$, and sample sizes $n \in \{5, 10, 30\}$ and $m \in \{5, 10, 30\}$. Regarding the parameters θ and λ, we consider a symmetric case $\theta = \lambda = 3$, and a non-symmetric case $\theta = 3$ and $\lambda = 2$. Note that in the symmetric case X and Y are equally distributed, and under the non-symmetric case the right tail of Y is heavier than that of X.

In Figure 3.1, we plotted distributions $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ under the various sample sizes, for the symmetric case $\theta = \lambda = 3$. Notice that since X and Y are equally distributed here, random variables $\hat{\Psi}_k(\theta, \lambda)$ and $\tilde{\Psi}_k(\theta, \lambda)$ remain equally distributed under a permutation of n and m, by symmetry, and we thus removed the redundant plots. In Figure 3.2, we again plotted distributions $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ under the various sample sizes, but for the non-symmetric case $\theta = 3$ and $\lambda = 2$. In both figures, the dots at the bottom of the plots indicate the expected joint survival probabilities $\mathbb{E}[\hat{\Psi}_k(\theta, \lambda)]$ and $\mathbb{E}[\tilde{\Psi}_k(\theta, \lambda)]$.

In order to help the understanding of these plots, we make the following observations:

- Distributions $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ always have the same right endpoint, which equals the theoretical maximal joint survival probability $\psi_k^*(\theta, \lambda)$.

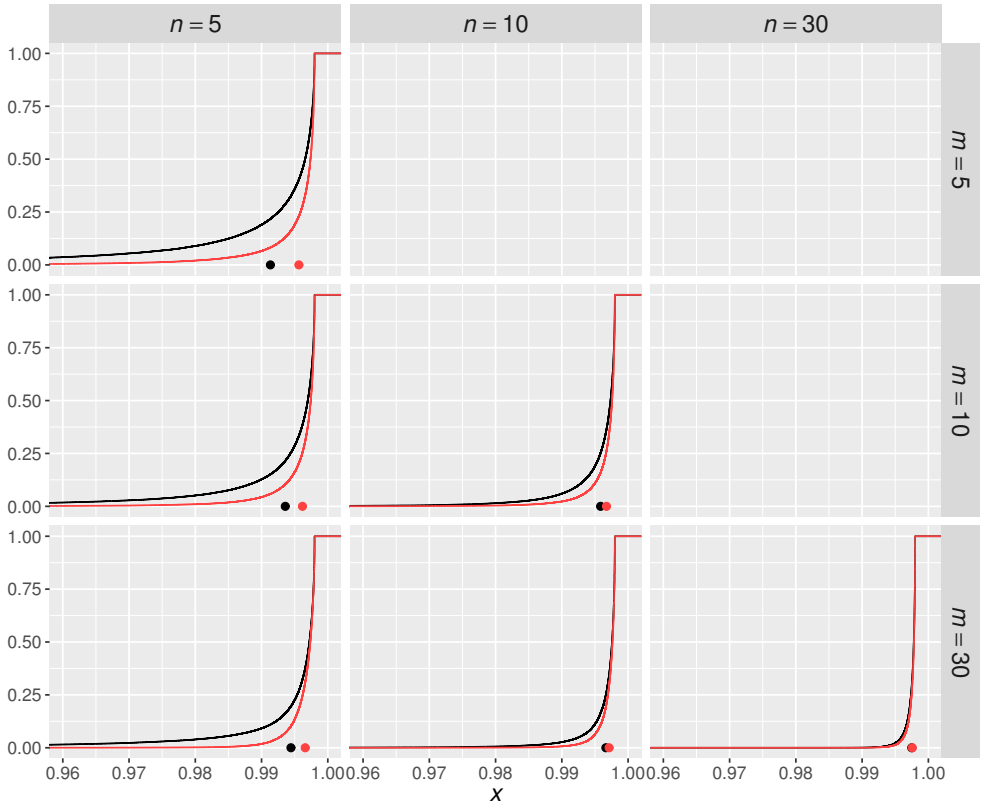

Figure 3.1: Black: $\mathbb{P}(\hat{\Psi}_{20}(3,3) \leq x)$ (curves) and $\mathbb{E}[\hat{\Psi}_{20}(3,3)]$ (dots). Red: $\mathbb{P}(\tilde{\Psi}_{20}(3,3) \leq x)$ (curves) and $\mathbb{E}[\tilde{\Psi}_{20}(3,3)]$ (dots). The sample sizes are $n \in \{5, 10, 30\}$ (columns) and $m \in \{5, 10, 30\}$ (rows).

- The left tail of $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ correspond to realizations $\mathbf{x}$ and $\mathbf{y}$ leading to inaccurate estimated distributions or predictive distributions, and hence to allocations $\hat{s}_k^*(\mathbf{x}, \mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x}, \mathbf{y})$ being far from the theoretical optimal allocation $s_k^*(\theta, \lambda)$.

- Both $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ are concentrated near their right endpoint, meaning that in most cases the allocations $\hat{s}_k^*(\mathbf{x}, \mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x}, \mathbf{y})$ are both close to their theoretical optimal value. From that point of view, the estimated and predictive distributions can be considered relatively accurate in general.

- The larger the sample sizes n and m, the less parameter uncertainty, so the more accurate the estimated and predictive distributions are. As a result, for larger n and m the allocations $\hat{s}_k^*(\mathbf{x}, \mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x}, \mathbf{y})$ are closer to their theoretical optimal value, and hence $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ are more concentrated near their right endpoint.

Regarding the comparison of $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$, the key points are:

- $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ is always under $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$, so compared to the allocation

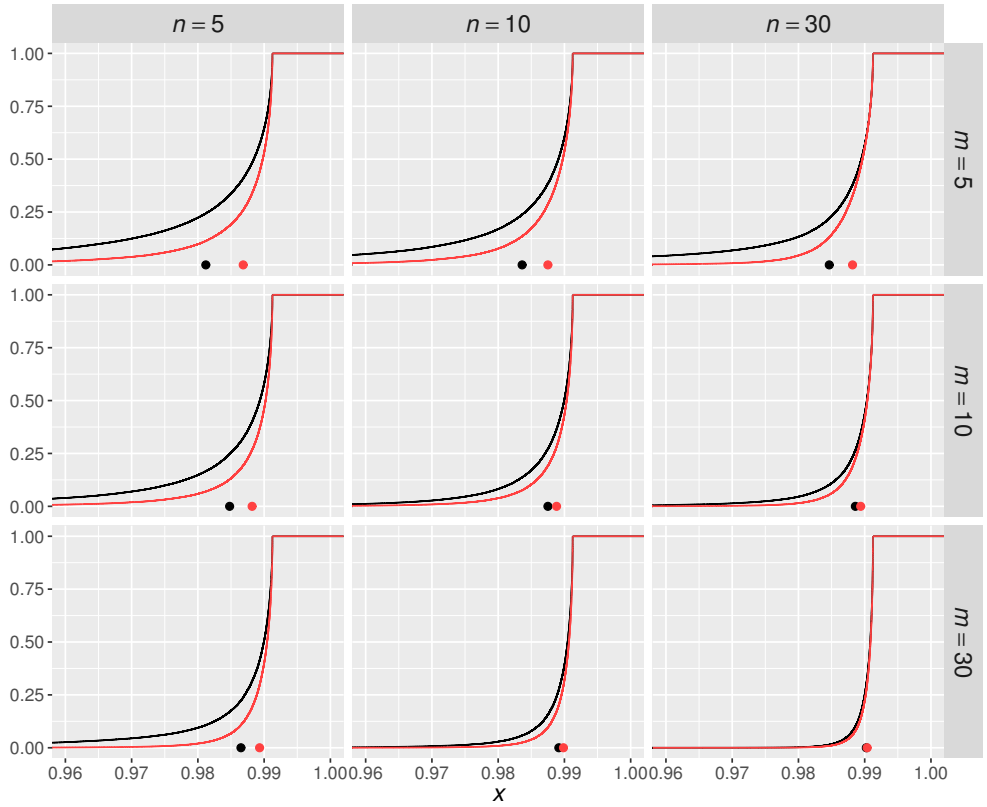

Figure 3.2: Black: $\mathbb{P}(\hat{\Psi}_{20}(3,2) \leq x)$ (curves) and $\mathbb{E}[\hat{\Psi}_{20}(3,2)]$ (dots). Red: $\mathbb{P}(\tilde{\Psi}_{20}(3,2) \leq x)$ (curves) and $\mathbb{E}[\tilde{\Psi}_{20}(3,2)]$ (dots). The sample sizes are $n \in \{5, 10, 30\}$ (columns) and $m \in \{5, 10, 30\}$ (rows).

done under the estimated distributions, the one resulting from the predictive distributions is less likely to yield a joint survival probability worse (smaller) than any value $x \in (0, 1)$. The predictive distributions thus lead to a better allocation in general, and this is well illustrated by the dots at the bottom of the plots, which indicate the expected joint survival probabilities $\mathbb{E}[\hat{\Psi}_k(\theta, \lambda)]$ and $\mathbb{E}[\tilde{\Psi}_k(\theta, \lambda)]$.

- As previously observed, when the sample sizes n and m increase, parameter uncertainty decreases, so both the estimated and predictive distributions are closer to each other, and hence so are the allocations $\hat{s}_k^*(\mathbf{x}, \mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x}, \mathbf{y})$ in general. As a result, for larger n and m the distributions $\mathbb{P}(\hat{\Psi}_k(\theta, \lambda) \leq x)$ and $\mathbb{P}(\tilde{\Psi}_k(\theta, \lambda) \leq x)$ are closer to one another. For $n = m = 30$, the difference almost vanished, in both Figures 3.1 and 3.2, so the improvements provided by the predictive distributions are really effective for small sample sizes.

Note that, while in this example the predictive distributions always lead to a better distribution of the joint survival probability than the estimated distributions, in other circumstances this may not necessarily hold. There are in fact several conflicting effects responsible for the difference between the two above approaches, and it would be interesting to study them explicitly and extensively in future research, and in a more general

framework. Yet, the present example shows that predictive distributions may lead to better resource allocations, and here this can be understood as follows.

First, note that while the estimated distributions $\hat{F}_X(x|\mathbf{x})$ and $\hat{F}_Y(x|\mathbf{y})$ are Pareto, the predictive distributions $\tilde{F}_X(x|\mathbf{x})$ and $\tilde{F}_Y(x|\mathbf{y})$ are "Log-Pareto". Since the true distributions are Pareto, using $\tilde{F}_X(x|\mathbf{x})$ and $\tilde{F}_Y(x|\mathbf{y})$ instead of $\hat{F}_X(x|\mathbf{x})$ and $\hat{F}_Y(x|\mathbf{y})$ can be considered as introducing some sort of bias. However, given that the predictive distributions lead to better resource allocations (i.e. larger joint survival probability) in general, this bias turns out to be negligible. Both shares $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$ thus vary around their theoretical optimal value $s_k^*(\theta,\lambda)$ as $\mathbf{x}$ and $\mathbf{y}$ vary, and the better performance of the predictive distributions is in fact principally due to that $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$ varies less than $\hat{s}_k^*(\mathbf{x},\mathbf{y})$. The reason for this is as follows. Since $\hat{F}_X(x|\mathbf{x})$ is Pareto and $\tilde{F}_X(x|\mathbf{x})$ is "Log-Pareto", the latter has a heavier tail (the same comment applies to Y, but for simplicity we will focus on X hereafter). This yields that for sufficiently large x, $\hat{F}_X(x|\mathbf{x})$ is flatter than $\tilde{F}_X(x|\mathbf{x})$. Then, in the above scenarios, the resource amount k is large, in that it generally makes $s_k^*(\theta,\lambda)$ fall in the region where $\hat{F}_X(x|\mathbf{x})$ is flatter than $\tilde{F}_X(x|\mathbf{x})$. Now, say that $\mathbf{y}$ varies. The objective functions $\hat{F}_X(s|\mathbf{x})\hat{F}_Y(k-s|\mathbf{y})$ and $\tilde{F}_X(s|\mathbf{x})\tilde{F}_Y(k-s|\mathbf{y})$ will consequently both vary, and hence so will the resulting shares $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$. Yet, since $\hat{F}_X(s|\mathbf{x})$ tends to be flatter than $\hat{F}_X(s|\mathbf{x})$ in the region of interest (i.e. for s close to $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ and $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$), it is less sensitive to changes in s, which lead to larger changes in $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ consecutive to variations in $\mathbf{y}$. So $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ is more sensitive to $\mathbf{y}$ than $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$. The same reasoning applies to $\mathbf{x}$, and hence the share $\hat{s}_k^*(\mathbf{x},\mathbf{y})$ tends to vary more than $\tilde{s}_k^*(\mathbf{x},\mathbf{y})$ around the theoretical optimal value $s_k^*(\theta,\lambda)$. As a result, the joint survival probability $\tilde{\Psi}_k(\theta,\lambda)$ deviates less than $\hat{\Psi}_k(\theta,\lambda)$ from the theoretical maximal joint survival probability $\psi_k^*(\theta,\lambda)$, and thus here the predictive distributions can be considered as leading to better resource allocations.

3.3 Conclusion

This note considered the problem of risk assessment in presence of parameter uncertainty. Two risk assessment techniques were compared, namely estimated distributions and predictive distributions. By means of a simple example, we illustrated the ability of the former to yield better allocations of resources, and hence, from that viewpoint, to assess risk in a more relevant way. Possible extensions of that work would be to assume more risks and consider model uncertainty, which will undoubtedly allow for larger improvements. Finally, it will be interesting to study other objective functions (e.g. more adapted to risk pricing), and also to compare predictive distributions with other distribution estimation (fitting) techniques.

Acknowledgements. I would like to thank Hansjörg Albrecher for his valuable inputs, which have helped me substantially improve this chapter. I would like to also thank Andreas Tsanakas for his valuable and insightful comments.

Chapter 4

An alternative path to extremes

This chapter is based on L. Vincent (2022): *An alternative path to extremes*. Preprint, University of Lausanne [77].

Abstract. The Fisher-Tippett-Gnedenko theorem characterizes the only possible non-degenerate limit distributions for the linearly normalised maximum of a large sample of independent and identically distributed random variables, and it is one of the foundational results of extreme value theory. That theorem naturally compares with the central limit theorem, which establishes the normality of linearly normalised sample sums, under some mild conditions. However, unlike the central limit theorem, the Fisher-Tippett-Gnedenko theorem can be applied in various ways in a given situation, and this potentially results in different limit distributions for the considered sample maximum. In this note, we discuss this aspect of the Fisher-Tippett-Gnedenko theorem, and we use it to propose an alternative family of extreme value distributions, which contains both the Weibull and Fréchet distributions, but where the Gumbel distribution is modified, and ends up being endowed with a shape parameter.

4.1 Introduction

One of the foundational results of extreme value theory is the Fisher-Tippett-Gnedenko theorem [40, 48], which is as follows (for an extensive treatment of extreme value theory, we refer to Embrechts et al. [32], Beirlant et al. [11] and de Haan and Ferreira [28], which are classical textbooks on that topic).

Theorem 4.1.1. *(Fisher-Tippett-Gnedenko theorem) Let $X_1, ..., X_n$ be a sequence of n mutually independent and identically distributed (iid) random variables, with common distribution F. Let further $M_n = \max\{X_1, ..., X_n\}$ be the corresponding sample maximum. If there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ and some non-degenerate distribution G such that*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - d_n}{c_n} \leq x\right) = F(c_n x + d_n)^n = G(x), \tag{4.1}$$

then G is necessarily one of the three following:

$$\begin{aligned} \textit{Weibull}: \ & \Psi_\alpha(x) = \exp(-(-x)^\alpha)\mathbb{1}\{x \leq 0\} + \mathbb{1}\{x > 0\}, \\ \textit{Gumbel}: \ & \Lambda(x) \ = \exp(-e^{-x}), \\ \textit{Fréchet}: \ & \Phi_\alpha(x) = \exp(-x^{-\alpha})\mathbb{1}\{x > 0\}, \end{aligned} \tag{4.2}$$

where $\alpha > 0$.

Accordingly, if G indeed appears as the limit distribution in (4.1), then F is said to be in the Maximum Domain of Attraction (MDA) of G, which is written as $F \in \text{MDA}(G)$, with $G \in \{\Psi_\alpha, \Lambda, \Phi_\alpha\}$.

Since the Fisher-Tippett-Gnedenko characterizes the potential limit behaviour of some linearly normalised combination of a large sample, it naturally compares with the central limit theorem.

Theorem 4.1.2. *(Classical central limit theorem) Let $X_1, ..., X_n$ be a sequence of n iid random variables, with common mean $\mathbb{E}[X] \in \mathbb{R}$ and variance $\mathbb{V}[X] \in (0, \infty)$. Let further $S_n = X_1 + ... + X_n$ be the corresponding sample sum, and write $a_n = \sqrt{n\mathbb{V}[X]}$ and $b_n = n\mathbb{E}[X]$. Then*

$$\lim_{n\to\infty} \frac{S_n - b_n}{a_n} \overset{d}{=} Z, \tag{4.3}$$

where Z is a Normal$(0, 1)$ random variable.

Yet, we may already remark one difference, namely that while the limit distribution of $a_n^{-1}(S_n - b_n)$ is unique, there are three different candidates for that of $c_n^{-1}(M_n - d_n)$. Therefore, on the one hand, for $\mathbb{V}[X] \in (0, \infty)$ and sufficiently large n, the central limit theorem allows to write

$$S_n \overset{d}{\approx} a_n Z + b_n, \text{ with } Z \sim \text{Normal}(0, 1). \tag{4.4}$$

On the other hand, for $F \in \text{MDA}(G)$ and sufficiently large n, according to the Fisher-Tippett-Gnedenko theorem we shall have

$$M_n \overset{d}{\approx} c_n W + d_n, \text{ with } W \sim G, \tag{4.5}$$

but here G could be either Ψ_α, Λ or Φ_α. As a consequence, if F is unknown, then the distribution of S_n can be approximated with that of a Normal(b_n, a_n) random variable, which thus only requires to estimate the parameters a_n and b_n. But for approximating the distribution of M_n, one must first estimate which of the three limit distributions (models) in (4.2) applies, and then estimate the parameters c_n and d_n (as well as α, potentially), which is a different and more complicated problem than simply estimating a_n and b_n.

There is, however, a convenient ingenuity, which allows to transform that model estimation problem into a parameter estimation problem: Firstly, the three distributions Ψ_α, Λ and Φ_α may be grouped into a single parametric family $\mathcal{G} = \{G_\xi, \xi \in \mathbb{R}\}$, by setting

$$G_\xi(x) = \begin{cases} \Psi_{-1/\xi}(x), & \xi < 0, \\ \Lambda(x), & \xi = 0, \\ \Phi_{1/\xi}(x), & \xi > 0. \end{cases} \tag{4.6}$$

We may call $\mathcal{G}$ the extreme value distribution family. Secondly, note that while G_ξ is the only possible non-degenerate limit for $F(c_n x + d_n)^n$, the choice of c_n and d_n is not unique, which allows to endow G_ξ with location and scale parameters. Indeed, if (4.1) holds, then for any $\sigma > 0$ and $\mu \in \mathbb{R}$, the constants $c'_n = c_n/\sigma$ and $d'_n = d_n - \mu/\sigma$ result in

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - d'_n}{c'_n} \leq x\right) = G_\xi\left(\frac{x-\mu}{\sigma}\right) = G_{\xi;\mu,\sigma}(x). \tag{4.7}$$

Thirdly, the freedom of choice for μ and σ when normalising M_n can be exploited to assemble the three distinct cases in (4.6) into a single and concise representation, due to von Mises [64] and Jenkinson [56]. The latter is called the Generalized Extreme Value distribution (GEV), and its formula is given by

$$H_\xi(x) = G_{\xi;-1/\xi,1/\xi}(x) = \exp(-(1+\xi x)_+^{-1/\xi}), \tag{4.8}$$

where $H_0(x)$ is interpreted as $\lim_{\xi\to 0} H_\xi(x) = \exp(-e^{-x})$.

