Modelling parameter uncertainty for risk capital calculation

Andreas Fröhlich and Annegret Weng*

February 2015

Abstract

For risk capital calculation within the framework of Solvency II the possible loss of basic own funds over the next business year of an insurance undertaking is usually interpreted as a random variable $\boldsymbol{X}$. If we assume that the parametric distribution family $\left\{\boldsymbol{X}(\theta)|\theta \in I \subseteq \mathbb{R}^d\right\}$ is known, but the parameter θ is unknown and has to be estimated from the available historical data, the undertaking faces parameter uncertainty. The parameter uncertainty has already been considered by several authors, see e.g. [2, 3, 6, 7, 11, 16].

To assess methods to model parameter uncertainty for risk capital calculations we apply a criterion going back to the theory of predictive inference which has already been used in the context of Solvency II. In particular, we show that the bootstrapping approach is not appropriate to model parameter uncertainty from the undertaking's perspective.

Based on ideas closely related to the concept of fiducial inference we introduce a new approach to model parameter uncertainty. For a wide class of distributions and for common estimators including the maximum likelihood method we prove that this approach is appropriate to model parameter uncertainty according to the criterion mentioned above.

Several examples demonstrate that our method can easily be applied in practice.

***Andreas Fröhlich**, R+V Versicherung AG, Zentrales Aktuariat Komposit, Raiffeisenplatz 1, 65189 Wiesbaden, Germany, **Annegret Weng**, Hochschule für Technik Schellingstr. 24 70174 Stuttgart, Germany

1 Introduction

The three-pillar project Solvency II pursued by the EU aims to harmonize insurance regulation. The first pillar requires an insurance undertaking to quantify its individual risks.
Solvency II defines the solvency risk capital requirement as the Value-at-Risk of basic own funds of the undertaking subject to a confidence level of 99.5% over a one-year period (cf. Article 101 of [18]). If we interpret the risk, i.e. the possible loss of basic own funds over the next business year, as a random variable $\boldsymbol{X}$, this requirement asks the undertaking to determine the 99.5%-quantile of $\boldsymbol{X}$.
Note that, in practice, not only the future realization of $\boldsymbol{X}$ but also the true distribution of $\boldsymbol{X}$ is unknown to the undertaking. In particular, the true 99.5%-quantile of $\boldsymbol{X}$ is not known. Therefore, in practice, the undertaking can only estimate its solvency risk capital requirement.

In this contribution we assume that the insurer sets up an internal model and estimates the model parameters using the available historical data. If we take the randomness of the historical observations into account, we view the solvency risk capital requirement **SCR** determined by the internal model as a random variable.
Thus there are two random variables, i.e. two sources of uncertainty:

- the random variable $\boldsymbol{X}$ and
- the modelled solvency risk capital requirement **SCR** whose randomness is caused by the randomness of the historical observations used for the calibration of the model.

Hence we interpret the solvency risk capital requirement as follows:

Definition 1.1. The required solvency risk capital **SCR** should be modelled in such a way that the losses $\boldsymbol{X}$ of the undertaking within the next business year will not exceed the **SCR** with a probability of 99.5% - taking into consideration the randomness of both $\boldsymbol{X}$ and **SCR**.

In the discussion above we think of $\boldsymbol{X}$ as the overall loss of the undertaking. However, all our arguments and the method described in this contribution can also be applied to a single risk factor or subrisk of the internal model for which the term "risk capital" might be meaningless (like the loss ratio, the severity of single losses or a chain ladder development factor in reserve risk etc.), if we just substitute the term "**SCR**" by "99.5%-quantile".

Note that the idea to consider the SCR as a random variable and to require that this random variable covers the losses with a probability of 99.5% is not new and goes back to the theory of predictive inference ([1], Chapter 10 in [19], [15] and [4]). Gerrard and Tsanakas [11] have applied this approach to parameter uncertainty for VaR-based solvency capital calculations. We discuss their results later in the paper.
In this contribution we concentrate on parameter uncertainty and ignore model uncertainty. That means, we assume that the parametric distribution family $\{\boldsymbol{X}(\theta) | \theta \in I \subset \mathbb{R}^d\}$ of $\boldsymbol{X}$ is known, in other words we can write $\boldsymbol{X} = \boldsymbol{X}(\theta_0)$ with the true, but unknown parameter (vector) θ_0. Throughout the paper we denote the parameter of the parametric distribution family by θ and the estimate of the unknown true parameter based on historical data by $\hat{\theta}$.