By construction, we have $F \in \text{MDA}(G) \Leftrightarrow F \in \text{MDA}(G_\xi) \Leftrightarrow F \in \text{MDA}(H_\xi)$, so (4.5) can be re-expressed using the GEV. That is, assuming $F \in \text{MDA}(H_\xi)$, for sufficiently large n we shall have

$$M_n \overset{d}{\approx} c_n W_\xi + d_n, \text{ with } W_\xi \sim H_\xi, \tag{4.9}$$

and the distribution of M_n can thus be approximated by

$$H_{\xi;d_n,c_n}(x) = H_\xi\left(\frac{x-d_n}{c_n}\right), \tag{4.10}$$

which now only requires to estimate the parameters c_n, d_n and ξ. Therefore, as previously mentioned, by relying on the GEV, the model estimation problem has indeed turned into a parameter estimation one.

The use of the GEV is the standard way by which the Fisher-Tippett-Gnedenko theorem is applied in practice, through the so-called block-maxima method (see e.g. Chapter 6 in Embrechts et al. [32] for details). And since the parameter ξ will rarely be estimated at exactly 0, if the true model is the Gumbel distribution Λ, i.e. $\xi = 0$, then the GEV will in general lead to approximate the latter with either a Weibull or a Fréchet distribution.

In this chapter, we show that there is not a unique way to apply the Fisher-Tippett-Gnedenko theorem for determining the limit distribution of a sample maximum. This is due to what we call the commutativity property of the maximum, which we explain in Section 4.2. In Section 4.3, we use this property to derive an alternative extreme value distribution family $\mathcal{G}^*$. The latter contains both the Weibull and Fréchet distributions, as $\mathcal{G}$ does, but it replaces Λ with a new type of distribution, which can be interpreted as a Gumbel distribution endowed with a shape parameter. We then discuss that result in Section 4.4, and finally Section 4.5 concludes this chapter.

4.2 On the commutativity of the maximum

We now consider a second difference between the Fisher-Tippett-Gnedenko and the central limit theorems, which is that the action of taking the maximum commutes with strictly increasing transformations of the sample, while summation does not: Let t be a strictly increasing function that maps the domain of X to a subset of $\mathbb{R}$. Let further t^{-1} be the inverse of t. Then we can write $M_n = t(t^{-1}(M_n)) = t(\tilde{M}_n)$ and $S_n = t(t^{-1}(S_n)) = t(\tilde{S}_n)$, where thus $\tilde{M}_n = t^{-1}(M_n)$ and $\tilde{S}_n = t^{-1}(S_n)$. Since t is strictly increasing, this yields

$$\tilde{M}_n = t^{-1}(\max\{X_1, ..., X_n\}) = \max(t^{-1}(X_1), ..., t^{-1}(X_n)), \tag{4.11}$$

which is what we call the commutative property of the maximum. Regarding $\tilde{S}_n$, we have

$$\tilde{S}_n = t^{-1}(X_1 + ... + X_n) \neq t^{-1}(X_1) + ... + t^{-1}(X_n), \tag{4.12}$$

in general (unless t is the identity function), and hence the sum does not satisfy the same commutative property than the maximum.

It stems from (4.12) that $\tilde{S}_n$ will not be a sum of iid random variables in general, in which case the central limit theorem cannot be applied to $\tilde{S}_n$, but only to S_n, i.e. through (4.3). On the other hand, since $X_1, ..., X_n$ are iid random variables, so are $t^{-1}(X_1), ..., t^{-1}(X_n)$. As a result, just as M_n, $\tilde{M}_n$ is the maximum of some iid random variables, and we can apply it the Fisher-Tippett-Gnedenko theorem. So in order to come up with a limit distribution for M_n, we may either try to find constants c_n and d_n specific to M_n, such that

$$\lim_{n\to\infty} \frac{M_n - d_n}{c_n} \overset{d}{=} W \sim G. \tag{4.13}$$

Or, alternatively, since $M_n = t(\tilde{M}_n)$, we may obtain a limit distribution for M_n by finding constants $\tilde{c}_n$ and $\tilde{d}_n$ specific to $\tilde{M}_n$, such that

$$\lim_{n\to\infty} t\left(\frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n}\right) \overset{d}{=} t(\tilde{W}). \tag{4.14}$$

And interestingly, the two resulting limit distributions (i.e. that of W and $t(\tilde{W})$) can differ. The two following examples illustrate this. For the explicit expressions of constants involved in (4.15) – (4.18), see Table 3.4.4 in Embrechts et al. [32].

Example 4.2.1. Let $X_1, ..., X_n$ be iid occurrences of $X \sim \text{LogNormal}(\mu, \sigma)$, with $\sigma > 0$ and $\mu \in \mathbb{R}$. Then constants c_n and d_n can be found such that

$$\lim_{n\to\infty} \frac{M_n - d_n}{c_n} \overset{d}{=} W_0 \sim G_0, \tag{4.15}$$

i.e. the LogNormal distribution is in the Gumbel MDA. Alternatively, since X has a LogNormal(μ, σ) distribution, we can take $t(x) = \exp(x)$ and write $X = t(\tilde{X})$, where $\tilde{X} = t^{-1}(X) = \ln(X) \sim \text{Normal}(\mu, \sigma)$, and hence $M_n = \exp(\tilde{M}_n)$, with $\tilde{M}_n =$

$\max\{\tilde{X}_1, ..., \tilde{X}_n\}$. It turns out that the Normal distribution is also in the Gumbel MDA, so constants $\tilde{c}_n$ and $\tilde{d}_n$ can be found such that

$$\lim_{n\to\infty} \frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n} \overset{d}{=} \tilde{W}_0 \sim G_0. \tag{4.16}$$

As a result, by properly normalising M_n, we can make it have a Gumbel limit distribution, cf. (4.15). On the other hand, if we normalise $\tilde{M}_n$ instead, $M_n = \exp(\tilde{M}_n)$ yields that M_n will in consequence behave as

$$\lim_{n\to\infty} \exp\left(\frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n}\right) \overset{d}{=} \exp(\tilde{W}_0),$$

so that in the limit its distribution will be that of a "Log-Gumbel", which is easily checked to be a Fréchet(1), i.e. $\tilde{W}_0 \sim \Lambda \Leftrightarrow \exp(\tilde{W}_0) \sim \Phi_1$.

Example 4.2.2. Let $X_1, ..., X_n$ be iid occurrences of $X \sim$ Weibull(κ, λ), with $\kappa > 0$ and $\lambda > 0$. Note that here we are considering the "classical" Weibull distribution, which can be expressed as $F(x) = 1 - \exp(\lambda x^\kappa)$, $x > 0$, and it must not be confused with the "extremal" (or "negative") Weibull in (4.2). The Weibull distribution is in the Gumbel MDA, meaning that constants c_n and d_n can be found, so as to obtain

$$\lim_{n\to\infty} \frac{M_n - d_n}{c_n} \overset{d}{=} W_0 \sim G_0. \tag{4.17}$$

On the other hand, since X has a Weibull distribution, we can take $t(x) = x^{1/\kappa}$ and write $X = t(\tilde{X})$, where thus $\tilde{X} = t^{-1}(X) = X^\kappa \sim$ Exponential(λ), so

$$M_n = \tilde{M}_n^{1/\kappa}.$$

Furthermore, since the Exponential distribution is also in the Gumbel MDA, there again exist constants $\tilde{c}_n$ and $\tilde{d}_n$, such that

$$\lim_{n\to\infty} \frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n} \overset{d}{=} \tilde{W}_0 \sim G_0. \tag{4.18}$$

Therefore, (4.17) and (4.18) allow to come up with two different limit distributions for $M_n = \tilde{M}_n^{1/\kappa}$. That is, either we apply the Fisher-Tippett-Gnedenko theorem to M_n and make it have a Gumbel distribution, as in (4.17). Or we do it to $\tilde{M}_n$ instead. For instance, for $\kappa = 1/2$ this leads to

$$\lim_{n\to\infty} \left(\frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n}\right)^2 \overset{d}{=} \tilde{W}_0^2,$$

i.e. a squared Gumbel random variable. The general case $\kappa > 0$ requires to be handled more carefully. Indeed, since the domain of $\tilde{W}_0$ is $\mathbb{R}$, the expression $\tilde{W}_0^{1/\kappa}$ could be undefined for $W_0 < 0$, e.g. when $\kappa = 2$. To solve this potential issue, we can take

$$t(x) = \text{sign}(x)\ |x|^{1/\kappa},$$

which is strictly increasing. Accordingly, we have

$$M_n = t(\tilde{M}_n) = \text{sign}(\tilde{M}_n)\ |\tilde{M}_n|^{1/\kappa}, \tag{4.19}$$

where thus $\tilde{M}_n = \text{sign}(M_n)\ |M_n|^{\kappa}$, from $t^{-1}(x) = \text{sign}(x)\ |x|^{\kappa}$. By then applying the Fisher-Tippett-Gnedenko theorem to $\tilde{M}_n$ in (4.19), we obtain

$$\lim_{n\to\infty} \text{sign}\left(\frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n}\right) \left|\frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n}\right|^{1/\kappa} \overset{d}{=} \text{sign}(\tilde{W}_0)\ |\tilde{W}_0|^{1/\kappa}, \tag{4.20}$$

where

$$\mathbb{P}(\text{sign}(\tilde{W}_0)\ |\tilde{W}_0|^{1/\kappa} \leq x) = \begin{cases} \exp(-e^{(-x)^{\kappa}}), & x \leq 0, \\ \exp(-e^{-x^{\kappa}}), & x > 0, \end{cases} \quad \text{for all } \kappa > 0. \tag{4.21}$$

Therefore, similarly to what we did in Example 4.2.1, we illustrated two different ways to come up with an asymptotic distribution for M_n. That is, either we apply the Fisher-Tippett-Gnedenko theorem to M_n directly (the left-hand side of (4.19)), which yields a Gumbel distribution, or to $\tilde{M}_n$ (in the right-hand side of (4.19)), which results in the distribution (4.21).

So in a given situation, unlike the central limit theorem, the Fisher-Tippett-Gnedenko theorem can be applied in different ways, which in fact correspond to different types of normalisation of M_n. For instance, in Example 4.2.1, the limit distribution of M_n is Gumbel under a proper location-scale (i.e. linear) normalisation, cf. (4.15). And since in that example we took $t(x) = \exp(x)$, the location-scale normalisation of $\tilde{M}_n$ in (4.16) corresponds to a scale-power normalisation of M_n, i.e.

$$t\left(\frac{t^{-1}(M_n) - \tilde{d}_n}{\tilde{c}_n}\right) = \exp\left(\frac{\ln(M_n) - \tilde{d}_n}{\tilde{c}_n}\right) = s_n M_n^{p_n},$$

where $p_n = 1/\tilde{c}_n$ and $s_n = \exp(-\tilde{d}_n/\tilde{c}_n)$, we have that M_n goes to a Fréchet distribution under a scale-power normalisation (see e.g. Pantcheva [67] and Barakat et al. [9] for contributions on this specific type of normalisation). The same comment applies to Example 4.2.2, but since there we considered $t(x) = \text{sign}(x)\ |x|^{1/\kappa}$, the normalising operation of M_n in (4.20) is

$$t\left(\frac{t^{-1}(M_n) - \tilde{d}_n}{\tilde{c}_n}\right) = \text{sign}\left(\frac{\text{sign}(M_n)\ |M_n|^{\kappa} - \tilde{d}_n}{\tilde{c}_n}\right) \left|\frac{\text{sign}(M_n)\ |M_n|^{\kappa} - \tilde{d}_n}{\tilde{c}_n}\right|^{1/\kappa}, \tag{4.22}$$

and hence it is not of a location-scale or scale-power type. For simplicity we call it hereafter a polynomial normalisation.

4.3 The alternative extreme value distribution family

Let us now extend the idea of Example 4.2.2, and consider applying the Fisher-Tippett-Gnedenko theorem within the polynomial normalisation (4.22), but in the general case,

i.e. when X is not necessarily a Weibull(κ, λ) random variable. We show that this leads to a new extreme value distribution family, where the Weibull and Fréchet are preserved, but the Gumbel distribution is replaced by a version of it endowed with a shape parameter.

So let us take t as in Example 4.2.2, i.e.

$$t(x) = \text{sign}(x)\ |x|^{1/\kappa},$$

for some $\kappa > 0$. We thus have $t^{-1}(x) = \text{sign}(x)\ |x|^{\kappa}$, and we write again $X = t(\tilde{X})$ and $M_n = t(\tilde{M}_n)$, where thus $\tilde{X} = t^{-1}(X) = \text{sign}(X)\ |X|^{\kappa}$ and $\tilde{M}_n = t^{-1}(M_n) = \text{sign}(M_n)\ |M_n|^{\kappa}$.

Let $\tilde{F}$ be the distribution of $\tilde{X}$. Assuming

$$\tilde{F} \in \text{MDA}(G_{\xi\kappa}),\ \text{with}\ \xi \in \mathbb{R}, \tag{4.23}$$

the Fisher-Tippett-Gnedenko theorem yields that there exist constants $\tilde{c}_n$ and $\tilde{d}_n$, such that

$$\lim_{n\to\infty} \frac{\tilde{M}_n - \tilde{d}_n}{\tilde{c}_n} \overset{d}{=} \tilde{W}_{\xi\kappa} \sim G_{\xi\kappa},$$

and hence

$$\lim_{n\to\infty} t\left(\frac{t^{-1}(M_n) - \tilde{d}_n}{\tilde{c}_n}\right) = \text{sign}(\tilde{W}_{\xi\kappa})\ |\tilde{W}_{\xi\kappa}|^{1/\kappa}.$$

Depending on ξ, the right-hand side of the above equation has distribution:

- Case $\xi < 0$: Since $\kappa > 0$, we have $\xi\kappa < 0$, so $\tilde{W}_{\xi\kappa}$ has distribution $\Psi_{-1/(\xi\kappa)}$ and hence domain $(-\infty, 0]$, which leads to

$$\mathbb{P}(\text{sign}(\tilde{W}_{\xi\kappa})\ |\tilde{W}_{\xi\kappa}|^{1/\kappa} \le x) = \Psi_{-1/(\xi\kappa)}(-(-x)^{\kappa}) = \Psi_{-1/\xi}(x).$$

- Case $\xi = 0$: In that case $\xi\kappa = 0$, so $\tilde{W}_{\xi\kappa} = \tilde{W}_0$ has distribution Λ, domain $\mathbb{R}$, and we obtain

$$\begin{aligned}\mathbb{P}(\text{sign}(\tilde{W}_0)\ |\tilde{W}_0|^{1/\kappa} \le x) &= \begin{cases}\exp(-e^{(-x)^{\kappa}}), & x \le 0,\\ \exp(-e^{-x^{\kappa}}), & x > 0,\end{cases}\\ &= \Upsilon_{\kappa}(x),\ \text{for all}\ \kappa > 0,\end{aligned}$$

  which is in fact as in (4.21).

- Case $\xi > 0$: Here $\tilde{W}_{\xi\kappa}$ has distribution $\Phi_{\xi\kappa}$ and hence domain $(0, \infty)$, resulting in

$$\mathbb{P}(\text{sign}(\tilde{W}_{\xi\kappa})\ |\tilde{W}_{\xi\kappa}|^{1/\kappa} \le x) = \Phi_{1/(\xi\kappa)}(x^{\kappa}) = \Phi_{1/\xi}(x).$$

By then reassembling those three distinct cases, we obtain that under Assumption (4.23) and the polynomial normalisation (4.22), the sample maximum M_n has the limit distribution

$$G^*_{\xi,\kappa}(x) = \mathbb{P}(\text{sign}(\tilde{W}_{\xi\kappa})\ |\tilde{W}_{\xi\kappa}|^{1/\kappa} \le x) = \begin{cases}\Psi_{-1/\xi}(x), & \xi < 0,\\ \Upsilon_{\kappa}(x), & \xi = 0,\\ \Phi_{1/\xi}(x), & \xi > 0,\end{cases}$$

resulting in the alternative extreme value distribution family $\mathcal{G}^* = \{G^*_{\xi,\kappa}, \xi \in \mathbb{R}, \kappa > 0\}$.

4.4 Discussion

It is well-known that the Gumbel MDA contains very different types of distributions, both light- and heavy-tailed, and this was partly illustrated in Examples 4.2.1 and 4.2.2, where we recalled that the Normal, LogNormal, Exponential and "classical" Weibull distributions are all in the Gumbel MDA. In addition, unlike the "extremal" Weibull and Fréchet distributions Ψ_α and Φ_α, the Gumbel distribution Λ has no shape parameter. Therefore, when applied to the maximum of a sample with marginal distribution $F \in \text{MDA}(\Lambda)$, the linear normalisation turns out to be a "Procrustean" operation, since it always makes it have a single parameterless distribution Λ, no matter of its original characteristics.

Interestingly, while the polynomial normalisation that we considered above let the Weibull and Fréchet MDAs unchanged, it creates a somewhat new MDA. The latter is characterized by the distribution Υ_κ, which can be interpreted as a Gumbel distribution endowed with a shape parameter, and it is thus more diverse than the Gumbel MDA. For instance, it allows to distinguish the limit distributions of sample maxima of two different "classical" Weibull distributions, as long as they have different shape parameters κ, see Example 4.2.2. Also, if F is unknown but believed to be in the MDA of Υ_κ, then κ will be an unknown parameter to be estimated, just as α (or ξ) for the Weibull and Fréchet MDAs, and this allows for a potentially better fit.

Moreover, it is worthwhile noting that Υ_κ can be interpreted as the distribution of the log of a random variable that has an "extremal" Weibull left-tail and a Fréchet right-tail, after a proper handling of negative values. This suggests the possibility to adapt the approach of Section 4.3, so as to obtain a concise version of $G^*_{\xi,\kappa}$, say $H^*_{\xi,\kappa}$, very much like the GEV, but where the Weibull and Fréchet distributions would be special cases of $H^*_{\xi,\kappa}$, instead of the Gumbel distribution being a special case of H_ξ. This is left for future research.

4.5 Conclusion

In this chapter, we outlined that there is not a unique way to apply the Fisher-Tippett-Gnedenko theorem in a given situation, contrarily to the central limit theorem. We used this aspect of the Fisher-Tippett-Gnedenko theorem, to propose an alternative family of extreme value distributions, which contains a version of the Gumbel distribution endowed with a shape parameter. We then discussed the eventual possibility to use the latter distribution to construct an alternative GEV. Given the close relationship between the Fisher-Tippett-Gnedenko and the Pickands-Balkema-de Haan theorems (which characterizes the possible limit distributions of the conditional excess of random variables over a large threshold, instead of sample maxima), it is natural to extend the results of this chapter to the latter theorem, and this will be the topic of future work. Finally, it will also be interesting to assess whether similar results can be obtained for limit theorems of multivariate extreme value theory.

Acknowledgements. I would like to thank Hansjörg Albrecher, for his valuable and insightful comments which have helped me substantially improve this chapter, and for the stimulating discussions on the topic.

Chapter 5

Structured reinsurance deals with reference to relative market performance

This chapter is based on L. Vincent, H. Albrecher and Y. Krvavych (2021): *Structured reinsurance deals with reference to relative market performance*. Published in *Insurance: Mathematics and Economics*, 101, 125-139 [80].

Abstract. In this chapter we study a specific type of structured reinsurance deals, for which the indemnification scheme is contingent upon the performance of the cedent, for instance measured in terms of his loss ratio compared to the average loss ratio of the market. We show that this type of deals may be efficiently used to manage risk in the presence of financial distress cost when the cover is provided to a cohort of insurers with positively correlated loss experience. In addition to theoretical results we quantitatively illustrate the potential performance improvement in a numerical example.

5.1 Introduction

Reinsurance is considered to be one of the key strategic levers used to manage risk and optimise capital and its cost in order to preserve and enhance shareholder value. Along with other risk levers, such as underwriting portfolio mix, diversification, asset mix, funding composition and dividend policy, it is often used to address the following key issues:

1. Risk-taking: what risks to write, where, and how much.

2. Risk retention: what part of risk to retain and what part to transfer.

3. Funding and cost of funding: how to fund the retained risk (i.e., debt, equity, reinsurance or hybrids) and at what cost.

Conventional forms of reinsurance – whilst being efficient in finding optimal risk management strategies to address mainly (2) and (3) and to some extent (1) – are uni-dimensional

focusing on insurance risk only and thus result in a local optimisation. In contrast, structured reinsurance deals are more flexible allowing to manage different parts of enterprise risk of the insurance company in a holistic way and thus enable global risk optimisation. Reinsurance companies offer such deals in various forms, concrete examples being catastrophe bonds, finite risk solutions or multi-line products. For a broad overview, we refer to Culp [25] and Albrecher et al. [1].

A frequently used type of reinsurance deal consists of an indemnification scheme that – in addition to the insurance risk – depends on the realization of one or more observable random variables (triggers) which are related to other sources of risk the insurer is exposed to. Structured deals of this type are sometimes referred to as *contingent covers*, i.e. the nature of the risk transfer applying is contingent upon the realization of such additional random variable(s). Contingent covers have already been studied by several authors and under different denominations. For example, Gründl and Schmeiser [49] discussed several approaches to pricing double-trigger contracts, which refer to reinsurance deals for which the indemnity is triggered if both the reinsured loss exceeds a fixed deductible and some capital market index falls below a pre-defined threshold value. Asimit et al. [6] considered a framework in which an insurer shares his risk with a reinsurer according to an indemnification scheme that may vary with the risk environment, and showed that for a wide class of risk measures layer-type indemnities with parameters that vary with the risk environment can minimize the resulting sum of risk measures of insurer and reinsurer. In other contributions, contingent covers are considered but the random variables they depend on are not related to any source of risk the insurer is already exposed to, and thus the contingent nature of the indemnification scheme is used as a purely mathematical tool. For instance, Gajek and Zagrodny [46] found that for an insurer being endowed with a fixed budget to purchase reinsurance, if the loss to be reinsured has discrete components, then covers contingent upon the realization of an independent random event can provide the insurer with a lower ruin probability than those which are not. Under the name of random treaties, Guerra and Centeno [50] used contingent reinsurance as a mathematical tool (and intermediate step) to solve an optimal reinsurance problem for deterministic treaties. In a framework with regulatory solvency constraints and cost of capital, Albrecher and Cani [2] studied a form of stop-loss cover, where the deductible is randomized according to an independent external mechanism and they showed that this can yield higher expected profits than traditional stop-loss covers.

While these three latter contributions reveal interesting properties of contingent covers, the main purpose of such covers in practice is to increase the efficiency of the risk transfer, by focusing it on the scenarios where the cedent expects to need it the most regarding his overall financial risk, and hence to appropriately mitigate that risk for a smaller reinsurance premium. It is therefore intuitive that for a contingent cover to be relevant, the random variable(s) it depends on should somehow be related to the overall financial result of the insurer. Among many different sources of potential losses contributing to the overall financial result, we would like to distinguish:

- Insurance (core) risk, which consists of the risks the insurer expects to make money on and generate return on capital (underwriting, reserving, catastrophe).

- Peripheral (non-core) risks, such as investment risk and operational risk, which are contained and actively managed to reduce leakages and additional drag on capital.
- Frictional cost, which emerges as the result of insufficient and/or inefficient control of risk.

In the actuarial literature, the peripheral risks of an insurer are sometimes referred to as background risks. While in practice peripheral risks and frictional cost should be considered as two distinct sources of potential losses, from a mathematical point of view the latter can also be treated as a materialisation of background risk. Multiple contributions have shown that the presence of background risk can influence the choice of reinsurance made by an insurer. For instance, Dana and Scarsini [26], Lu et al. [61] and Chi and Wei [22] showed that the optimal traditional (non-contingent) reinsurance cover can be influenced by the presence of a background risk. Fan [35] considered an insurer maximizing his expected utility and facing a background risk, and proved the stop-loss reinsurance with a deductible being contingent upon the realization of the background risk to be optimal in this setup.

In this chapter we consider a structured reinsurance deal whose indemnification scheme is contingent upon the performance of the insurer buying it, for instance measured in terms of his loss ratio relative to the average loss ratio of the market, or relative performance for short. There are several arguments for doing so, which we explain hereafter.
On the one hand, as the insurer incurs larger insurance losses, his solvency is at stake, which may put him into financial distress. During such periods, the insurer will face additional expenses (the path to ruin is costly), such as the ones related to the intervention of the regulator or the increased difficulty of issuing new debt or acquiring new business. These expenses, once combined together, are referred to as the *financial distress cost* (see e.g. Froot et al. [44]), which is a particular type of non-negligible frictional cost. Moreover, since the market stakeholders partly assess the performance of insurers by benchmarking them to one another, the ones who perform below the average will typically go through worse financial distress periods and thus have a larger financial distress cost. For this reason, but also simply because a bad relative performance is in general the consequence of large insurance losses, the overall financial risk of an insurer can be expected to be well related with his relative performance, which makes the latter an interesting candidate to be used in a contingent cover.
On the other hand, since the worse the relative performance of a particular insurer is, the better the one of the other insurers will be, a reinsurance company selling covers of this kind to several insurers in a given market will benefit from some degree of hedging.

It is hence assumed that a key driver of the need (demand) for this special type of reinsurance is the financial distress cost (this is also in line with findings in Krvavych and Sherris [58]). We will consider the above simple structured reinsurance deal as a toy model and assess the benefits it might bring when used for optimising capital resources to enhance firm value. In particular, we want to quantify the potential improvements, for both the insurer(s) and the reinsurer, when the latter offers such covers simultaneously to various insurers. To do so, in a simple yet realistic model, we consider a representative insurer

who manages his risk according to a scenario-based approach and seeks to minimize his reinsurance premium. We solve the resulting optimal reinsurance problem and explicitly derive the optimal reinsurance cover, both when contingent covers are available and when they are not. The optimal contingent and traditional covers are then compared, first theoretically and then quantitatively, by means of a numerical application.

The rest of the chapter is organised as follows. In Section 5.2, we formulate a model in which the insurer faces both an insurance risk and financial distress cost, and we then describe the two forms of reinsurance he has at his disposal, namely contingent and traditional covers. In Section 5.3, we formulate the optimization problem that allows us to determine both the optimal contingent cover and the optimal traditional one. In Section 5.4, the optimal contingent and traditional reinsurance covers are explicitly derived, and their properties are discussed and compared. In Section 5.5, we give a concrete application in which we are able to determine the cases where the optimal contingent cover outperforms the traditional one, and we quantify the eventual improvements. Finally, we provide a conclusion in Section 5.6.

5.2 Preliminaries

5.2.1 The Model

In this paper, all random variables are defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $P_X \geq 0$ be the total premium received by an insurer for covering an (annual aggregate) insurance loss. This insurance loss is represented by the non-negative random variable X with distribution function $F_X(x) = \mathbb{P}(X \leq x)$.

As a tool to manage his risk, the insurer may purchase a reinsurance cover, under which he will cede the portion R (the ceded loss) of his insurance loss to a reinsurance company (the reinsurer), and in turn pay the corresponding reinsurance premium P_R. That reinsurance premium is defined as $P_R = \pi[R]$, where the functional $\pi : \Omega \to \mathbb{R}_+$ satisfying $\pi[0] = 0$ is the premium principle, that is, the rule determining the amount to be paid by the insurer for ceding R.

Under a reinsurance cover, the part of X and P_X being retained by the insurer are thus $D = X - R$ (the retained loss) and $P_D = P_X - P_R$ (the retained premium). Here, for simplicity the insurer is assumed to face only one other source of loss, namely the financial distress cost. The latter is assumed to depend on the retained loss and is therefore denoted Y_D. The cash-flows are considered to occur according to the following sequence: the insurer first receives P_X and pays P_R at the same time ($t = 0$), and then one year later ($t = 1$) he pays D and Y_D. We refer to the sum

$$H_D := D + Y_D$$

as the insurer's retained risk.

The financial distress cost is modelled as a function of the excess of the retained loss over the retained premium, and its concrete shape depends on the insurer's relative performance described by the discrete random variable Z, leading to

$$Y_D = g_Z\big((D - P_D)_+\big) = \sum_{z \in \mathcal{Z}} g_z\big((D - P_D)_+\big) \cdot \mathbb{1}\{Z = z\},$$

where $g_z : \mathbb{R}_+ \to \mathbb{R}_+$ is the financial distress cost function for the risk scenario z and $\mathbb{1}\{Z = z\}$ the indicator function of the event $Z = z$. We assume that

$$g_z(x) \text{ is continuous and increasing in } x, \text{ with } g_z(0) = 0, \text{ for all } z. \tag{A1}$$

(throughout this paper, we write "increasing" for "non-decreasing" and "decreasing" for "non-increasing"). As a consequence, the financial distress cost is non-negative, and can only be greater than 0 if the retained loss of the insurer exceeds the retained premium. In practice, the shape of each function g_z would typically be linked to the market capitalization of the insurer, as studied by Froot [45].

The insurer is part of an insurance market with a total of $n \geq 2$ insurers. Let V and $\overline{V}_n$ be variables measuring at $t = 1$ the performance realized during the elapsed year by the insurer and the market, respectively. Typically V is a function of X, and $\overline{V}_n$ is based on public data. For instance, in the concrete example in Section 5.5, we will consider the insurer's loss ratio $V = X/P_X$, and $\overline{V}_n$ will be the average loss ratio of the market.

The relative performance of the insurer is now modelled as $Z := s(V, \overline{V}_n)$, where $s : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ is a bivariate step function. It is assumed that both the insurer and the reinsurer know the function s. The relative performance Z is thus a random variable at $t = 0$, whose realization can be known with no ambiguity by both the insurer and the reinsurer at $t = 1$. Correspondingly, Z is a discrete random variable that represents mutually exclusive scenarios for the realization of the relative performance of the insurer at $t = 1$, and we denote its domain by $\mathcal{Z}$.

Consider a risk scenario $z \in \mathcal{Z}$. The random variable $(X \mid Z = z)$ has the conditional distribution function $F_{X|Z=z}(x) = \mathbb{P}(X \leq x \mid Z = z)$, and we denote its Value-at-Risk at one particular $\beta \in [0, 1]$ by

$$\rho_z[X] := \mathrm{VaR}_\beta[X|Z = z] = \inf\{x : F_{X|Z=z}(x) \geq \beta\}.$$

Then, we make the two additional assumptions

$$\rho_z[X] \text{ increases in } z, \tag{A2}$$

and

$$g_z(x) \text{ increases in } z, \text{ for all } x, \tag{A3}$$

so Z can be seen as measuring how bad the relative performance of the insurer has been. From this, we finally define $\sup \mathcal{Z}$ as the worst-case scenario.

Remark 5.2.1. Studies on the optimal choice of reinsurance in the presence of background risk (or a related risk factor) often consider the property of stochastic increasingness (see e.g. Dana and Scarsini [26], Lu et al. [61] and Chi and Wei [22]). Note that in our framework the stochastic increasingness of X in Z – i.e. having $\mathbb{P}(X > x \mid Z = z)$ that increases in z, for all x – is a sufficient, but not a necessary condition for (A2).

5.2.2 Traditional and Contingent Reinsurance

The ceded loss of a traditional reinsurance cover is computed as

$$R = f(X)$$

where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a pre-defined deterministic function, referred to as the ceded loss function. For instance, the ceded loss function of a quota-share (qs) cover at $x \in \mathbb{R}_+$ is $f^{qs}(x) = a \cdot x$, where $a \in [0, 1]$ is the proportionality factor, and the one of a (traditional) bounded stop-loss (*tbsl*) (also referred to as limited stop-loss or one-layer reinsurance cover) is

$$f^{tbsl}(x; d, \ell) = \min\{(x - d)_+, \ell\},$$

where $d \geq 0$ is the deductible (or retention) and $\ell \geq 0$ the limit (or layer). The reader interested in a broader overview of traditional reinsurance covers can refer to Albrecher et al. [1].

In contrast to traditional reinsurance, under a contingent cover the type of the ceded loss function (or the values of the parameters involved in it) can depend on the realization of one or more random variables. For a reinsurance cover being contingent upon the realization of Z, the ceded loss is computed as

$$R = f_Z(X) = \sum_{z \in \mathcal{Z}} f_z(X) \cdot \mathbb{1}\{Z = z\},$$

where f_Z denotes the contingent ceded loss function that depends on Z (so f_z is the ceded loss function when $Z = z$).

Remark 5.2.2. Any traditional ceded loss function f is a special case of a contingent one with $f_z = f$ for all z. Consequently, any result that holds in general for contingent ceded loss functions, also holds for the traditional ones.

The contingent version of a bounded stop-loss ($cbsl$) has ceded loss function

$$f_Z^{cbsl}(x; d_Z, \ell_Z) = \min\{(x - d_Z)_+, \ell_Z\} = \sum_{z \in \mathcal{Z}} \min\{(x - d_z)_+, \ell_z\} \cdot \mathbb{1}\{Z = z\},$$

where $d_z \geq 0$ and $\ell_z \geq 0$ are the pre-defined deductible and limit that apply if $Z = z$, and $d_Z = \sum_{z \in \mathcal{Z}} d_z \cdot \mathbb{1}\{Z = z\}$ and $\ell_Z = \sum_{z \in \mathcal{Z}} \ell_z \cdot \mathbb{1}\{Z = z\}$ are the resulting contingent deductible and contingent limit.

5.3 Choice of Reinsurance

5.3.1 The Optimization Problem

Sometimes insurers are willing (or constrained by law) to measure and manage their risk in several distinct risk scenarios, rather than on the average over all the risk scenarios. When an insurer does so, but contingent reinsurance is not available (for instance because the realized risk scenario is not observable), he usually has no other choice than purchasing the traditional cover that fits his needs under the worst-case risk scenario and therefore tends to pay a high reinsurance premium. A typical example of a non-observable realized risk scenario is the one of model risk (or model uncertainty, ambiguity), that is, when an insurer considers multiple probability models to describe his risk but cannot determine which is the correct one (for a suggestion of mixing of quantile levels over different such scenarios, see e.g. Cohignac and Kazi-Tani [24]). In that case, if the insurer wants to manage his risk over all the considered models by purchasing a reinsurance cover, then he will have to make his choice under the worst one. In the literature, this situation is referred to as worst-case or minimax (maxmin) optimization. For recent studies on that topic we refer to Asimit et al. [5], Birghila and Pflug [15] and Birghila et al. [14]. If, by contrast, the realized risk scenario is observable (for instance, whether a hailstorm occurs or not in a particular geographical region and time period), then using it in a contingent reinsurance cover allows the insurer to choose a potentially different indemnification scheme for each risk scenario, which lowers the reinsurance premium but still achieves the desired mitigation of risk.

The nature of the financial distress cost motivates a scenario-based measurement of risk for each z separately. For the purposes of this paper, we thus consider, for all z, the concrete conditional risk measure

$$\rho_z[H_D] = \text{VaR}_\beta[H_D|Z = z],$$

where β is typically large. If the insurer enters into a reinsurance deal at $t = 0$, then $\rho_z[H_D]$ corresponds to the funds needed at $t = 1$ in order to limit his ruin probability to at most $1 - \beta$ in scenario z. The insurer then sets his maximal acceptable level of riskiness $k \geq 0$, and chooses the reinsurance cover so that none of the resulting conditional risk measures exceeds it. Concretely, if the insurer does so and has the funds k (and hence the capital $k - P_D$) available at $t = 1$, then for any realization of Z his ruin probability is at most $1 - \beta$, which represents an additional level of safety over simply having a ruin probability of at most $1 - \beta$.

Remark 5.3.1. For an insurer managing his risk according to the approach described above, the choice of the maximal acceptable level of risk k will depend on several factors, such as the resulting retained premium, the capital available and the objective function. While the determination of this choice is a question of interest, it is outside the scope of this paper. Therefore, in order to keep our results general, we will consider the maximal acceptable level of risk to be chosen exogenously and assume k to be given. However, the

optimality results from Section 5.4 can be used in a second step to determine k endogenously.

Since in general there will be several covers reducing all the conditional risk measures to at most k, we are led to the mathematical problem of choosing the contingent cover with the smallest premium. Also, in order to avoid ending up with a cover that has undesirable properties, as in Asimit et al. [6] we restrict this choice to ceded loss functions in the set

$$\mathcal{C}^{(1)} = \Big\{ f_Z : 0 \leq f_z(x) \leq x \text{ and both } f_z(x) \text{ and } x - f_z(x) \text{ are increasing functions, for all } z \Big\},$$

where the first condition ensures that the insurer cedes a loss that is neither negative nor greater than the original loss, and the second one reduces moral hazard from both parties within each risk scenario. The constraints in $\mathcal{C}^{(1)}$ are in fact an extension of those suggested by Huberman et al. [53], that were for traditional reinsurance covers only.

Remark 5.3.2. While restricting the choice of ceded loss functions to $\mathcal{C}^{(1)}$ reduces moral hazard within each risk scenario, the insurer may still have an incentive to misreport X and hence V in order to increase his ceded loss $f_Z(X)$. In practice, there may, however, be ways to deal with this issue applying a proper level of governance and auditing.

In mathematical terms, the optimal contingent reinsurance cover can be formulated as

$$f_Z^{(1)}(\,\cdot\,; k) = \underset{f_Z(\,\cdot\,;k) \in \mathcal{C}^{(1)}(k)}{\arg\min} \; \pi[f_Z(X)], \tag{5.1}$$

where

$$\mathcal{C}^{(1)}(k) = \Big\{ f_Z \in \mathcal{C}^{(1)} : \rho_z[H_{X - f_Z(X)}] \leq k, \text{ for all } z \Big\}$$

is the set of admissible contingent ceded loss functions for a given maximal acceptable level of riskiness k.