For simplicity, we assume that the cumulative distribution function $F_{\boldsymbol{X}(\theta)}$ is invertible and that the α-quantile of $\boldsymbol{X}(\theta)$ is given by the expression $F^{-1}_{\boldsymbol{X}(\theta)}(\alpha)$. Without the consideration of parameter uncertainty, the practitioner is used to model the risk capital as the 99.5%-quantile of the random fluctuation (process risk) based on the estimate $\hat{\theta}$. Of course, for a fixed estimate $\hat{\theta}$ this risk capital is a real number and not a random variable. However, the randomness of the historical sample to estimate the parameter induces a random deviation of the estimate $\hat{\boldsymbol{\theta}}$ from the true parameter θ_0. In this sense $F^{-1}_{\boldsymbol{X}(\hat{\boldsymbol{\theta}})}(0.995)$ should be viewed as a random variable. If the true risk is underestimated, i.e. if $F^{-1}_{\boldsymbol{X}(\hat{\theta})}(0.995) < F^{-1}_{\boldsymbol{X}(\theta_0)}(0.995)$, the calculated risk capital $F^{-1}_{\boldsymbol{X}(\hat{\theta})}(0.995)$ will not guarantee solvency with the required probability of 99.5%. Therefore, the potential underestimation of the risk capital induces an additional risk reducing the probability of solvency of the undertaking.
This motivates the question: What is the probability of solvency if we take the potential under-/overestimation of the true risk into account?
This question has already been raised in [11], where the authors showed that the probability of insolvency can be substantially higher than 0.5% if the undertaking calculates its solvency risk capital without taking parameter uncertainty into account. As a consequence, the confidence level of 99.5% required by Solvency II can only be achieved by a risk capital reflecting the parameter uncertainty. Furthermore, there is an argument already mentioned in [11] that backtesting can not reveal the required confidence level without taking parameter uncertainty into account (see Section 3.1, Remark 3.1).
Note that an undertaking managed on the basis of an underestimated risk capital faces an increased risk of insolvency: An overestimated value of the solvency coverage ratio may lead to wrong management decisions (e.g. con-

cerning the distribution of dividend payments). Furthermore, depending on the parameter uncertainty related to the respective subrisks, the capital allocation might be misleading. This yields a true economic risk.

The importance of the consideration of parameter risk has already been pointed out by many authors. There are two main lines for modelling parameter uncertainty recommended in the literature:

1. The approach most commonly considered in the existing literature models the underlying loss distribution as a predictive distribution, i.e. a mixture distribution of the loss distribution over a parameter space. It considers the parameter as a random variable, and determines the risk capital as the 99,5% of this predictive distribution (see e.g. [2, 3, 6, 7, 8, 11, 16]). The existing literature proposes two approaches to determine the parameter distribution: the bootstrapping approach (see Section 3.2) and the Bayesian approach (see Section 3.3).

2. Another approach uses only the process distribution as the underlying loss distribution, but takes parameter uncertainty into account by modifying the risk measure. This is achieved in [11] by raising the confidence level. In [2] the calculation of a capital add-on in terms of "residual estimation risk" is suggested. We discuss these approaches in Section 3.4.

In our contribution we concentrate on the first approach and consider a predictive distribution. We distinguish two different points of view:

1. The true parameter is assumed to be fixed but unknown and its estimate is assumed to be random. We call this point of view the "theoretical perspective".

2. The estimate is fixed (since there is only one observed sample) but the true parameter is assumed to be random. We call this point of view the "undertaking's perspective".

For example, the Bayesian approach models the undertaking's perspective. The bootstrapping approach approximates the distribution of the estimate $\hat{\boldsymbol{\theta}}$ and corresponds to the "theoretical perspective".
Note that the uncertainty of the undertaking about its future economic losses over the next business year can be characterized as follows:

a) The true risk $\boldsymbol{X}(\theta)$ of the undertaking according to the next business year does not depend on the estimate $\hat{\theta}$.

b) At the beginning of the business year the historical data have already been observed and are, therefore, fixed realizations of the corresponding random variables. Thus the estimate $\hat{\theta}$ is a fixed value and not a random variable.

c) There is uncertainty about the true parameter θ.

d) The undertaking has no knowledge about a prior distribution on θ, i.e. a distribution on θ without taking the observed data into account.

By a) and b) the theoretical perspective is not directly applicable: A predictive distribution with a parameter distribution of the estimate $\hat{\boldsymbol{\theta}}$ would not model the economic risk of the undertaking.
The combination of a) and c) shows that the "undertaking's perspective" is the economically relevant perspective in the sense that the uncertainty about the true parameter θ increases the risk of high potential losses. As an example, assume for simplicity that θ is a real number and the risk of high potential losses is increasing in θ. If $\hat{\theta} \approx \theta$, the modelled risk would be a good approximation of the true risk. But the uncertainty about θ yields the additional risk that $\theta \gg \hat{\theta}$, i.e. an additional risk of high losses in the next business year. This shows that the parameter risk is an economically relevant risk for the next business year, although the historical sample and thus the estimate are known - and, therefore, fixed - realizations at the beginning of the business year. Hence the undertaking's perspective is the economically relevant perspective referring to potential true losses in the next business year.
Therefore, the aim of this contribution is to model the parameter risk from the undertaking's perspective respecting the requirements a)-d) above.
The above arguments lead to the following definition:

Definition 1.2. The **parameter risk** from the undertaking's perspective refers to the uncertainty about the true parameter θ_0 given the fixed sample $(d_1, \ldots, d_n)$ resp. the estimate $\hat{\theta}$ of the true parameter θ_0 gained from the sample.

Hence from the undertaking's perspective the true parameter appears to be random.
After formalizing Definition 1.1 in Section 2 and summarizing results of [11] concerning the risk capital calculation without taking parameter uncertainity into account in Subsection 3.1 we show in Subsection 3.2 that the bootstrapping approach is not an appropriate method to model parameter uncertainty, since it does not guarantee a sufficient level of solvency. In Subsections 3.3 and 3.4 we discuss the results by Gerrard and Tsanakas [11] on the Bayesian

approach and their additional idea to raise the confidence level for the risk capital calculation to reflect the parameter uncertainty. Subsection 3.4 also covers the idea of Bignozzi and Tsanakas [2] of calculating a capital add-on in terms of the residual risk.

In Section 4 we introduce a new method closely related to Fisher's concept of fiducial inference (cf. [9]) taking all requirements a)-d) into account. By a slight modification of Fisher's "one-parameter concept" of fiducial inference we deduce a generalization to the multi-parameter case. Using the wording introduced above: Our method yields a transformation from the theoretical perspective into the economically relevant perspective from the undertaking. This change of perspective is achieved in a straight forward way: Roughly spoken, the idea is to express the estimate $\hat{\theta}$ by the true parameter θ and then to solve this equation for θ.

Despite the theoretical criticism on fiducial inference (see [10, 20]) we show that our method works well in the context of parameter uncertainty for risk capital calculation: For a wide class of distributions and for common estimators including the maximum likelihood method, we show that the method introduced in Section 4.3 is appropriate to model parameter uncertainty in the sense of Definition 1.1 and 1.2 (and their formalization given in Section 2, Definition 2.4).

In particular, we do not need any additional assumptions, which can not be observed in practice (like e.g. the assumption of a prior distribution, cf. Condition d) above).

For transformed location-scale families we derive explicit formulas for the parameter distribution, which guarantee that our method is easy to apply, well performing and compatible with models based on Monte Carlo simulation. Note that transformed location-scale families cover many distributions relevant in practice like the normal/lognormal, the exponential/Pareto, the Gumbel or the logistic distributions.

Notation 1. Random variables are printed in bold.

Throughout the article let $\boldsymbol{\zeta}$ resp. $\boldsymbol{\xi}$ be independent, on $[0;1]^n$ resp. on $[0;1]$ uniformly distributed random variables. By ζ resp. ξ we denote fixed realizations of these random variables.

Let $I \subseteq \mathbb{R}^d$ be a set of parameters and let $\{\boldsymbol{X}(\theta)|\theta \in I \subseteq \mathbb{R}^d\}$ and $\{F_{\boldsymbol{X}(\theta)}|\theta \in I \subseteq \mathbb{R}^d\}$ be the corresponding set of random variables resp. the set of corresponding distribution functions. For the sake of a clear and comprehensive presentation of our ideas and to avoid technical difficulties we assume throughout this paper that the distribution of $\boldsymbol{X}$ is continuous and that the inverse $F^{-1}_{\boldsymbol{X}(\theta)}$ of the cumulative distribution function of $\boldsymbol{X}$ exists. We set $X(\boldsymbol{\xi}, \theta) := F^{-1}_{\boldsymbol{X}(\theta)}(\boldsymbol{\xi})$ and use $X(\boldsymbol{\xi}, \theta)$ to denote the random variable $\boldsymbol{X}(\theta)$.