Remark 5.3.3. If instead of reducing the conditional risk measure $\rho_z[H_D]$ to at most k for each risk scenario z, we considered it to be done for $\text{VaR}_\beta[H_D]$ globally, then in many cases ceding less risk in more dangerous risk scenarios is the optimal strategy for minimizing the reinsurance premium (see e.g. some solutions in Asimit et al. [6]). A contingent cover with such a property might, however, be viewed as a theoretical optimization tool rather than a risk management instrument that could be implemented in practice. So among other advantages, the scenario-based approach above can be seen as a way to prevent the optimal contingent cover to have this possibly undesirable property.

Regarding the traditional reinsurance cover that will be used as a benchmark for the optimal contingent one, we consider the same problem as (5.1), but with $\mathcal{C}^{(1)}$ replaced by

$$\mathcal{C}^{(2)} = \Big\{ f : 0 \leq f(x) \leq x \text{ and both } f(x) \text{ and } x - f(x) \text{ are increasing functions} \Big\}.$$

The best choice of the benchmark cover can then be formulated as

$$f^{(2)}(\,\cdot\,; k) = \underset{f(\,\cdot\,;k) \in \mathcal{C}^{(2)}(k)}{\arg\min} \; \pi[f(X)], \tag{5.2}$$

where

$$\mathcal{C}^{(2)}(k) = \Big\{ f(x) \in \mathcal{C}^{(2)} : \rho_z[H_{X - f(X)}] \leq k, \text{ for all } z \Big\}.$$

Remark 5.3.4. When the relative performance and the financial distress cost are ignored, the constraints on the conditional risk measures in $\mathcal{C}^{(2)}(k)$ can be replaced by the single condition $\text{VaR}_\beta[X - f(X)] \leq k$. If reinsurance is priced according to the expected value principle, the optimal reinsurance problem (5.2) then becomes the dual problem of (3.10) in Chi and Tan [21], which consists in the minimization of $\text{VaR}[X - f(X)] + \pi[f(X)]$ under the set of admissible ceded loss functions $\mathcal{C}^{(2)}$. Problem (5.2) is thus closely related to (3.10) in Chi and Tan [21], and as a consequence so is problem (5.1), as it is simply a relaxed version of (5.2) (the difference being the set of admissible ceded loss functions, with $\mathcal{C}^{(2)}(k) \subseteq \mathcal{C}^{(1)}(k)$). Note that problems (5.1) and (5.2) slightly differ in spirit from the formulation of many other optimal reinsurance problems in the literature, in that the measure of risk is to be made acceptable rather than minimal (see also [1, Sec.8], the review paper Cai and Chi [17] and references therein). In our formulation, the insurer can make an explicit trade-off between the reinsurance premium and his maximal acceptable measure of riskiness k, see also Remark 5.3.1.

5.3.2 The Premium Principle

In this paper, problems (5.1) and (5.2) will be solved for monotone premium principles.

Definition 5.3.1. *The premium principle π is said to be monotone if for any two ceded losses R and $\tilde{R}$, having $R \leq \tilde{R}$ almost surely always results in $\pi[R] \leq \pi[\tilde{R}]$.*

An important property of monotone premium principles is that they preserve the order relationships of contingent ceded loss functions.

Definition 5.3.2. *If two contingent ceded loss functions f_Z and $\tilde{f}_Z$ satisfy $f_z(x) \leq \tilde{f}_z(x)$ for all x and z, then f_Z is said to be pointwise smaller than $\tilde{f}_Z$, and we write $f_Z \preceq \tilde{f}_Z$.*

We state this property formally:

Property 5.3.1. *If π is a monotone premium principle, then for any two ceded loss functions f_Z and $\tilde{f}_Z$ satisfying $f_Z \preceq \tilde{f}_Z$, we have $\pi[f_Z(X)] \leq \pi[\tilde{f}_Z(X)]$.*

A well-known instance of a monotone premium principle is the expected value premium principle

$$\pi[R] = (1 + \theta)\, \mathbb{E}[R],$$

where $\mathbb{E}[R]$ is the expectation of R and $\theta \geq 0$ the safety loading factor. Another example is the risk-adjusted premium principle introduced by Wang [81] (also referred to as Wang or distortion premium principle)

$$\pi[R] = \int_0^\infty w(\mathbb{P}(R > x))\, \mathrm{d}x,$$

with $w(u)$ being a non-negative increasing and concave function such that $w(0) = 0$ and $w(1) = 1$.

5.4 Optimality Results

5.4.1 Preliminaries

We begin by recalling Theorem 1 from Dhaene et al. [30], according to which for any increasing and continuous function $\eta(x)$ and $\beta \in [0,1]$, the Value-at-Risk of a random variable X satisfies

$$\text{VaR}_\beta[\eta(X)] = \eta(\text{VaR}_\beta[X]). \tag{5.3}$$

Given the definition of the financial distress costs Y_D, the retained risk of the insurer can alternatively be expressed as

$$H_D = h_Z(D; P_D) = \sum_{z \in \mathcal{Z}} h_z(D; P_D) \cdot \mathbb{1}\{Z = z\},$$

where

$$h_z(x; p) = x + g_z\big((x-p)_+\big)$$

is a function of x that depends on both p and z. More specifically, we have that

$$h_z(x; p) \text{ is a continuous and strictly increasing function of } x, \tag{5.4}$$

from (A1),

$$h_z(x; p) \text{ decreases in } p, \tag{5.5}$$

and finally

$$h_z(x; p) \text{ increases in } z, \tag{5.6}$$

from (A3).

Note that in absence of reinsurance, the insurer's retained risk is $H_X = h_Z(X; P_X)$.

We then define the random variables

$$h_Z^{-1}(u; p) = \sum_{z \in \mathcal{Z}} h_z^{-1}(u; p) \cdot \mathbb{1}\{Z = z\}$$

where $h_z^{-1}(u; p) = \inf\{x : h_z(x; p) \geq u\}$ is the inverse function of $h_z(x; p)$, and

$$\rho_Z[X] = \sum_{z \in \mathcal{Z}} \rho_z[X] \cdot \mathbb{1}\{Z = z\}.$$

For $i = 1, 2$, the minimal reinsurance premium and the related retained premium resulting from problem (i) are denoted

$$P_R^{(i)}(k) = \pi[f_Z^{(i)}(X; k)] \quad \text{and} \quad P_D^{(i)}(k) = P_X - P_R^{(i)}(k),$$

and the retained risk of the insurer is

$$H_D^{(i)}(k) = h_Z\big(X - f_Z^{(i)}(X; k); P_D^{(i)}(k)\big),$$

where $f_Z^{(2)}(\,\cdot\,; k) = f^{(2)}(\,\cdot\,; k)$.

Let us derive the following results.

Proposition 5.4.1. *For $i = 1, 2$, let π be a monotone premium principle and k the maximal acceptable level of riskiness. If the set $\mathcal{C}^{(i)}(k)$ of candidates is non-empty, then problem (i) admits a solution.*

Proof. For $i = 1, 2$, let $j \geq 1$ be an integer and $z \in \mathcal{Z}$. Let further $f_Z(\,\cdot\,; k, x_{j-1,z}, x_{j,z})$ be the pointwise smallest element of $\mathcal{C}^{(i)}(k)$ for $x \in [x_{j-1,z}, x_{j,z})$ and $Z = z$ (or an arbitrarily chosen one of them, if there are several), where $x_{0,z} = 0$ and $x_{j-1,z} \leq x_{j,z}$ for all j and z. Since by definition

$$f_Z(\,\cdot\,; k, x_{j-1,z}, x_{j,z}) \in \mathcal{C}^{(i)}(k), \text{ for all } j \text{ and } z, \tag{5.7}$$

the functions $f_z(x; k, x_{j-1,z}, x_{j,z})$ and $x - f_z(x; k, x_{j-1,z}, x_{j,z})$ are both increasing in x, and hence they are both continuous in x too, which yields

$$f_z(x_{j,z}; k, x_{j-1,z}, x_{j,z}) = f_z(x_{j,z}; k, x_{j,z}, x_{j+1,z}), \text{ for all } j \text{ and } z. \tag{5.8}$$

Let then

$$f_Z(x; k) = \sum_{z \in \mathcal{Z}} f_z(x; k) \cdot \mathbb{1}\{Z = z\},$$

where

$$f_z(x; k) = \sum_{j \geq 1} f_z(x; k, x_{j-1,z}, x_{j,z}) \cdot \mathbb{1}\{x \in [x_{j-1,z}, x_{j,z})\}.$$

Given (5.7), we know that $f_z(x; k)$ and $x - f_z(x; k)$ are both increasing in each interval $(x_{j-1,z}, x_{j,z})$. Therefore, since (5.8) yields that $f_z(x; k)$ is continuous in x at $x = x_{j,z}$, for all j and z, we have that $f_z(x; k)$ and $x - f_z(x; k)$ are both increasing in x, which yields

$$f_Z(\,\cdot\,; k) \in \mathcal{C}^{(i)}(k). \tag{5.9}$$

By definition, we have that $f_Z(\,\cdot\,; k) \preceq \tilde{f}_Z(\,\cdot\,; k)$, for any $\tilde{f}_Z(\,\cdot\,; k) \in \mathcal{C}^{(i)}(k)$, and hence from (5.9) we deduce that $f_Z(\,\cdot\,; k)$ is the pointwise smallest element in $\mathcal{C}^{(i)}(k)$. Since π is a monotone premium principle, Property 5.3.1 yields $f_Z^{(i)}(\,\cdot\,; k) = f_Z(\,\cdot\,; k)$, which proves the result. □

Proposition 5.4.2. *For $i = 1, 2$, let π be a monotone premium principle and $k_{\inf}^{(i)}$ the smallest maximal acceptable level of riskiness for which problem (i) admits a solution. Then problem (i) admits a solution for any $k \geq k_{\inf}^{(i)}$, and the resulting minimal reinsurance premium $P_R^{(i)}(k)$ decreases in k, down to 0 for $k \geq \rho_{\sup \mathcal{Z}}[H_X]$.*

Proof. For $i = 1, 2$, consider the maximal acceptable levels of riskiness k and $\tilde{k}$ satisfying $k_{\inf}^{(i)} \leq k \leq \tilde{k}$. By definition, for any $f_Z(\,\cdot\,; k) \in \mathcal{C}^{(i)}(k)$ we have $f_Z(\,\cdot\,; k) \in \mathcal{C}^{(i)}(\tilde{k})$, leading to

$$\mathcal{C}^{(i)}(k) \subseteq \mathcal{C}^{(i)}(\tilde{k}). \tag{5.10}$$

Since by definition problem (i) admits a solution for the maximal acceptable level of riskiness $k_{\inf}^{(i)}$, the set $\mathcal{C}^{(i)}(k_{\inf}^{(i)})$ is non-empty and hence from (5.10), neither is $\mathcal{C}^{(i)}(k)$, which, given Proposition 5.4.1, proves that problem (i) admits a solution for any k satisfying $k \geq k_{\inf}^{(i)}$.

Therefore, since k and $\tilde{k}$ satisfy $k_{\inf}^{(i)} \leq k \leq \tilde{k}$, the solutions $f_Z^{(i)}(\,\cdot\,;k)$ and $f_Z^{(i)}(\,\cdot\,;\tilde{k})$ are both defined, and from (5.10), they satisfy $f_Z^{(i)}(\,\cdot\,;k) \preceq f_Z^{(i)}(\,\cdot\,;\tilde{k})$. Given that π is monotone, from Property 5.3.1 this last relationship yields $P_R^{(i)}(k) \leq P_R^{(i)}(\tilde{k})$, which proves that $P_R^{(i)}$ decreases in k for any $k \geq k_{\inf}^{(i)}$.

Finally, consider the case where the insurer cedes no loss to the reinsurer, and hence $R = 0$ for any realization of X and Z. The corresponding contingent ceded loss function is

$$f_Z(x) = 0, \text{ for all } x \text{ and any realization of } Z, \tag{5.11}$$

which belongs to both $\mathcal{C}^{(i)}$. Since by definition $\pi[0] = 0$, the reinsurance premium resulting from (5.11) is

$$P_R = 0, \tag{5.12}$$

and for each z, the related conditional risk measure is thus $\rho_z[H_X]$. By virtue of (5.3) and (5.4), the latter can be rewritten as

$$\rho_z[H_X] = h_z(\rho_z[X]; P_X). \tag{5.13}$$

By considering properties (5.4) and (5.6) in (5.13), and since $\rho_z[X]$ is assumed to increase in z, the conditional risk measure $\rho_z[H_X]$ increases in z, which leads to

$$\rho_z[H_X] \leq \rho_{\sup \mathcal{Z}}[H_X], \text{ for all } z. \tag{5.14}$$

As a consequence, if

$$k \geq \rho_{\sup \mathcal{Z}}[H_X], \tag{5.15}$$

then from (5.14) we know that (5.11) belongs to $\mathcal{C}^{(i)}(k)$. Moreover, since by the definition of π the reinsurance premium must be non-negative, from (5.12) we deduce that under (5.15), the contingent ceded loss function (5.11) is a solution to problem (i), which proves that $P_R^{(i)}(k) = 0$ whenever $k \geq \rho_{\sup \mathcal{Z}}[H_X]$. □

Lemma 5.4.1. *For any $k \geq 0$ and $v \geq 0$, the pointwise smallest $f \in \mathcal{C}^{(2)}$ satisfying $f(v) \geq v - k$ is $f(x) = \min\{(x-k)_+, \ell\}$, where $\ell = (v-k)_+$.*

Proof. Let us first consider the case $0 \leq k < v$: the smallest value $f(v)$ satisfying $f(v) \geq v - k$ then is $f(v) = v - k$. For any $f \in \mathcal{C}^{(2)}$, we have

$$0 \leq f(x) - f(y) \leq x - y, \text{ for all } y \leq x,$$

which can be partitioned into

$$0 \leq f(v) - f(x) \leq v - x, \text{ for } x \in [0, v), \tag{5.16}$$

and

$$0 \leq f(x) - f(v) \leq x - v, \text{ for } x \in [v, \infty). \tag{5.17}$$

With $f(v) = v - k$, inequalities (5.16) and (5.17) become

$$x - k \leq f(x) \leq v - k, \text{ for } x \in [0, v),$$

and

$$v - k \leq f(x) \leq x - k, \text{ for } x \in [v, \infty),$$

from what we deduce that the pointwise smallest $f \in \mathcal{C}^{(2)}$ satisfying $f(v) \geq v - k$ is

$$\begin{aligned} f(x) &= \begin{cases} 0 & , \text{ for } x < k, \\ x - k & , \text{ for } k \leq x < v, \\ v - k & , \text{ for } x \geq v, \end{cases} \\ &= \min\{(x-k)_+, v-k\}. \end{aligned} \tag{5.18}$$

For the case $0 \leq v \leq k$, we have $v - k \leq 0$, which means that the pointwise smallest $f \in \mathcal{C}^{(2)}$ satisfying $f(v) \geq v - k$ is

$$f(x) = 0, \text{ for all } x. \tag{5.19}$$

Finally, if we let $\ell = (v-k)_+$, reassembling (5.18) and (5.19) establishes the result. □

5.4.2 Optimal Contingent Reinsurance Cover

In the following theorem, we derive the optimal reinsurance cover when contingent covers are allowed.

Theorem 5.4.1. *Let π be a monotone premium principle and k the maximal acceptable level of riskiness satisfying $k \geq k_{\inf}^{(1)} = \inf\big\{k : h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0\big\}$. Then $f_Z^{(1)}(\,\cdot\,; k) = f_Z^{cbsl}\big(\,\cdot\,; d_Z(k), \ell_Z(k)\big)$, i.e. a contingent bounded stop-loss cover is optimal, with contingent deductible $d_Z(k) = h_Z^{-1}\big(k; P_D^{(1)}(k)\big)$ and contingent limit $\ell_Z(k) = \big(\rho_Z[X] - d_Z(k)\big)_+$.*

Proof. By Proposition 5.4.2, we know that for $k \geq k_{\inf}^{(1)}$, the solution $f_Z^{(1)}(\,\cdot\,; k)$ is defined and thus belongs to $\in \mathcal{C}^{(1)}(k)$, meaning that

$$\rho_z[H_D^{(1)}(k)] \leq k, \text{ for all } z. \tag{5.20}$$

but also

$$0 \leq f_z^{(1)}(x; k) \leq x, \text{ for all } z, \tag{5.21}$$

and

$$f_z^{(1)}(x; k) \text{ and } x - f_z^{(1)}(x; k) \text{ are increasing functions, for all } z. \tag{5.22}$$

From (5.3) and (5.4), each inequality $\rho_z[H_D^{(1)}(k)] \leq k$ can be rewritten as

$$\rho_z[X - f_z^{(1)}(X; k)] \leq h_z^{-1}\big(k; P_D^{(1)}(k)\big),$$

which, since (5.3) and (5.22) yield $\rho_z[X - f_z^{(1)}(X; k)] = \rho_z[X] - f_z^{(1)}(\rho_z[X]; k)$, means that (5.20) is equivalent to

$$f_z^{(1)}(\rho_z[X]; k) \geq \rho_z[X] - h_z^{-1}\big(k; P_D^{(1)}(k)\big), \text{ for all } z. \tag{5.23}$$

By reassembling (5.21) for $x = \rho_z[X]$ with (5.23), we obtain

$$\rho_z[X] \geq f_z^{(1)}(\rho_z[X]; k) \geq \rho_z[X] - h_z^{-1}\big(k; P_D^{(1)}(k)\big), \text{ for all } z,$$

from what we deduce that since $f_Z^{(1)}(\,\cdot\,; k)$ is defined, we have

$$h_z^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0, \text{ for all } z. \tag{5.24}$$

Given (5.6), for the inequalities in (5.24) to be satisfied, it is necessary and sufficient that

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0. \tag{5.25}$$

Moreover, from (5.4), (5.5) and Proposition 5.4.2, we have that

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(i)}(k)\big) \text{ is strictly increasing in } k, \text{ for } i = 1, 2, \tag{5.26}$$

and therefore (5.25) is equivalent to $k \geq \inf\big\{k : h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0\big\}$, which proves that

$$k_{\inf}^{(1)} = \inf\big\{k : h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0\big\}.$$

Subsequently, since X is a positive random variable, we have

$$\rho_z[X] \geq 0, \text{ for all } z. \tag{5.27}$$

If we now define $f_z^{cbsl}\big(x; d_z(k), \ell_z(k)\big) = \min\big\{\big(x - d_z(k)\big)_+, \ell_z(k)\big\}$, where $d_z(k) = h_z^{-1}\big(k; P_D^{(1)}(k)\big)$ and $\ell_z(k) = \big(\rho_z[X] - d_z(k)\big)_+$, then from Lemma 5.4.1 and given (5.24) and (5.27), we have that $f_z^{cbsl}\big(\,\cdot\,; d_z(k), \ell_z(k)\big)$ is, for all z, the pointwise smallest f_z satisfying (5.21), (5.22) and (5.23). As a result,

$$f_Z^{cbsl}\big(\,\cdot\,; d_Z(k), \ell_Z(k)\big) = \sum_{z \in \mathcal{Z}} f_z^{cbsl}\big(\,\cdot\,; d_z(k), \ell_z(k)\big) \cdot \mathbb{1}\{Z = z\}$$

is the pointwise smallest contingent ceded loss function in $\mathcal{C}^{(1)}(k)$, which, from Property 5.3.1 and since π is a monotone premium principle, proves that $f_Z^{(1)}(\,\cdot\,; k) = f_Z^{cbsl}\Big(\,\cdot\,; d_Z(k), \ell_Z(k)\Big)$ for $k \geq k_{\inf}^{(1)}$. □

Several observations concerning the optimal contingent ceded loss function are in order. Firstly, for each scenario z, the deductible $d_z(k)$ corresponds to the threshold amount at which the insurance loss X yields a retained risk of k for the insurer. Since $d_z(k) = h_z^{-1}(k; P_D^{(1)}(k))$, from (5.4) we have $d_z(k) \leq k$ for all z. Moreover, if $k \leq P_D^{(1)}(k)$, or if for a particular scenario z there is no financial distress cost ($g_z(x) = 0$ for all x), then we have $d_z(k) = k$. Subsequently, if $d_z(k) < \rho_z[X]$, then the related limit $\ell_z(k)$ is strictly positive ($\ell_z(k) > 0$), whereas whenever $d_z(k) \geq \rho_z[X]$ one has no risk transfer for the scenario z ($\ell_z(k) = 0$).