Analogously we write the data vector in the form $\boldsymbol{D} = \boldsymbol{D}(\theta) = D(\boldsymbol{\zeta}, \theta)$ where $\boldsymbol{\zeta}$ is uniformly distributed on $[0;1]^n$: According the Theorem of Sklar [17] there exists an n-dimensional copula $\mathcal{C}$ with $F_{\boldsymbol{D}}(d_1, \ldots, d_n) = \mathcal{C}(F_{\boldsymbol{D}_1}(d_1), \ldots, F_{\boldsymbol{D}_n}(d_n))$ with the marginal distribution functions $F_{\boldsymbol{D}_i} = F_{\boldsymbol{D}_i(\theta)}$, $i = 1, \ldots, n$. Let $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n) \in [0;1]^n$ be an uniformly distributed random vector whose distribution function is equal to $\mathcal{C}$. We define $D : [0,1]^n \times I \to \mathbb{R}^n$ by

$$D(\zeta, \theta) := \left(F^{-1}_{\boldsymbol{D}_1}(\zeta_1), \ldots, F^{-1}_{\boldsymbol{D}_n}(\zeta_n)\right).$$

In fact we have

$$\begin{aligned} F_{D(\boldsymbol{\zeta},\theta)}(d_1, \ldots, d_n) &= P(F^{-1}_{\boldsymbol{D}_1}(\boldsymbol{\zeta}_1) \leq d_1, \ldots, F^{-1}_{\boldsymbol{D}_n}(\boldsymbol{\zeta}_n) \leq d_n) \\ &= P(\boldsymbol{\zeta}_1 \leq F_{\boldsymbol{D}_1}(d_1), \ldots, \boldsymbol{\zeta}_n \leq F_{\boldsymbol{D}_n}(d_n)) \\ &= \mathcal{C}(F_{\boldsymbol{D}_1}(d_1), \ldots, F_{\boldsymbol{D}_n}(d_n)) = F_{\boldsymbol{D}}(d_1, \ldots, d_n). \end{aligned}$$

If the random historical data is given by $\boldsymbol{D} = D(\boldsymbol{\zeta}, \theta)$ and $\boldsymbol{X} = X(\boldsymbol{\xi}, \theta)$ is the true risk, we assume that $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ are independent.

2 The modelled risk capital

Let $\boldsymbol{X}$ be the random variable with distribution function $F_{\boldsymbol{X}}$ describing the true risk for the next business year. We assume that the parametric familiy $\{\boldsymbol{X}(\theta) | \theta \in I \subseteq \mathbb{R}^d\}$ of $\boldsymbol{X}$ resp. the parametric class of distribution functions $\{F_{\boldsymbol{X}(\theta)} | \theta \in I \subseteq \mathbb{R}^d\}$ of $F_{\boldsymbol{X}}$ are known. Let $\theta_0 \in I \subseteq \mathbb{R}^d$ be the unknown true parameter corresponding to the true risk $\boldsymbol{X}$, i.e. $\boldsymbol{X} = \boldsymbol{X}(\theta_0)$ and $F_{\boldsymbol{X}} = F_{\boldsymbol{X}(\theta_0)}$.
We assume that the historical data is a sample $(d_1, \ldots, d_n)$ drawn from the random variable $\boldsymbol{D} = (\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$ whose distribution function $F_{\boldsymbol{D}} = F_{\boldsymbol{D}(\theta)}$ is known except for the true, but unknown parameter $\theta_0 \in I \subseteq \mathbb{R}^d$. More precisely, the marginal distribution function of $\boldsymbol{D}_i$ belongs to the set $\{F_{\boldsymbol{D}_i(\theta)} | \theta \in I \subseteq \mathbb{R}^d\}$ for $i = 1, \ldots, n$, and the n-dimensional copula representing the dependencies between the $\boldsymbol{D}_i$ is known.

Remark 2.1. (Special case) The previous paragraph describes the most general setting. In most discussions on parameter uncertainty the authors restrict to the important, but special case $\boldsymbol{D}_i = \boldsymbol{X}_i$ where $\boldsymbol{X}_i$ are independent variables with $\boldsymbol{X}_i \sim \boldsymbol{X}$ for $i = 1, \ldots, n$ and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is independent of $\boldsymbol{X}$.
However, in general, the data used to estimate the parameters of the model is given on a more granular level than the modelled risk. In particular, the special case $\boldsymbol{D}_i = \boldsymbol{X}_i$ does not cover the situation where $\boldsymbol{X}$ is the overall loss of the undertaking.

For a given set of data $(d_1, \ldots, d_n)$ we assume that the undertaking takes parameter uncertainty into account by modelling its risk as a predictive distribution (cf. [13], Section 15). This means that the undertaking models its risk by the following steps:

1. Parameter distribution: Using some method M generate a probability distribution $\mathcal{P} = \mathcal{P}(d_1, \ldots, d_n; M)$ for the parameter $\boldsymbol{\theta}_{sim}$ depending on the sample $(d_1, \ldots, d_n)$.

2. Modelled risk: Set $\boldsymbol{Y} := X(\boldsymbol{\xi}, \boldsymbol{\theta}_{sim})$ (cf. Notation 1) where $\boldsymbol{\xi}$ is an uniformly distributed random variable on $[0; 1]$ and independent of $\boldsymbol{\theta}_{sim}$.

Note that the random variable $\boldsymbol{Y}$ depends on the sample $(d_1, \ldots, d_n)$ and the method M resp. the probability distribution $\mathcal{P}$. In practice, the procedure above will be performed using a Monte Carlo simulation method.

Definition 2.2. The random variable $\boldsymbol{Y}$ given by the two step procedure described above is called the **modelled risk**. It depends on the method M and the sample $(d_1, \ldots, d_n)$. We call

$$SCR(\alpha; d_1, \ldots, d_n; M) := F_{\boldsymbol{Y}}^{-1}(\alpha)$$

the **modelled risk capital required with confidence level** α.

Example 2.3. Note that $\boldsymbol{\theta}_{sim} \equiv \hat{\theta}$ in case the undertaking does not model its parameter risk. A non-trivial example for the method M is non-parametric bootstrapping (see Subsection 3.2).

Note that the modelled risk capital $\boldsymbol{SCR} := SCR(\alpha; \boldsymbol{D}_1, \ldots, \boldsymbol{D}_n; M)$ is itself a random variable whose distribution depends on the distribution of the historical data $\boldsymbol{D}$. We reformulate Definition 1.1 as follows:

Definition 2.4. The method M resp. the probability distribution $\mathcal{P}$ are called **appropriate for the confidence level** α if

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{D}_1, \ldots, \boldsymbol{D}_n; M)\right) = \alpha. \tag{1}$$

The method M resp. the probability distribution $\mathcal{P}$ are called **appropriate** if they are appropriate for all α with $0 < \alpha < 1$.

As pointed out in the introduction, the criterion (1) has already been used by several authors. For an application in the specific context of VaR-based solvency capital calculation see Equation (1) in [11].

3 Previous approaches

Existing literature recommends to approach the measurement of parameter uncertainty either by bootstrapping (see e.g. [6, 7, 14, 16]), by the Bayesian approach (see e.g. [3, 6, 7, 11, 16]), by the raise of confidence level [11] or by a capital add-on (residual risk) [2]. Before proceeding by a discussion of these ideas we summarize results of [11] demonstrating that parameter uncertainty cannot be neglected.
In Subsections 3.1 and 3.2 we consider the special case where the data points are independent historical observations of $\boldsymbol{X}$, i.e. $\boldsymbol{D}_i := \boldsymbol{X}_i \sim \boldsymbol{X}$ for $i = 1, \ldots, n$, and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is independent of $\boldsymbol{X}$.

3.1 Ignoring the parameter uncertainty

Setting $\boldsymbol{\theta}_{sim} \equiv \hat{\theta}$ does not yield an appropriate probability distribution:
LThe following table compares α with $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ for different confidence levels α and different n where M is the maximum likelihood method for parameter estimation and $\boldsymbol{X}$ is either a lognormal or normal distribution or the one-parameter Pareto distribution with shape parameter θ. In these cases it can be shown that $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ does not depend on the parameter ([11], Section 2.1).