Secondly, given property (5.6), the deductible $d_z(k)$ decreases in z, and thus, as $\rho_z[X]$ is assumed to increase in z, the limit $\ell_z(k)$ also increases in z. From this we get

$$f_z^{(1)}(\,\cdot\,;k) \preceq f_{\tilde{z}}^{(1)}(\,\cdot\,;k), \text{ whenever } z \leq \tilde{z}, \tag{5.28}$$

meaning that the worse the relative performance of the insurer is, the more extended his optimal reinsurance cover will be.

Thirdly, under $f_Z^{(1)}(\,\cdot\,;k)$, the retained risk of the insurer has, for each scenario z, the conditional distribution

$$\begin{aligned} F_{H_D^{(1)}(k)|Z=z}(x) &= \mathbb{P}\big(H_D^{(1)}(k) \leq x \;\big|\; Z=z\big) \\ &= \mathbb{P}\Big(X - f_z^{(1)}(X;k) \leq h_z^{-1}\big(x; P_D^{(1)}(k)\big) \;\Big|\; Z=z\Big) \\ &= \begin{cases} F_{X|Z=z}\Big(h_z^{-1}\big(x; P_D^{(1)}(k)\big)\Big) & , \text{ if } x < k, \\ F_{X|Z=z}\Big(h_z^{-1}\big(x; P_D^{(1)}(k)\big) + \ell_z(k)\Big) & , \text{ otherwise.} \end{cases} \end{aligned}$$

At $x = k$, this conditional distribution amounts to

$$\begin{aligned} F_{H_D^{(1)}(k)|Z=z}(k) &= F_{X|Z=z}\Big(h_z^{-1}\big(k; P_D^{(1)}(k)\big) + \ell_z(k)\Big) \\ &= F_{X|Z=z}\big(d_z(k) + \ell_z(k)\big) \\ &= F_{X|Z=z}\Big(d_z(k) + \big(\rho_z[X] - d_z(k)\big)_+\Big) \\ &\begin{cases} = 1-\alpha & , \text{ if } \ell_z(k) > 0, \\ \geq 1-\alpha & , \text{ otherwise,} \end{cases} \end{aligned} \tag{5.29}$$

and has the probability mass

$$F_{X|Z=z}\big(d_z(k) + \ell_z(k)\big) - F_{X|Z=z}\big(d_z(k)\big).$$

The unconditional distribution function of the retained risk of the insurer is then

$$\begin{aligned} F_{H_D^{(1)}(k)}(x) &= \mathbb{P}\big(H_D^{(1)}(k) \leq x\big) \\ &= \sum_{z\in\mathcal{Z}} F_{H_D^{(1)}(k)|Z=z}(x) \cdot \mathbb{P}(Z=z), \end{aligned}$$

which, at $x = k$, amounts to

$$F_{H_D^{(1)}(k)}(k) \quad \begin{cases} = 1-\alpha & , \text{ if } \ell_z(k) > 0 \text{ for all } z, \\ \geq 1-\alpha & , \text{ otherwise} \end{cases}$$

and has the probability mass

$$\sum_{z\in\mathcal{Z}} \Big(F_{X|Z=z}\big(d_z(k) + \ell_z(k)\big) - F_{X|Z=z}\big(d_z(k)\big)\Big) \cdot \mathbb{P}(Z=z).$$

Finally regarding the conditional risk measures, from (5.29) we have

$$\rho_z[H_D^{(1)}(k)] \quad \begin{cases} = k & \text{, if } \ell_z(k) > 0, \\ \leq k & \text{, otherwise.} \end{cases}$$

Hence, for each scenario z for which reinsurance is required ($\ell_z(k) > 0$), the optimal contingent bounded stop-loss only just satisfies the related constraint on the conditional risk ($\rho_z[H_D^{(1)}(k)] = k$).

5.4.3 Optimal Benchmark Traditional Reinsurance Cover

In the following theorem, we derive the optimal reinsurance cover when contingent covers are not allowed.

Theorem 5.4.2. *Let π be a monotone premium principle and k be the maximal acceptable level of riskiness satisfying $k \geq k_{\inf}^{(2)} = \inf\big\{k : h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big) \geq 0\big\}$. Then $f^{(2)}(\,\cdot\,; k) = f^{tbsl}\big(\,\cdot\,; d(k), \ell(k)\big)$, i.e. a traditional bounded stop-loss cover is optimal among all traditional covers, with deductible $d(k) = h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big)$ and limit $\ell(k) = \big(\rho_{\sup \mathcal{Z}}[X] - d(k)\big)_+$.*

Proof. By Proposition 5.4.2, we know that for $k \geq k_{\inf}^{(2)}$, the solution $f^{(2)}(\,\cdot\,; k)$ is defined and thus belongs to $\in \mathcal{C}^{(2)}(k)$, so that

$$\rho_z[H_D^{(2)}(k)] \leq k, \text{ for all } z, \tag{5.30}$$

but also

$$0 \leq f^{(2)}(x; k) \leq x, \tag{5.31}$$

and

$$f^{(2)}(x; k) \text{ and } x - f^{(2)}(x; k) \text{ are increasing functions.} \tag{5.32}$$

Following the same steps as in the proof of Theorem 5.4.1, we obtain that $\rho_z[H_D^{(2)}(k)] \leq k$ can be rewritten as

$$\rho_z[X] - f^{(2)}(\rho_z[X]) \leq h_z^{-1}\big(k; P_D^{(2)}(k)\big).$$

On the one hand, given (A2) and (5.32), the left-hand side of that last inequality increases in z. On the other hand, from (5.6) its right-hand side decreases in z. As a result, in order to fulfil (5.30), it is necessary and sufficient that $f^{(2)}(\,\cdot\,; k)$ satisfies

$$f^{(2)}(\rho_{\sup \mathcal{Z}}[X]) \geq \rho_{\sup \mathcal{Z}}[X] - h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big). \tag{5.33}$$

By reassembling (5.31) for $x = \rho_{\sup \mathcal{Z}}[X]$ and $z = \sup \mathcal{Z}$ with (5.33), we obtain

$$\rho_{\sup \mathcal{Z}}[X] \geq f^{(2)}(\rho_{\sup \mathcal{Z}}[X]; k) \geq \rho_{\sup \mathcal{Z}}[X] - h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big),$$

from which we deduce

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big) \geq 0. \tag{5.34}$$

With this, by applying the same reasoning as in the proof of Theorem 5.4.1, we can prove that

$$k_{\inf}^{(2)} = \inf\left\{k : h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big)\right\}. \tag{5.35}$$

If we now define $f^{tbsl}\big(x; d(k), \ell(k)\big) = \min\left\{\big(x - d(k)\big)_+, \ell(k)\right\}$, where $d(k) = h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big)$ and $\ell(k) = \big(\rho_{\sup \mathcal{Z}}[X] - d(k)\big)_+$, then from Lemma 5.4.1 and given (5.27) and (5.34), we have that $f^{tbsl}\big(\,\cdot\,; d(k), \ell(k)\big)$ is the pointwise smallest f satisfying (5.31), (5.32) and (5.33), and hence it is the pointwise smallest ceded loss function in $\mathcal{C}^{(2)}(k)$, which by Property 5.3.1 for a monotone premium principle proves the result. □

Remark 5.4.1. In their Theorem 3.2, Chi and Tan [21] proved the traditional bounded stop-loss to solve the problem of minimizing $\text{VaR}_\beta[X - f(X)] + \pi[f(X)]$, when the set of admissible ceded loss functions is $\mathcal{C}^{(2)}$ and reinsurance is priced according to the expected value premium principle. As outlined in Remark 5.3.4, that problem is similar in spirit to both Problems (5.1) and (5.2) of the present paper, and the appearance of bounded stop-loss structures in our optimal solutions is therefore intuitive.

We conclude this section with the following observations concerning the optimal benchmark cover:
Firstly, the deductible $d(k)$ corresponds to the amount at which the insurance loss X yields a retained risk of k for the insurer, given that he has the relative performance $Z = \sup \mathcal{Z}$. Since $d(k) = h_{\sup \mathcal{Z}}^{-1}(k; P_D^{(2)}(k))$, from (5.4) we have $d(k) \leq k$, and if $k \leq P_D^{(2)}(k)$, or if there is no financial distress cost ($g_z(x) = 0$ for all x and z), then we have $d(k) = k$. Subsequently, if the deductible $d(k)$ satisfies $d(k) < \rho_{\sup \mathcal{Z}}[X]$, then the limit $\ell(k)$ is strictly positive ($\ell(k) > 0$). On the other hand, if $d(k) \geq \rho_{\sup \mathcal{Z}}[X]$, then the benchmark traditional bounded stop-loss implies no risk transfer ($\ell(k) = 0$).

Secondly, under $f^{(2)}(\,\cdot\,; k)$, the retained risk of the insurer has, for each scenario z, the conditional distribution

$$\begin{aligned}
F_{H_D^{(2)}(k)|Z=z}(x) &= \mathbb{P}\big(H_D^{(2)}(k) \leq x \mid Z = z\big) \\
&= \mathbb{P}\Big(X - f^{(2)}(X; k) \leq h_z^{-1}\big(x; P_D^{(2)}(k)\big) \;\Big|\; Z = z\Big) \\
&= \begin{cases} F_{X|Z=z}\Big(h_z^{-1}\big(x; P_D^{(2)}(k)\big)\Big) & , \text{ if } x < h_z\big(d(k); P_D^{(2)}(k)\big), \\ F_{X|Z=z}\Big(h_z^{-1}\big(x; P_D^{(2)}(k)\big) + \ell(k)\Big) & , \text{ if } x \geq h_z\big(d(k); P_D^{(2)}(k)\big), \end{cases}
\end{aligned}$$

where $h_{\sup \mathcal{Z}}\big(d(k); P_D^{(2)}(k)\big) = k$ and hence

$$h_z\big(d(k); P_D^{(2)}(k)\big) \leq h_{\tilde{z}}\big(d(k); P_D^{(2)}(k)\big) \leq k, \text{ whenever } z \leq \tilde{z} \leq \sup \mathcal{Z},$$

from (5.6).

At $x = h_z\big(d(k); P_D^{(2)}(k)\big)$, this conditional distribution amounts to

$$\begin{aligned} F_{H_D^{(2)}(k)|Z=z}\Big(h_z\big(d(k); P_D^{(2)}(k)\big)\Big) &= F_{X|Z=z}\big(d(k)+\ell(k)\big) \\ &= F_{X|Z=z}\Big(d(k)+\big(\rho_z[X]-d(k)\big)_+\Big) \\ &\begin{cases} = 1-\alpha & , \text{ if } \ell(k)>0 \text{ and } z=\sup\mathcal{Z}, \\ \geq 1-\alpha & , \text{ otherwise,} \end{cases} \end{aligned} \tag{5.36}$$

and has the probability mass

$$F_{X|Z=z}\big(d(k)+\ell(k)\big) - F_{X|Z=z}\big(d(k)\big).$$

The unconditional distribution function of the retained risk of the insurer is then

$$\begin{aligned} F_{H_D^{(2)}(k)}(x) &= \mathbb{P}\big(H_D^{(2)}(k) \leq x\big) \\ &= \sum_{z\in\mathcal{Z}} F_{H_D^{(2)}(k)|Z}(x|z)\cdot\mathbb{P}(Z=z), \end{aligned}$$

and has, at each

$$x \in \big\{h_z\big(d(k); P_D^{(2)}(k)\big),\ z\in\mathcal{Z}\big\},$$

the probability masses of at least

$$\Big\{\Big(F_{X|Z=z}\big(d(k)+\ell(k)\big) - F_{X|Z=z}\big(d(k)\big)\Big)\cdot\mathbb{P}(Z=z),\ z\in\mathcal{Z}\Big\}.$$

Finally regarding the conditional risk measures, from (5.36) we have

$$\rho_z[H_D^{(2)}(k)] \quad \begin{cases} = k & , \text{ if } \ell(k)>0 \text{ and } z=\sup\mathcal{Z}, \\ \leq k & , \text{ otherwise,} \end{cases} \tag{5.37}$$

meaning that when reinsurance is required ($\ell(k)>0$), the benchmark cover only just satisfies the constraint on the conditional risk measure for the worst-case scenario ($\rho_{\sup\mathcal{Z}}[H_D^{(2)}(k)] = k$).

5.4.4 Comparison

We now compare the optimal contingent bounded stop-loss and its benchmark, when the same monotone premium principle π and maximal acceptable level of riskiness k apply. We assume the latter to be such that the solutions $f_Z^{(1)}(\,\cdot\,;k)$ and $f^{(2)}(\,\cdot\,;k)$ are both defined.

We start by showing that an optimal contingent cover always leads to a smaller reinsurance premium than an optimal traditional cover:

Proposition 5.4.3. *Let π be a monotone premium principle and k the maximal acceptable level of riskiness. Then $k_{\inf}^{(1)} \leq k_{\inf}^{(2)}$ and for all $k \geq k_{\inf}^{(2)}$*

$$P_R^{(1)}(k) \leq P_R^{(2)}(k). \tag{5.38}$$

Proof. By Proposition 5.4.2, problems (5.1) and (5.2) both admit a solution for $k \geq \max\{k_{\min}^{(1)}, k_{\min}^{(2)}\}$. As $\mathcal{C}^{(2)}(k) \subseteq \mathcal{C}^{(1)}(k)$, problem (5.2) is hence just a constrained version of problem (5.1), so that

$$P_R^{(1)}(k) \leq P_R^{(2)}(k), \text{ for } k \geq \max\{k_{\inf}^{(1)}, k_{\inf}^{(2)}\}. \tag{5.39}$$

To complete the proof, it only remains to show that $k_{\inf}^{(1)} \leq k_{\inf}^{(2)}$. By the definitions of $k_{\inf}^{(1)}$ and $k_{\inf}^{(2)}$ and from (5.26), we have

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq 0, \text{ for } k \geq k_{\inf}^{(1)},$$

and

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big) \geq 0, \text{ for } k \geq k_{\inf}^{(2)},$$

and hence (5.5) and (5.39) yield

$$h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(1)}(k)\big) \geq h_{\sup \mathcal{Z}}^{-1}\big(k; P_D^{(2)}(k)\big) \geq 0, \text{ for } k \geq \max\{k_{\inf}^{(1)}, k_{\inf}^{(2)}\}. \tag{5.40}$$

Let us now assume that $k_{\inf}^{(1)} > k_{\inf}^{(2)}$. In that case, since

$$h_{\sup \mathcal{Z}}\big(k_{\inf}^{(1)}; P_D^{(1)}(k_{\inf}^{(1)})\big) = 0 \text{ and } h_{\sup \mathcal{Z}}\big(k_{\inf}^{(2)}; P_D^{(2)}(k_{\inf}^{(2)})\big) = 0,$$

from (5.26) and (5.40) we would have

$$0 \geq h_{\sup \mathcal{Z}}\big(k_{\inf}^{(1)}; P_D^{(2)}(k_{\inf}^{(1)})\big) > 0,$$

which is not possible, so that necessarily $k_{\inf}^{(1)} \leq k_{\inf}^{(2)}$. □

Remark 5.4.2. Proposition 5.4.3 proves that for any given maximal acceptable level of riskiness $k \geq k_{\inf}^{(2)}$, the optimal contingent bounded stop-loss is worth a smaller reinsurance premium than its benchmark, which yields a larger retained premium and hence a greater risk bearing capacity for the insurer. Under the optimal contingent bounded stop-loss, the insurer can thus typically afford to choose a larger maximal acceptable level of riskiness, which, according to Proposition 5.4.2, will make that reinsurance cover even cheaper than its benchmark. However, since the choice of the maximal acceptable level of riskiness is not dealt with in this paper (cf. Remark 5.3.1), we will pursue our analysis by comparing the optimal contingent cover and its benchmark under the same k.

With this result, we are now able to compare in more detail $f_Z^{(1)}(\,\cdot\,; k)$ and $f^{(2)}(\,\cdot\,; k)$. Indeed, since (5.38) yields $P_D^{(1)}(k) \geq P_D^{(2)}(k)$, given (5.5) we have $d_{\sup \mathcal{Z}}(k) \geq d(k)$ and thus $\ell_{\sup \mathcal{Z}}(k) \leq \ell(k)$, which results in

$$f_{\sup \mathcal{Z}}^{(1)}(\,\cdot\,; k) \preceq f^{(2)}(\,\cdot\,; k) \tag{5.41}$$

and hence

$$f_Z^{(1)}(\,\cdot\,;k) \preceq f^{(2)}(\,\cdot\,;k), \tag{5.42}$$

from (5.28). That last relationship implies by definition

$$f_Z^{(1)}(X;k) \leq f^{(2)}(X;k), \text{ for any realization of } X \text{ and } Z, \tag{5.43}$$

which can be seen as the counterpart of (5.38). That is, while the optimal contingent bounded stop-loss is cheaper than its benchmark, it also yields a smaller ceded loss.

Remark 5.4.3. We can distinguish two factors responsible for the difference between $P_R^{(1)}(k)$ and $P_R^{(2)}(k)$. On the one hand, by definition the optimal contingent bounded stop-loss varies with Z, which makes it being adapted to the need of the insurer (in terms of reduction of the conditional risk measures) for each scenario. In contrast, the benchmark cover cannot vary with Z and, as shown by (5.37), it fits the need of the insurer only for the worst-case scenario, leaving him over-reinsured for the other scenarios, which contributes to a higher premium. On the other hand, the optimal contingent bounded stop-loss cover results in a larger retained premium than its benchmark, which for the worst-case scenario yields a smaller financial distress cost. As a result, while one could expect the ceded loss functions $f_{\sup \mathcal{Z}}^{(1)}(\,\cdot\,;k)$ and $f^{(2)}(\,\cdot\,;k)$ to be equal (they both serve the insurer's need under the worst-case scenario), their order relationship is given by (5.41), and the latter contributes to making the benchmark cover being more expensive than the optimal contingent bounded stop-loss.