For $\alpha = 99\%$ and $\alpha = 99.5\%$ the probabilities $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ stated below can be deduced from Table 1 in [11]. We added results for lower quantiles to allow for the comparison with the bootstrapping approach in Section 3.2.

		Confidence level α			
n	Distribution	90%	95%	99.0%	99.5%
$n = 10$	Normal/Lognormal Distribution	86.2%	91.5%	96.8%	97.8%
	Pareto Distribution	87.4%	92.7%	97.7%	98.6%
$n = 20$	Normal/Lognormal Distribution	88.1%	93.3%	98.0%	98.8%
	Pareto Distribution	88.7%	93.9%	98.4%	99.1%
$n = 50$	Normal/Lognormal Distribution	89.2%	94.3%	98.7%	99.2%
	Pareto Distribution	89.5%	94.5%	98.8%	99.4%
$n = 100$	Normal/Lognormal Distribution	89.6%	94.7%	98.8%	99.4%
	Pareto Distribution	89.7%	94.8%	98.9%	99.4%

Table 1: $P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ without the consideration of parameter risk, i.e. $\boldsymbol{\theta}_{sim} \equiv \hat{\theta}$ for different α and different n (cf. Table 1 in [11] for similar probabilities)

Note that $P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)) < \alpha$ for every α and every n. Hence $\boldsymbol{\theta}_{sim} \equiv \hat{\theta}$ is not appropriate in the sense of Definition 2.4. In particular, for $\alpha = 99.5\%$ and small sample size n the failure probability $1 - P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ is substantially higher than 0.5%. For example, for $n = 10$ the probability that the risk capital does not cover the losses is 2.2% resp. 1.4% for the lognormal resp. the Pareto distribution instead of 0.5% as required by Solvency II.
As mentioned in [11], the shape and the heaviness of the tail of the distribution does not influence the solvency probability in an obvious way.

Remark 3.1. (Backtesting) For another interpretation of the results consider a backtesting scenario (cf. [11], p. 731 or [4, 15]). Let us assume that we have observed N realizations for the data and the random variable $\boldsymbol{X}$ given by $(X_1^{(i)}, \ldots, X_n^{(i)}; X^{(i)})$, $1 \leq i \leq N$. For each set of data $(X_1^{(i)}, \ldots, X_n^{(i)})$ we calculate a risk capital $SCR(\alpha; X_1^{(i)}, \ldots, X_n^{(i)}; M)$ and compare it with $X^{(i)}$. For large N the probability $P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ represents the fraction of scenarios for which the risk capital $SCR(\alpha; X_1^{(i)}, \ldots, X_n^{(i)}; M)$ covers the loss $X^{(i)}$. Note that the approach for generating the results in Table 1 is equivalent to a simulation of backtesting histories, i.e Table 1 represents the fraction of scenarios for which the calculated risk capital covers the losses. Recall that Article 101 of the EU framework directive for Solvency

II [18] requires the undertaking to hold a risk capital that the undertaking survives in 99.5% of all cases. As the results in Table 1 show, in general, this requirement is not met by a risk capital calculation ignoring parameter uncertainty.

3.2 The bootstrapping approach

Bootstrapping allows to approximate the probability distribution or - in other words - the uncertainty of the random estimate $\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ (cf. [6, 7, 8, 14, 16]). Recall the notion introduced in the introduction: It is important to note that the bootstrapping approach does not correspond to the undertaking's perspective, since it models a distribution of the estimate rather than the uncertainty of the undertaking about the true parameter given a fixed estimate $\hat{\theta}$.
Let $\boldsymbol{X} = \boldsymbol{X}(\theta_0)$ be the true risk with fixed, but unknown parameter θ_0. We distinguish between the parametric and the non-parametric version of the bootstrapping algorithm. The **non-parametric version** (as recommended in [6, 16]) given a sample $x_1, \ldots, x_n$ drawn from $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ can be defined by the following two step procedure:

1. Draw a sample $x_1^*, \ldots, x_n^*$ of size n from $x_1, \ldots, x_n$ where the x_i^* need not be pairwise distinct.

2. For each sample we determine a simulated parameter θ^* using an estimation method (e.g. maximum likelihood estimation) and draw from $\tilde{\boldsymbol{X}} = \boldsymbol{X}(\theta^*)$.