At this point, it is interesting to note that whereas (5.43) results in

$$X - f_Z^{(1)}(X;k) \geq X - f^{(2)}(X;k), \text{ for any realization of } X \text{ and } Z,$$

the optimal contingent bounded stop-loss does not necessarily yield a larger retained risk for the insurer. Indeed, due to (5.38), it can happen that the financial distress cost

$$Y_D^{(2)}(k) = g_Z\Big(\big(X - f^{(2)}(X;k) - P_D^{(2)}(k)\big)_+\Big)$$

exceeds

$$Y_D^{(1)}(k) = g_Z\Big(\big(X - f_Z^{(1)}(X;k) - P_D^{(1)}(k)\big)_+\Big),$$

for some X and Z, in which case the difference $Y_D^{(2)}(k) - Y_D^{(1)}(k)$ may be large enough to result in $H_D^{(1)}(k) < H_D^{(2)}(k)$. However, when that happens, the difference $H_D^{(2)}(k) - H_D^{(1)}(k)$ will be non-negligible only if the retained risk of the insurer is dominated by the financial distress cost.

In any case, by the design of problems (5.1) and (5.2), the optimal contingent bounded stop-loss and its benchmark both bring all the conditional risk measures to at most the maximal acceptable level of riskiness k and hence, from this viewpoint, they mitigate the risk equivalently. However, the contingent cover does it for a smaller reinsurance premium and thus leaves more potential profits for the insurer. In Section 5.5, we will quantify the difference between $P_R^{(1)}(k)$ and $P_R^{(2)}(k)$ in a concrete example.

Remark 5.4.4. In the light of Remark 5.4.3, it is intuitive that the main factors determining how small $P_R^{(1)}(k)$ will be relative to $P_R^{(2)}(k)$ are, on the one hand, how over-reinsured the insurer will be for the scenarios different from $\sup \mathcal{Z}$, and on the other hand, how much weight is given to these scenarios. If reinsurance is priced according to the expected value premium principle (as it will be in the concrete example from Section 5.5), the difference between $P_R^{(2)}(k)$ and $P_R^{(1)}(k)$ can be expressed as

$$P_R^{(2)}(k) - P_R^{(1)}(k) = (1+\theta) \cdot \sum_{z \in \mathcal{Z}} \Delta_z(k),$$

where $\Delta_z(k) = \big(\mathbb{E}[f^{(2)}(X;k)|Z = z] - \mathbb{E}[f_z^{(1)}(X;k)|Z = z]\big) \cdot \mathbb{P}(Z = z)$. Correspondingly, this difference will be substantial if $\Delta_z(k)$ is large for some z (which occurs when $\rho_z[X]$ is significantly smaller than $\rho_{\sup \mathcal{Z}}[X]$, as that increases the difference between $\ell_z(k)$ and $\ell(k)$ and subsequently between $f_z^{(1)}(\,\cdot\,;k)$ and $f^{(2)}(\,\cdot\,;k)$) and at the same time $\mathbb{P}(Z = z)$ is large.

Finally, from the reinsurer's perspective, while (5.38) means that offering $f_Z^{(1)}(\,\cdot\,;k)$ instead of $f^{(2)}(\,\cdot\,;k)$ yields less potential profits, it also yields less risk, cf. (5.43). Whether selling contingent covers instead of traditional ones will improve the risk-to-profit of the reinsurer will hence depend on the concrete situation, particularly on the degree of negative dependence between the relative performance of each insurer. However, we will show in the next section that in several realistic cases, selling contingent covers can indeed improve the risk-to-profit of the reinsurer.

Remark 5.4.5. Note that in the absence of financial distress costs ($g_z(x) = 0$ and $h_z(x;p) = h_z^{-1}(x;p) = x$ for all p, x and z), there is in fact no need for Z to still model the relative performance of the insurer. Any other contingent cover based on an external (discrete) Z can then also be considered, with the above results still being applicable, as long as one assures (A2) to hold.

5.5 Numerical Illustration

In this section we will consider a concrete numerical illustration in detail. We assume an insurance market with $n = 3$ (and later $n = 5$) insurers. They are all assumed to be identical in distribution.

5.5.1 Concrete Model Specifications

For the marginal distribution of X_i ($i = 1, \ldots, n$) representing the aggregate loss of insurer i, we consider the following composite (splicing) model (see e.g. Scollnik [73]) with density function

$$b(x) = \lambda \cdot \varphi(x;\mu,\sigma,s) + (1-\lambda) \cdot \nu(x;\alpha,s),$$

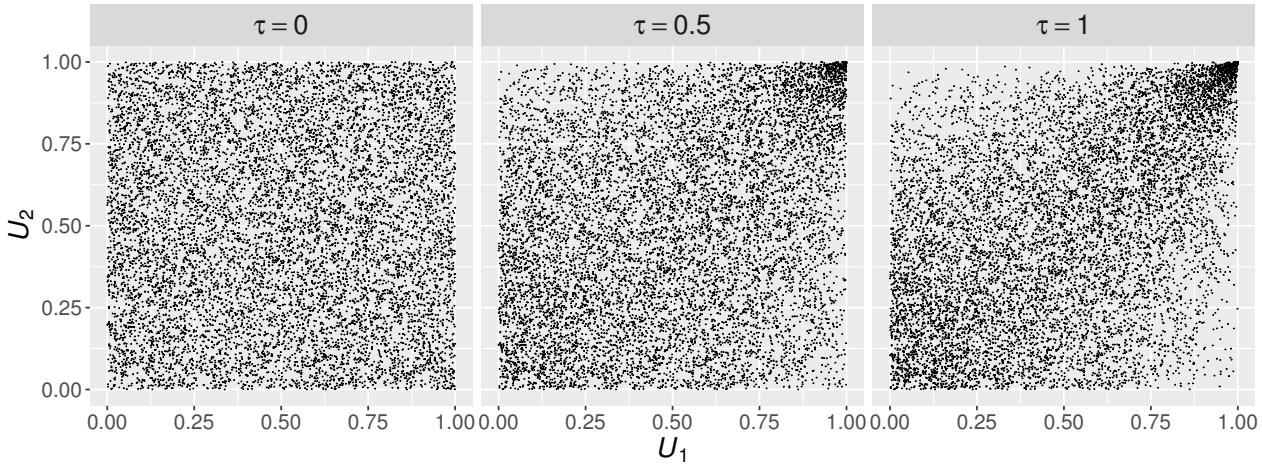

Figure 5.1: Random samples of the random vector (U_1, U_2), with Uniform$[0, 1]$ marginals and F-Clayton(τ) copula, for $\tau = 0, 0.5, 1$. On these plots, each random sample contains 10^4 realizations.

where

$$\varphi(x; \mu, \sigma, s) = \frac{\frac{1}{x} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{\ln x - \mu}{\sigma} \right)^2 \right)}{\int_0^s \frac{1}{y} \cdot \exp\left(-\frac{1}{2} \cdot \left(\frac{\ln y - \mu}{\sigma} \right)^2 \right) \mathrm{d}y} \cdot \mathbb{1}\{0 < x \leq s\},$$

is the density function of a Log-Normal(μ, σ) random variable, upper-truncated at $s > 0$, and

$$\nu(x; \alpha, s) = \alpha \cdot \frac{s^\alpha}{x^{\alpha+1}} \cdot \mathbb{1}\{x > s\},$$

the density function of a Pareto(α,s) random variable. This density function b allows to model the loss with a Log-Normal bulk of the distribution and a Pareto tail, which is often considered to be realistic. For the parameters, we set $\alpha = 2.2$ and $s = 1'800$, and then choose λ, μ and σ so that $\mathbb{E}[X_i] = 1'000$ and b is continuous and differentiable at $x = s$, which yields $\lambda \approx 0.9009$, $\mu \approx 6.5728$ and $\sigma \approx 0.6476$. The resulting standard deviation is $\sqrt{\text{Var}[X_i]} \approx 1'780$.

The dependence structure of the random vector $(X_1, X_2, ..., X_n)$ is modelled according to an Archimedean survival copula with generator $\phi(t) = t^{-1/\tau} - 1$, which is an n-dimensional Flipped-Clayton (or F-Clayton, for short) copula with parameter τ (see e.g. Nelsen [66]). In addition to interpolate between independence ($\tau = 0$) and comonotonicity ($\tau = \infty$), this copula allows for right-tail dependence and is therefore appropriate to our context, since it is not rare for reinsurers to face tail dependence when reinsuring losses. In the following applications, we will consider three dependence scenarios: mutual independence ($\tau = 0$), medium tail dependence ($\tau = 0.5$) and strong tail dependence ($\tau = 1$). Figure 5.1 illustrates random samples drawn from a bivariate F-Clayton(τ) copula for these three cases.

As the above specifications do not allow an explicit expression of the joint distribution function of $(X_1, ..., X_n)$, we will consider the respective empirical joint distribution based on 10^7 sample points instead for all calculations.

For simplicity of notation, we now re-identify X_1 with X. The total premium P_X and the reinsurance premium P_R are both assumed to be computed according to the expected value principle (with safety loading $\theta_X = 0.2$ and $\theta_R = 0.5$, respectively). The total premium is thus $P_X = 1'200$ and the retained premium after reinsurance amounts to

$$P_D = 1'200 - 1.5 \cdot \mathbb{E}[R].$$

As mentioned in Section 5.3.2, the expected value principle is monotone and hence $f_Z^{cbsl}\big(\,\cdot\,; d_Z(k), \ell_Z(k)\big)$ and $f^{tbsl}\big(\,\cdot\,; d(k), \ell(k)\big)$ are indeed the solutions to problems (5.1) and (5.2), respectively.

In practice, a conditional distribution $F_{X|Z=z}$ and a financial distress cost function g_z must be estimated for all $z \in \mathcal{Z}$. If too many risk scenarios are considered in Z, this can be a difficult task and may significantly increase model risk. For this reason, we consider here the simple case of the binary measure of relative performance

$$Z = \mathbb{1}\{V > 1.5 \cdot \frac{n}{0.5 + n} \cdot \overline{V}_n\} = \mathbb{1}\{V > 1.5 \cdot \overline{V}_{n-1}\},$$

where $V_i = X_i/P_{X_i}$ is the loss ratio of insurer i (and hence $V = X/P_X$), $\overline{V}_n = \frac{1}{n} \cdot \sum_{i=1}^n V_i$ the average loss ratio of the market and $\overline{V}_{n-1} = \frac{1}{n-1} \cdot \sum_{i=2}^n V_i$ the average loss ratio of the insurer's $n-1$ competitors. The insurer is thus viewed as realizing a notably bad relative performance if his loss ratio is greater than 1.5 times the average loss ratio of his competitors. The resulting domain of Z is correspondingly $\mathcal{Z} = \{0, 1\}$. Notice that, since the insurers are assumed to be identical, they all receive the same total premium and hence the relative performance simplifies to

$$Z = \mathbb{1}\{X > 1.5 \cdot \overline{X}_{n-1}\},$$

where $\overline{X}_{n-1} = \frac{1}{n-1} \cdot \sum_{i=2}^n X_i$.

For the financial distress cost, we assume

$$Y_D = 0.5 \cdot Z \cdot (D - P_D)_+.$$

Therefore, if both a bad relative performance occur ($Z = 1$) and the retained loss exceeds the retained premium, then each two additional monetary units of retained loss result in one monetary unit of financial distress cost. The resulting financial distress cost function $g_z(x) = 0.5 \cdot z \cdot x$ is continuous in x and increasing in both x and z with $g_z(0) = 0$, in accordance with Assumptions (A1) and (A3). Also, we have $h_z(x; p) = x + 0.5 \cdot z \cdot (x - p)_+$, which yields that the optimal contingent bounded stop-loss has deductibles

$$d_z(k) = \begin{cases} k & , \text{ for } z = 0, \\ \frac{2}{3} \cdot k + \frac{1}{3} \cdot \min\{k, P_D^{(1)}(k)\} & , \text{ for } z = 1, \end{cases}$$

	$n=3$			$n=5$		
τ	0	0.5	1	0	0.5	1
$\mathbb{P}(Z=1)$	0.2713	0.2440	0.2103	0.2253	0.2066	0.1749
$\rho_0[H_X]$	2′238	4′339	5′814	2′092	4′227	5′782
$\rho_1[H_X]$	18′351	16′932	14′475	19′962	18′529	15′780

Table 5.1: Conditional risk measures $\rho_0[H_X]$ and $\rho_1[H_X]$ resulting from the model inputs, together with the the probability that the insurer incurs a bad relative performance $\mathbb{P}(Z=1)$, for $n=3,5$ and $\tau=0,0.5,1$.

and limits

$$\ell_z(k)=\begin{cases}\big(\rho_0[X]-d_0(k)\big)_+ & \text{, for } z=0,\\ \big(\rho_1[X]-d_1(k)\big)_+ & \text{, for } z=1,\end{cases}$$

while the benchmark has deductible

$$d(k)=\frac{2}{3}\cdot k+\frac{1}{3}\cdot\min\{k,P_D^{(2)}(k)\},$$

and limit

$$\ell(k)=\big(\rho_1[X]-d(k)\big)_+,$$

cf. Section 5.4. For the operators ρ_0 and ρ_1, we set $\beta=0.995$.

In order to illustrate the model described above, we show in Table 5.1 the values of the conditional risk measures $\rho_0[H_X]$ and $\rho_1[H_X]$, together with the probability $\mathbb{P}(Z=1)$, for $n=3,5$ and $\tau=0,0.5,1$. Note that in all cases $\rho_z[H_X]$ increases in z, in accordance with Assumption (A2).

Figure 5.2 depicts the plots of the conditional distribution functions $F_{H_X|Z=0}$ and $F_{H_X|Z=1}$, as well as the unconditional distribution function F_{H_X} for each choice of n and τ. We observe from Table 5.1 and Figure 5.2 that there is a significant difference between the conditional distributions $F_{H_X|Z=0}$ and $F_{H_X|Z=1}$ (and hence between their 99.5%-quantile $\rho_0[H_X]$ and $\rho_1[H_X]$), with the unconditional F_{H_X} being in between. The significant difference between $F_{H_X|Z=0}$ and $F_{H_X|Z=1}$ indicates that the measure Z of the relative performance distinguishes two risk scenarios in which the risk faced by the insurer (and hence his need for reinsurance) is clearly distinct, which makes Z being an appropriate candidate to be used in a contingent reinsurance cover. The distance between the curves naturally decreases with τ and increases with n: The parameter τ models the strength of the right-tail dependence between the X_i's. For larger τ, the realizations of the X_i's

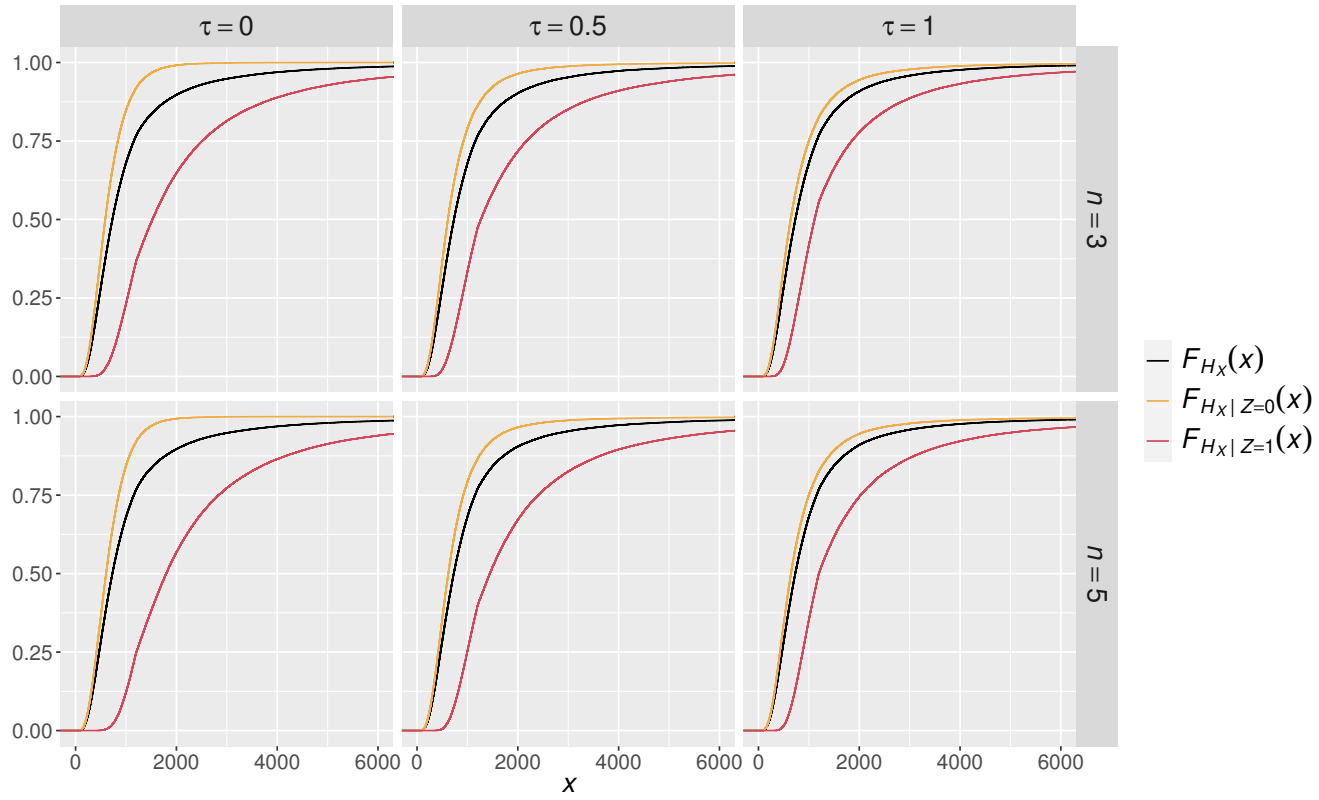

Figure 5.2: Conditional distribution functions $F_{H_X|Z=0}$ and $F_{H_X|Z=1}$, together with the unconditional distribution function F_{H_X}, for $n = 3, 5$ and $\tau = 0, 0.5, 1$.

will be closer to each other and hence less information on X will be carried by the events $Z = 0$ and $Z = 1$. At the same time, increasing the market size n lowers the variance of $\overline{X}_{n-1}$ and hence allows for potentially larger deviations of X from $1.5 \cdot \overline{X}_{n-1}$, which explains that the distance between $F_{H_X|Z=1}$ and F_{H_X} increases with n.

The same effects also drive the probability to experience a bad relative performance $\mathbb{P}(Z = 1)$ as given in Table 5.1. That probability decreases in both τ and n. For larger τ the realizations of the X_i's will be closer to each other, making the exceedance of X over $1.5 \cdot \overline{X}_{n-1}$ less likely. Since increasing n lowers the variance of $\overline{X}_{n-1}$, that also makes it less likely that $Z = 1$ occurs caused by a small realization of $\overline{X}_{n-1}$.

5.5.2 The insurer's viewpoint

In this section, we analyse the position of the insurer, when he purchases either the optimal contingent bounded stop-loss or its benchmark. The quantities to follow turn out to vary considerably with the given maximal acceptable level of riskiness. For the clarity of the plots, we thus consider the intermediate range of maximal acceptable levels of riskiness $k \in [1'200, 10'000]$. For $n = 3, 5$ and $\tau = 0, 0.5, 1$, the lower bound is greater than both $k_{\text{inf}}^{(1)}$ and $k_{\text{inf}}^{(2)}$, which ensures that both $f_Z^{(1)}(\,\cdot\,; k)$ and $f^{(2)}(\,\cdot\,; k)$ are defined.

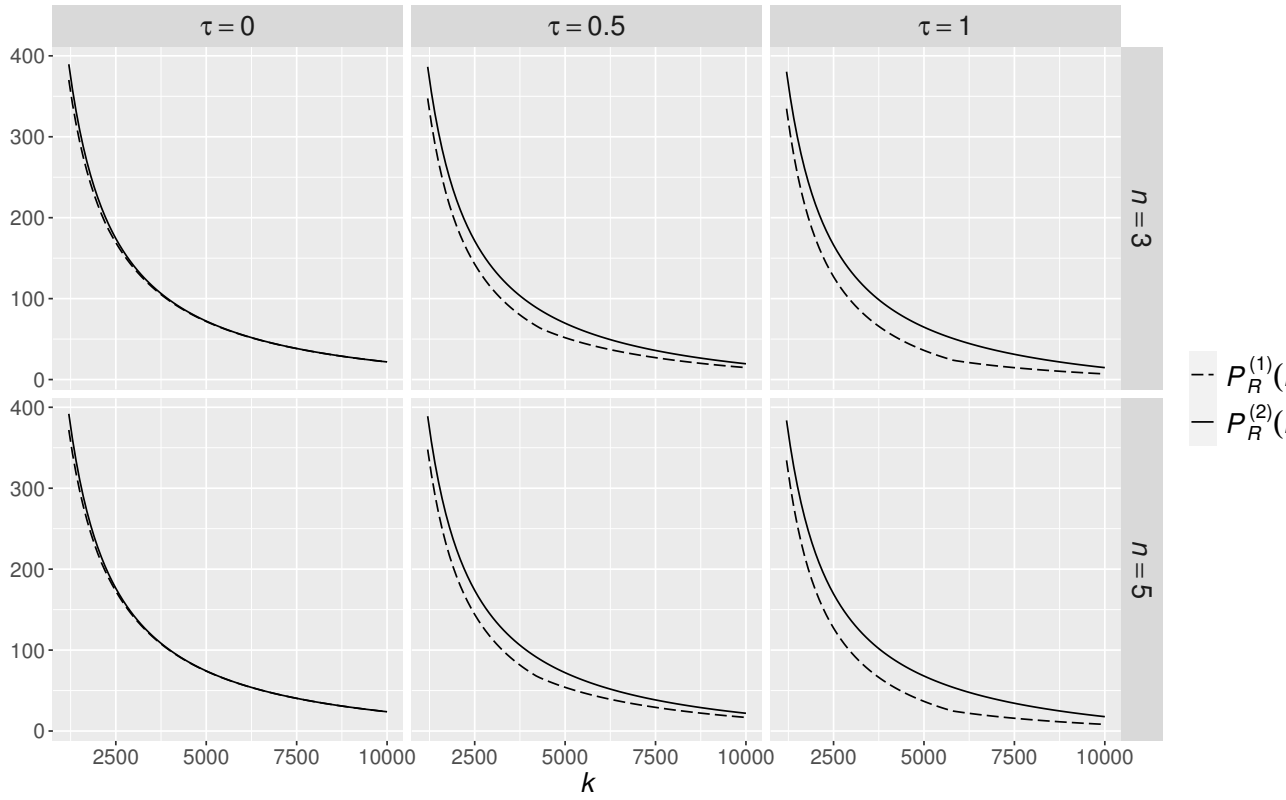

Figure 5.3: Reinsurance premiums, for $n = 3, 5$ and $\tau = 0, 0.5, 1$.

In Figure 5.3, we plot the reinsurance premiums $P_R^{(1)}(k)$ and $P_R^{(2)}(k)$ for $n = 3, 5$ and $\tau = 0, 0.5, 1$. In accordance with Proposition 5.4.3, it shows that for all the considered maximal acceptable levels of riskiness, the reinsurance premium for the optimal contingent bounded stop-loss is smaller than the one for its benchmark. If we then compare the difference between $P_R^{(1)}(k)$ and $P_R^{(2)}(k)$ for the various values of τ and n, we notice that it notably increases with τ. The reason for this is the following: By (5.28) and (5.43), for any given monotone premium principle π and maximal acceptable level of riskiness k we have

$$f_0^{(1)}(\,\cdot\,;k) \preceq f_1^{(1)}(\,\cdot\,;k) \preceq f^{(2)}(\,\cdot\,;k), \tag{5.44}$$

and the only difference between $f_1^{(1)}(\,\cdot\,;k)$ and $f^{(2)}(\,\cdot\,;k)$ is the retained premium involved in the related parameters $d_1(k)$, $\ell_1(k)$, $d(k)$ and $\ell(k)$. In the present example, for all the considered maximal acceptable levels of riskiness the distance between $f_1^{(1)}(\,\cdot\,;k)$ and $f^{(2)}(\,\cdot\,;k)$ is very small, and hence so is $\Delta_1(k)$. As a result, given that here $P_R^{(2)}(k) - P_R^{(1)}(k) = 1.5 \cdot \big(\Delta_0(k) + \Delta_1(k)\big)$, what prevails in the latter difference is

$$\Delta_0(k) = \underbrace{\big(\mathbb{E}[f^{(2)}(X;k)|Z=0] - \mathbb{E}[f_0^{(1)}(X;k)|Z=0]\big)}_{(a)} \cdot \underbrace{\mathbb{P}(Z=0)}_{(b)}.$$

where (a) quantifies how much the benchmark makes the insurer being over-reinsured with respect to the contingent cover and (b) is the related weight. As shown in Table 5.1,

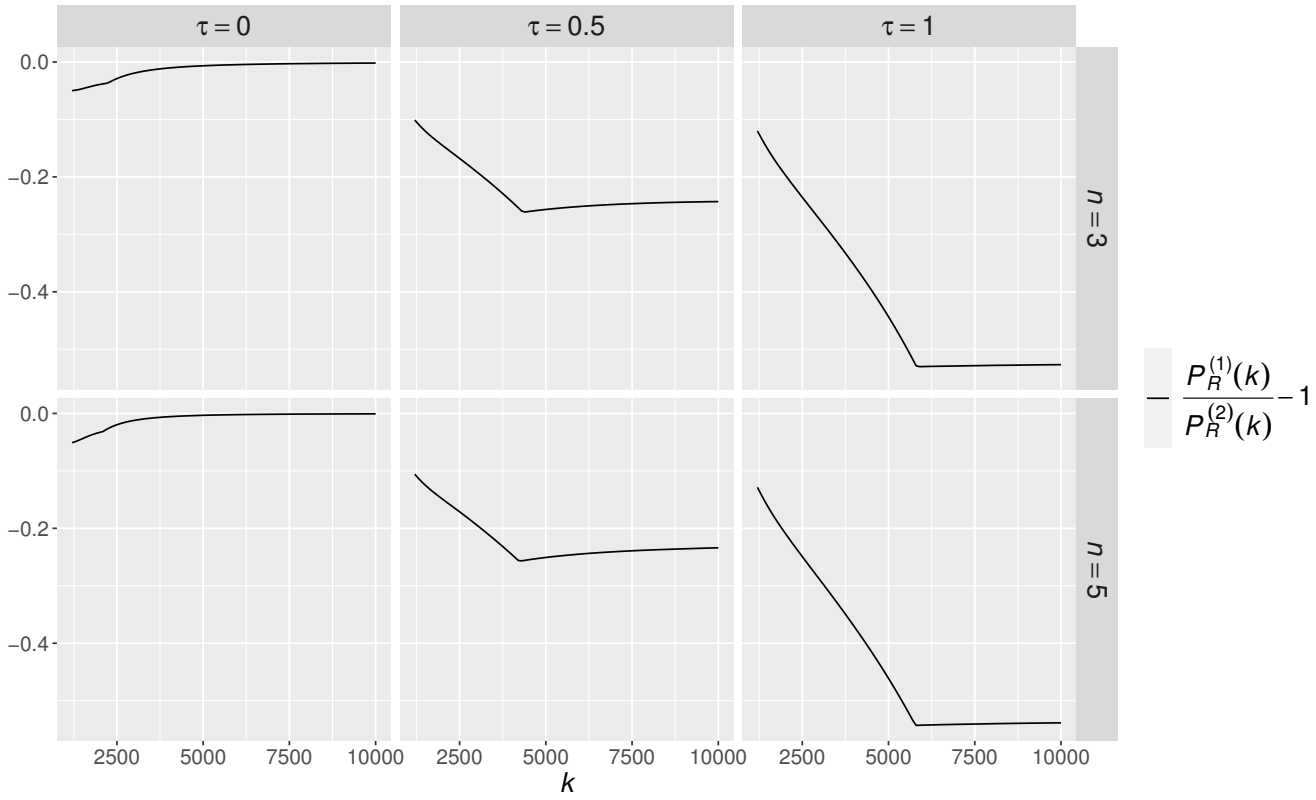

Figure 5.4: Relative difference of the reinsurance premiums, for $n = 3, 5$ and $\tau = 0, 0.5, 1$.

the difference between $\rho_0[X]$ and $\rho_1[X]$ decreases in τ, which leads to (a) being decreasing in τ (cf. Remark 5.4.4). On the other hand, the probability $\mathbb{P}(Z = 0) = 1 - \mathbb{P}(Z = 1)$ increases in τ (see Table 5.1 and the respective discussion above). While these two effects are conflicting, it turns out that the increase in τ of (b) dominates, which leads $P_R^{(2)}(k) - P_R^{(1)}(k)$ to increase with τ.

Since in absolute terms, the above curves are quite close to each other, it may be more instructive to consider the relative difference $P_R^{(1)}(k)/P_R^{(2)}(k) - 1$ instead, which is plotted in Figure 5.4 for $n = 3, 5$ and $\tau = 0, 0.5, 1$. The bend appearing in each plot occurs at $k = \rho_0[X]$, and its presence can be understood as follows. While the deductibles $d_0(k)$, $d_1(k)$ and $d(k)$ are all increasing in k, the limits $\ell_0(k)$, $\ell_1(k)$ and $\ell(k)$ are all decreasing in k, which makes the reinsurance premiums $P_R^{(1)}(k)$ and $P_R^{(2)}(k)$ both to be decreasing in k, as shown by Figure 5.3 and in accordance with Proposition 5.4.2. However, whereas for $k < \rho_0[X]$, the limits $\ell_0(k)$, $\ell_1(k)$ and $\ell(k)$ are all strictly decreasing in k, for $k \in [\rho_0[X], \rho_1[H_X])$, the limit $\ell_0(k)$ is constant at 0 and hence only $\ell_1(k)$ and $\ell(k)$ remain strictly decreasing. As a result, for $k \in [\rho_0[X], \rho_1[H_X])$ the reinsurance premium $P_R^{(1)}(k)$ decreases less in k than $P_R^{(2)}(k)$ does, resulting in that particular bend in the relative difference curve $P_R^{(1)}(k)/P_R^{(2)}(k) - 1$ at $k = \rho_0[X]$.

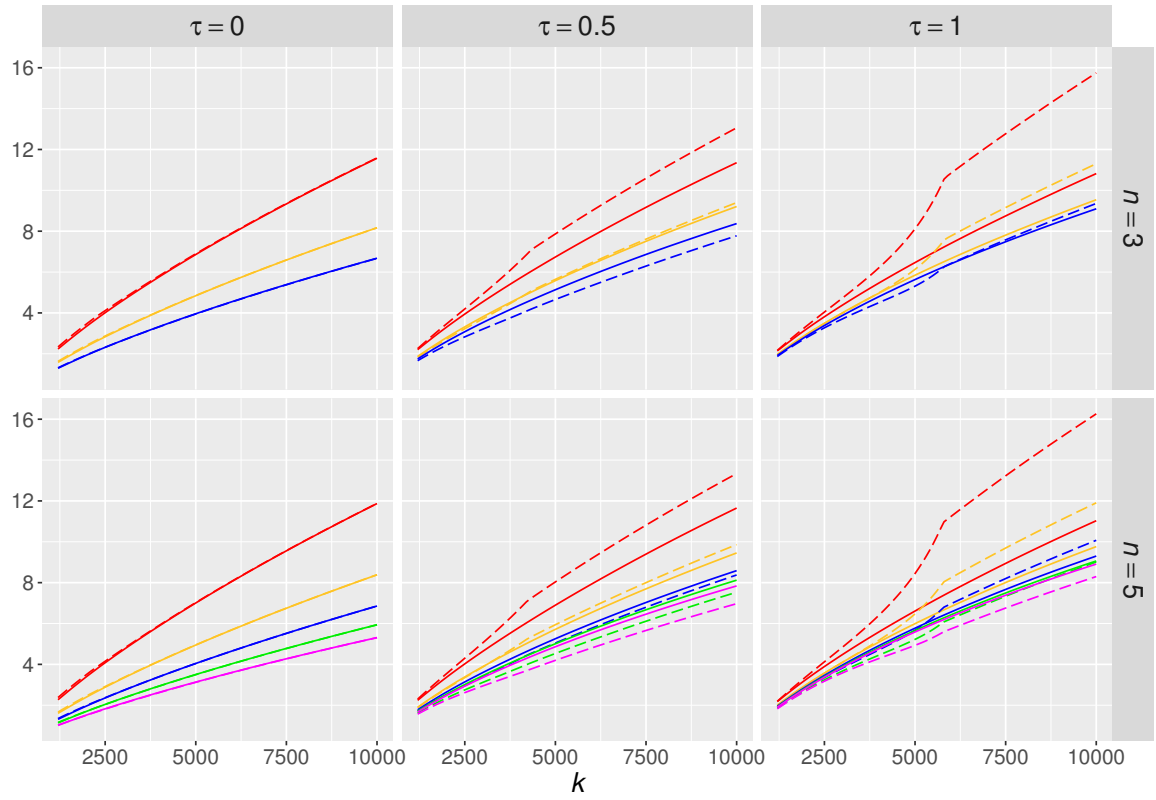

Figure 5.5: Standard-deviation of the reinsurer's loss ratio, for $n = 3, 5$ and $\tau = 0, 0.5, 1$.

Finally, Figure 5.3 already showed that the more likely it is for the worst-case scenario not to occur, the cheaper (and hence the more advantageous for the insurer) the optimal contingent bounded stop-loss will be in comparison to its benchmark. Figure 5.4 confirms it, showing that the relative difference can be about -25% for $\tau = 0.5$ and -55% for $\tau = 1$. That is, an increased degree of dependence among insurers is in fact advantageous for the performance of this cover. Also, we would like to emphasize that similar results could be obtained using other measures of relative performance, as long as they imply one or more large $\Delta_z(k)$ (cf. Remark 5.4.4).

5.5.3 The reinsurer's viewpoint

In order to assess whether selling contingent covers instead of traditional ones improves the risk-to-profit measure of the reinsurer or not, we will consider his loss ratio in two distinct (extreme) situations: either the reinsurance covers he sells are all of optimal contingent bounded stop-loss type, or they are all benchmark covers. The resulting reinsurer's loss ratios are then

$$W^{(1)}_{m:n}(k) = \frac{\sum_{i=1}^{m} f^{(1)}_{Z_i}(X_i;k)}{\sum_{i=1}^{m} P^{(1)}_{R_i}(k)} \quad \text{and} \quad W^{(2)}_{m:n}(k) = \frac{\sum_{i=1}^{m} f^{(2)}_{Z_i}(X_i;k)}{\sum_{i=1}^{m} P^{(2)}_{R_i}(k)},$$

where $m \in \{1, ..., n\}$ is the number of insurers that the reinsurer sold covers to and Z_i is the relative performance of the i^{th} insurer.

At $t = 0$, these two loss ratios are random variables which have, for any m and $k \geq k_{\text{inf}}^{(2)}$ the common expectation

$$\mathbb{E}[W_{m:n}^{(1)}(k)] = \mathbb{E}[W_{m:n}^{(2)}(k)] = 2/3. \tag{5.45}$$

In order to characterise the difference between $W_{m:n}^{(1)}(k)$ and $W_{m:n}^{(2)}(k)$, we will thus focus on their standard-deviation, denoted by

$$Q_{m:n}^{(1)}(k) = \sqrt{\text{Var}[W_{m:n}^{(1)}(k)]} \quad \text{and} \quad Q_{m:n}^{(2)}(k) = \sqrt{\text{Var}[W_{m:n}^{(2)}(k)]}.$$

Remark 5.5.1. Since reinsurance is priced according to the expected value principle with a unique risk margin of 0.5, the standard deviations of the reinsurer's loss ratio $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ can alternatively be expressed as

$$Q_{m:n}^{(1)}(k) = \frac{\sqrt{\text{Var}\big[\sum_{i=1}^m f_{Z_i}^{(1)}(X_i;k)\big]}}{1.5 \cdot \mathbb{E}\big[\sum_{i=1}^m f_{Z_i}^{(1)}(X_i;k)\big]} \quad \text{and} \quad Q_{m:n}^{(2)}(k) = \frac{\sqrt{\text{Var}\big[\sum_{i=1}^m f^{(2)}(X_i;k)\big]}}{1.5 \cdot \mathbb{E}\big[\sum_{i=1}^m f^{(2)}(X_i;k)\big]}, \tag{5.46}$$

meaning that $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ are just the scaled (by a factor $2/3$) coefficients of variation of the total reinsurance claims $\sum_{i=1}^m f_{Z_i}^{(1)}(X_i;k)$ and $\sum_{i=1}^m f^{(2)}(X_i;k)$, respectively.

In Figure 5.5, we plot $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ for $n = 3, 5$ and $\tau = 0, 0.5, 1$. We firstly observe that for all n and τ, $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ are both decreasing in m, which reveals and quantifies the diversification effect of pooling risks for the reinsurer. We notice that for all n and τ, $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ are both increasing in k. The reason for this is the following: As k increases, the part of the X_i's that is transferred to the reinsurer decreases ($\ell_0(k)$, $\ell_1(k)$ and $\ell(k)$ are all decreasing in k) and is shifted to the right tail ($d_0(k)$, $d_1(k)$ and $d(k)$ are all increasing in k), which reduces proportionally more the expectations of $\sum_{i=1}^m f_{Z_i}^{(1)}(X_i;k)$ and $\sum_{i=1}^m f^{(2)}(X_i;k)$ than their standard deviations and hence makes $Q_{m:n}^{(1)}(k)$ and $Q_{m:n}^{(2)}(k)$ both increase in k, from (5.46).

In Figure 5.6 we consider the relative difference $Q_{m:n}^{(1)}(k)/Q_{m:n}^{(2)}(k) - 1$ for $n = 3, 5$ and $\tau = 0, 0.5, 1$, which like for the premium differences before may be more instructive to study. We firstly notice that, as for $P_R^{(1)}(k)/P_R^{(2)}(k) - 1$, the curves $Q_{m:n}^{(1)}(k)/Q_{m:n}^{(2)}(k) - 1$ all contain a bend, which occurs at $k = \rho_0[X]$ in every case. The reason for this is the following: For the considered maximal acceptable levels of riskiness, the limit $\ell_0(k)$ decreases faster than $\ell_1(k)$ in k when $k < \rho_0[X]$, and slower when $k \geq \rho_0[X]$. The difference between $\ell_0(k)$ and $\ell_1(k)$ thus increases in k for $k < \rho_0[X]$, and decreases for $k \geq \rho_0[X]$. Then, given that for all i the potential difference between $f_0^{(1)}(X_i;k)$ and $f_1^{(1)}(X_i;k)$ depends directly and positively on the one between $\ell_0(k)$ and $\ell_1(k)$, increasing k when $k < \rho_0[X]$ will add some variability to each $f_{Z_i}^{(1)}(X_i;k)$ and hence also to $\sum_{i=1}^m f_{Z_i}^{(1)}(X_i;k)$, while for $k \geq \rho_0[X]$ it will remove some. As a result, $Q_{m:n}^{(1)}(k)$ tends to increase in k faster for $k < \rho_0[X]$ than it does for $k \geq \rho_0[X]$, which results in the notable

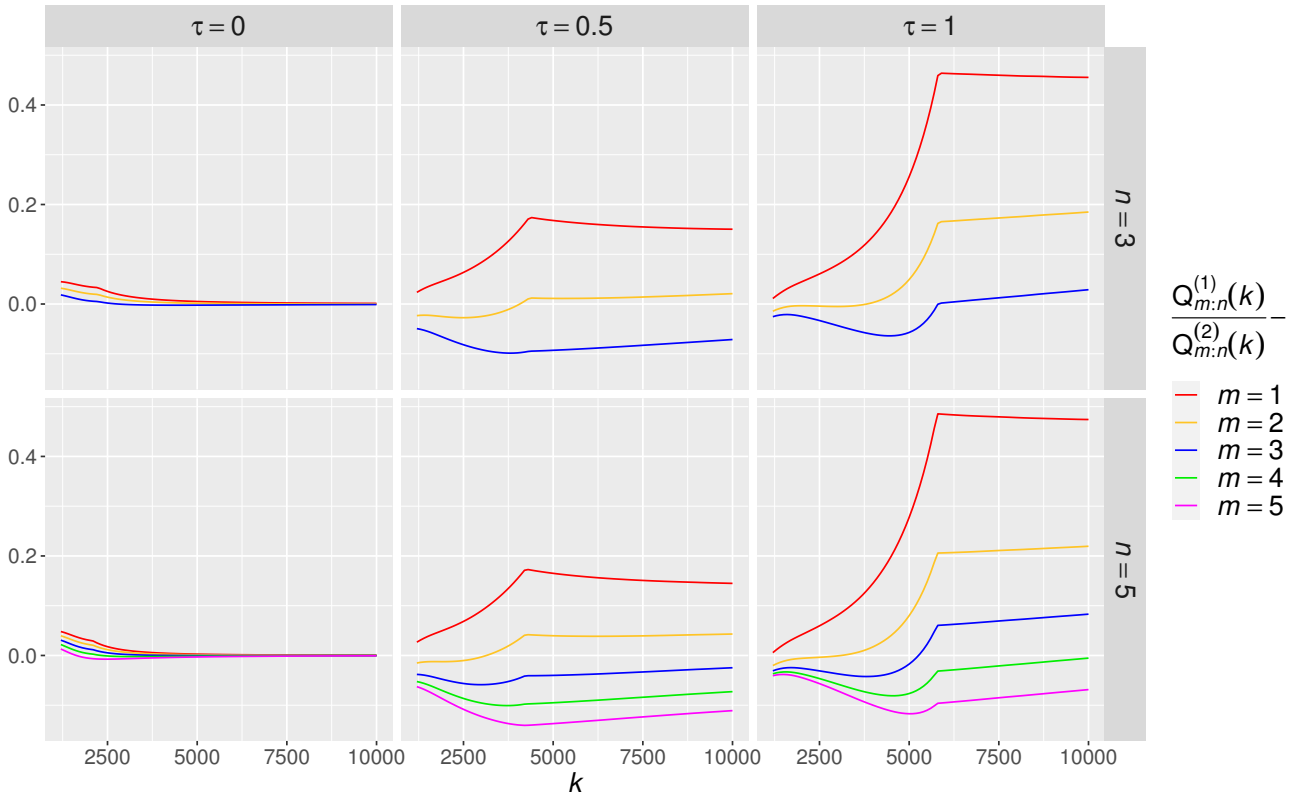

Figure 5.6: Relative difference of the standard-deviations of the reinsurer's loss ratio, for $n = 3, 5$ and $\tau = 0, 0.5, 1$.

bend that occurs in the relative difference $Q^{(1)}_{m:n}(k)/Q^{(2)}_{m:n}(k) - 1$ at $k = \rho_0[X]$.

We observe that for a single reinsurance deal ($m = 1$), this relative difference is always positive, meaning that the loss ratio of each optimal contingent bounded stop-loss cover has a greater standard deviation than the one of the benchmark. This comes from the fact the optimal contingent bounded stop-loss inherits from its property of varying with the relative performance some variability that the benchmark does not have. As a result, while on the one hand the contingent bounded stop-loss reduces the expectation of the ceded loss over the one of the benchmark, on the other hand it reduces proportionally less its standard deviation. The ceded loss $f^{(1)}_Z(X; k)$ therefore has a greater coefficient of variation than $f^{(2)}(X; k)$, which from (5.46) yields $Q^{(1)}_{1:n}(k)/Q^{(2)}_{1:n}(k) - 1 > 0$.

We note that $Q^{(1)}_{m:n}(k)/Q^{(2)}_{m:n}(k) - 1$ decreases in m. The explanation for that is as follows: On the one hand, by construction the Z_i's tend to be negatively correlated, which introduces some degree of negative dependence between the ceded losses $f^{(1)}_{Z_i}(X_i; k)$. On the other hand, since the ceded loss function $f^{(2)}(\,\cdot\,; k)$ is increasing, the ceded losses $f^{(2)}(X; k)$, ..., $f^{(2)}(X_m; k)$ have the same dependence structure as X, ..., X_m. Because of this, when the reinsurer sells reinsurance covers to more insurers in the market (when m increases), if these covers are the optimal contingent bounded stop-loss,

then he benefits from a larger diversification effect than if they are the benchmark ones. The standard deviation $Q^{(1)}_{m:n}(k)$ thus decreases faster than $Q^{(2)}_{m:n}(k)$ in m, which makes $Q^{(1)}_{m:n}(k)/Q^{(2)}_{m:n}(k) - 1$ to be decreasing in m.

Finally, when m approaches n, the relative difference $Q^{(1)}_{m:n}(k)/Q^{(2)}_{m:n}(k) - 1$ turns negative. Therefore, if the reinsurer has a large market share, then while keeping the same expectation of the loss ratio (cf. (5.45)), in several cases selling contingent covers rather than the traditional benchmark ones makes him benefit from a smaller standard deviation of his loss ratio. Figure 5.6 illustrates that the improvement is substantial when the insurers' losses are positively dependent. This is particularly noteworthy, as in this case positive dependence has a favourable impact for both the insurers and the reinsurer, which is rather uncommon in risk sharing constructions. Also, it suggests that such a favourable effect for the reinsurer may still be obtained using another measure of relative performance Z, as long as one ensures that it sufficiently introduce negative dependence among the contingent covers.

5.6 Conclusion

In this paper we studied the efficiency of contingent reinsurance covers as a particular example of structured reinsurance deals. Since for insurers the performance relative to other market participants is quite important in terms of potential financial distress costs, we investigated a reinsurance form that pays more in scenarios where the financial distress cost is increased. On the marginal side of the insurer, this can lead to a performance improvement, and for a reinsurer offering similar covers to several market participants there also can be a beneficial diversification effect. Under certain assumptions on the performance and risk measures involved, we proved optimality results of such a cover from the viewpoint of the insurer. We further illustrated the results in a detailed numerical example, where we also showed the hedging effect for the reinsurer writing several simultaneous such contracts to market participants.
It was the purpose of this paper to propose a new perspective for the analysis and the intuitive understanding of potential advantages of this structured reinsurance deal, which is why we deliberately chose a rather simple model that allowed to keep the calculations tractable and led to explicit results. Naturally, there are various directions in which the present results can be extended. Next to possibly different performance and risk measures than the ones considered in the paper, it could also be interesting to generalize the analysis to other reinsurance premium principles, and to reinsurance pricing techniques that are more specific to the individual reinsurer's situation rather than applying a general principle. Furthermore, it will be interesting to see to what extent the results of this paper still hold in more heterogeneous (re)insurance markets.

Acknowledgements. We would like to thank two anonymous referees for their careful reading and interesting suggestions to improve the presentation of the manuscript.

Bibliography

[1] Hansjörg Albrecher, Jan Beirlant, and Jozef L Teugels. *Reinsurance: Actuarial and Statistical Aspects*. John Wiley & Sons, 2017.

[2] Hansjörg Albrecher and Arian Cani. On randomized reinsurance contracts. *Insurance: Mathematics and Economics*, 84:67–78, 2019.

[3] Barry C Arnold, Narayanaswamy Balakrishnan, and Haikady Navada Nagaraja. *A first course in order statistics*. SIAM, 2008.

[4] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[5] Alexandru V Asimit, Valeria Bignozzi, Ka Chun Cheung, Junlei Hu, and Eun-Seok Kim. Robust and Pareto optimality of insurance contracts. *European Journal of Operational Research*, 262(2):720–732, 2017.

[6] Alexandru Vali Asimit, Tim J Boonen, Yichun Chi, and Wing Fung Chong. Risk sharing with multiple indemnity environments. *Available at SSRN 3616746*, 2020.

[7] Terje Aven and Enrico Zio. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety*, 96(1):64–74, 2011.

[8] August A Balkema and Laurens De Haan. Residual life time at great age. *The Annals of probability*, 2(5):792–804, 1974.

[9] HM Barakat, EM Nigm, and OM Khaled. Extreme value modeling under power normalization. *Applied Mathematical Modelling*, 37(24):10162–10169, 2013.

[10] Debabrata Basu. The family of ancillary statistics. In *Selected Works of Debabrata Basu*, pages 81–90. Springer, 2011.

[11] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. *Statistics of extremes: theory and applications*, volume 558. John Wiley & Sons, 2004.

[12] Valeria Bignozzi and Andreas Tsanakas. Model uncertainty in risk capital measurement. *Journal of Risk, Forthcoming*, 2014.

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

[14] Corina Birghila, Tim J Boonen, and Mario Ghossoub. Optimal insurance under maxmin expected utility. *Available at SSRN 3711743*, 2020.

[15] Corina Birghila and Georg Ch Pflug. Optimal XL-insurance under Wasserstein-type ambiguity. *Insurance: Mathematics and Economics*, 88:30–43, 2019.

[16] Tim J Boonen. Optimal reinsurance with heterogeneous reference probabilities. *Risks*, 4(3):26, 2016.

[17] Jun Cai and Yichun Chi. Optimal reinsurance designs based on risk measures: A review. *Statistical Theory and Related Fields*, 4(1):1–13, 2020.

[18] Jun Cai and Ken Seng Tan. Optimal retention for a stop-loss reinsurance. 2006.

[19] Andrew JG Cairns. A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, 27(3):313–330, 2000.

[20] Ka Chun Cheung, Wing Fung Chong, and Ambrose Lo. Budget-constrained optimal reinsurance design under coherent risk measures. *Scandinavian Actuarial Journal*, 2019(9):729–751, 2019.

[21] Yichun Chi and Ken Seng Tan. Optimal reinsurance under var and cvar risk measures: a simplified approach. *ASTIN Bulletin: The Journal of the IAA*, 41(2):487–509, 2011.

[22] Yichun Chi and Wei Wei. Optimal insurance with background risk: An analysis of general dependence structures. *Finance and Stochastics*, 24(4):903–937, 2020.

[23] VLADIMIR P Chistyakov. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability & Its Applications*, 9(4):640–648, 1964.

[24] Thierry Cohignac and Nabil Kazi-Tani. Quantile mixing and model uncertainty measures. *Bulletin Francais d'Actuariat*, 36(1):103–121, 2019.

[25] Christopher L Culp. *Structured finance and insurance: the ART of managing capital and risk*, volume 339. John Wiley & Sons, 2011.

[26] Rose-Anne Dana and Marco Scarsini. Optimal risk sharing with background risk. *Journal of Economic Theory*, 133(1):152–176, 2007.

[27] Philip Dawid. Fiducial inference then and now. *arXiv preprint arXiv:2012.10689*, 2020.

[28] Laurens De Haan, Ana Ferreira, and Ana Ferreira. *Extreme value theory: an introduction*, volume 21. Springer, 2006.

[29] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? *Structural safety*, 31(2):105–112, 2009.

[30] Jan Dhaene, Michel Denuit, Marc J Goovaerts, Rob Kaas, and David Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

[31] Jan Dhaene, Roger JA Laeven, Steven Vanduffel, Gregory Darkiewicz, and Marc J Goovaerts. Can a coherent risk measure be too subadditive? *Journal of Risk and Insurance*, 75(2):365–386, 2008.

[32] Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

[33] Paul Embrechts, Alexander Schied, and Ruodu Wang. Robustness in the optimization of risk measures. *Operations Research*, 70(1):95–110, 2022.

[34] EU-LEX. Commission delegated regulation (eu) 2015/35 of 10 october 2014 supplementing directive 2009/138/ec of the european parliament and of the council on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii). *Official Journal of the European Union*, 58, 2015.

[35] Yulian Fan. The solution of the optimal insurance problem with background risk. *Journal of Function Spaces*, 2019.

[36] Ronald A Fisher. Inverse probability. In *Mathematical proceedings of the Cambridge philosophical society*, volume 26, pages 528–535. Cambridge University Press, 1930.

[37] Ronald A Fisher. The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 139(838):343–348, 1933.

[38] Ronald A Fisher. The fiducial argument in statistical inference. *Annals of eugenics*, 6(4):391–398, 1935.

[39] Ronald A Fisher. Statistical methods and scientific inference. 1956.

[40] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical proceedings of the Cambridge philosophical society*, volume 24, pages 180–190. Cambridge University Press, 1928.

[41] Ivo Francioni and Florian Herzog. Probability-unbiased value-at-risk estimators. *Quantitative Finance*, 12(5):755–768, 2012.

[42] Donald AS Fraser. On fiducial inference. *The Annals of Mathematical Statistics*, 32(3):661–676, 1961.

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

[44] Kenneth Froot, Gary Venter, and John Major. Capital and value of risk transfer. In *14th Annual International AFIR Colloquium*, pages 181–195, 2004.

[45] Kenneth A Froot. The intermediation of financial risks: Evolution in the catastrophe reinsurance market. *Risk Management and Insurance Review*, 11(2):281–294, 2008.

[46] Lesław Gajek and Dariusz Zagrodny. Reinsurance arrangements maximizing insurer's survival probability. *Journal of Risk and Insurance*, 71(3):421–435, 2004.

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

[48] Boris Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of mathematics*, pages 423–453, 1943.

[49] Helmut Gründl and Hato Schmeiser. Pricing double-trigger reinsurance contracts: financial versus actuarial approach. *Journal of Risk and Insurance*, 69(4):449–468, 2002.

[50] Manuel Guerra and Maria L Centeno. Are quantile risk measures suitable for risk-transfer decisions? *Insurance: Mathematics and Economics*, 50(3):446–461, 2012.

[51] Jan Hannig. On generalized fiducial inference. *Statistica Sinica*, pages 491–544, 2009.

[52] Jan Hannig, Hari Iyer, Randy CS Lai, and Thomas CM Lee. Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111(515):1346–1361, 2016.

[53] Gur Huberman, David Mayers, and Clifford W Smith Jr. Optimal insurance policy indemnity schedules. *The Bell Journal of Economics*, pages 415–426, 1983.

[54] Stuart Jarvis, James Sharpe, and Andrew D Smith. Ersatz model tests. *British Actuarial Journal*, 22(3):490–521, 2017.

[55] Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.

[56] Arthur F Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, 1955.

[57] Robert E Kass and Larry Wasserman. The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435):1343–1370, 1996.

[58] Yuriy Krvavych and Michael Sherris. Enhancing insurer value through reinsurance optimization. *Insurance: Mathematics & Economics*, 38(3):495–517, 2006.

[59] Andreas Kull. Sharing risk–an economic perspective. *ASTIN Bulletin: The Journal of the IAA*, 39(2):591–613, 2009.

[60] Zinoviy Landsman and Andreas Tsanakas. Parameter uncertainty in exponential family tail estimation. *ASTIN Bulletin: The Journal of the IAA*, 42(1):123–152, 2012.

[61] Zhiyi Lu, Shengwang Meng, Leping Liu, and Ziqi Han. Optimal insurance design under background risk with dependence. *Insurance: Mathematics and Economics*, 80:15–28, 2018.

[62] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[63] Pietro Millossovich, Andreas Tsanakas, and Ruodu Wang. A theory of multivariate stress testing. *Available at SSRN 3966204*, 2021.

[64] R von Mises. La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1:141–160, 1936.

[65] Rahul Mukerjee and Dipak K Dey. Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika*, 80(3):499–505, 1993.

[66] Roger B. Nelsen. *An introduction to copulas*. Springer Series in Statistics. Springer, New York, second edition, 2006.

[67] Elisaveta Pantcheva. Limit theorems for extreme order statistics under nonlinear normalization. In *Stability problems for stochastic models*, pages 284–309. Springer, 1985.

[68] Georg Ch Pflug and Werner Romisch. *Modeling, measuring and managing risk*. World Scientific, 2007.

[69] James Pickands III. Statistical inference using extreme order statistics. *the Annals of Statistics*, pages 119–131, 1975.

[70] Marcin Pitera and Thorsten Schmidt. Unbiased estimation of risk. *Journal of Banking & Finance*, 91:133–145, 2018.

[71] Alfréd Rényi. On the theory of order statistics. *Acta Math. Acad. Sci. Hung*, 4(2), 1953.

[72] Matthias Scherer and Gerhard Stahl. The standard formula of solvency ii: a critical discussion. *European Actuarial Journal*, 11(1):3–20, 2021.

[73] David PM Scollnik. On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 2007(1):20–33, 2007.

[74] Thomas A Severini, Rahul Mukerjee, and Malay Ghosh. On an exact probability matching property of right-invariant priors. *Biometrika*, 89(4):952–957, 2002.

[75] M Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

[76] Jozef L Teugels. The class of subexponential distributions. *The Annals of Probability*, pages 1000–1011, 1975.

[77] Léonard Vincent. An alternative path to extremes. *Preprint, University of Lausanne*, 2022.

[78] Léonard Vincent. A note on risk assessment under parameter uncertainty. *Preprint, University of Lausanne*, 2022.

[79] Léonard Vincent. Risk estimation under parameter and model uncertainty. *Preprint, University of Lausanne*, 2022.

[80] Léonard Vincent, Hansjörg Albrecher, and Yuriy Krvavych. Structured reinsurance deals with reference to relative market performance. *Insurance: Mathematics and Economics*, 101:125–139, 2021.

[81] Shaun Wang. Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92, 1996.

[82] Sandy L Zabell. R.a. fisher and the fiducial argument. *Statistical Science*, pages 369–387, 1992.

[83] Yunzhou Zhu, Yichun Chi, and Chengguo Weng. Multivariate reinsurance designs for minimizing an insurer's capital requirement. *Insurance: Mathematics and Economics*, 59:144–155, 2014.