The iterative combination of step 1 and 2 defines a random variable $\boldsymbol{Y} = X(\boldsymbol{\xi}, \boldsymbol{\theta}^*)$ which depends on the fixed sample $x_1, \ldots, x_n$. For the **parametric version** we replace the first step by:

1.′ Given an estimate $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ we draw a random sample $x_1^*, \ldots, x_n^*$ from $\hat{\boldsymbol{X}} = \boldsymbol{X}(\hat{\theta})$.

According to Definition 2.2 we calculate $SCR(\alpha; x_1, .., x_n; M)$ as the α-quantile of the distribution of $\boldsymbol{Y}$ generated by a Monte Carlo simulation. More precisely, we set $SCR(\alpha; x_1, \ldots, x_n; M)$ equal to the $\lfloor (1-\alpha) \cdot N \rfloor$-th largest value of $\boldsymbol{Y}$ where N is the number of scenarios. As opposed to this, without modelling parameter risk $SCR(\alpha; x_1, \ldots, x_n; M)$ is calculated as the α-quantile of the distribution of $\boldsymbol{X}(\hat{\theta})$ generated by Monte Carlo simulation where $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ is estimated from the sample $x_1, \ldots, x_n$.
The bootstrapping method measures the uncertainty of the estimated parameter. As pointed out above, it does not answer the question, how to determine

the parameter uncertainty from the undertaking's perspective given the estimate $\hat{\theta}$.

These qualitative objections can be supported by quantitative results: In the sequel, we consider the random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ of historical data and examine whether the bootstrapping approach defines an appropriate probability distribution in the sense of Definition 2.4 or if it is at least conservative for the estimate of the α-quantile meaning that

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right) \geq \alpha. \tag{2}$$

Thus for a given confidence level α we are interested in the probability

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right). \tag{3}$$

As an example we consider an undertaking with two different lines of non-life insurance business. We assume that $\boldsymbol{X}$ describes the combined ratio (loss and expense ratio) of the respective line of business. For the sake of simplicity we suppress the subscript of $\boldsymbol{X}$ indicating the line of business. Note that in practice, mean and standard deviation of the combined ratio depend on the line of business and on the size of the portfolio.

1. For the first line of business the random variable $\boldsymbol{X}$ is lognormally distributed. The undertaking does not know the parameters of the distribution and estimates them using the maximum likelihood method.

2. The second line of business consists of heavy tail risks. We assume that $\boldsymbol{X}$ follows a Pareto Type II distribution with distribution function $F(x) = 1 - (1 + x)^{-(\theta_0+1)}$. The undertaking estimates the parameter θ_0 by the maximum likelihood method.

Since the normal/lognormal distribution and the Pareto distribution are examples of transformed location-scale families it follows from Corollary **??** in Appendix B that $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ is independent of the true parameter θ.

For the non-parametric bootstrapping we get the following results for the probability $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ by using a Monte Carlo simulation technique with 10.000 simulations for the sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and 10.000 simulations for $SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)$ given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ for different level of confidence α and different n:

		Confidence level α			
n	Distribution	90%	95%	99.0%	99.5%
$n = 10$	Normal/Lognormal Distribution	85.7%	91.5%	97.1%	98.1%
	Pareto Distribution	87.8%	93.4%	98.3%	99.1%
$n = 20$	Normal/Lognormal Distribution	87.8%	93.3%	98.2%	99.0%
	Pareto Distribution	88.8%	94.2%	98.7%	99.3%
$n = 50$	Normal/Lognormal Distribution	89.2%	94.3%	98.7%	99.3%
	Pareto Distribution	89.5%	94.7%	98.9%	99.4%
$n = 100$	Normal/Lognormal Distribution	89.6%	94.6%	98.9%	99.4%
	Pareto Distribution	89.8%	94.9%	98.9%	99.46%

Table 2: $P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ with non-parametric bootstrapping

The next table displays the results for parametric bootstrapping (again using a Monte Carlo simulation technique with 10.000 simulations for the sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and 10.000 simulations for $SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$):

		Confidence level α			
n	Distribution	90%	95%	99.0%	99.5%
$n = 10$	Normal/Lognormal Distribution	85.6%	91.4%	97.2%	98.3%
	Pareto Distribution	87.8%	93.5%	98.5%	99.1%
$n = 20$	Normal/Lognormal Distribution	87.8%	93.2%	98.3%	99.0%
	Pareto Distribution	88.9%	94.3%	98.7%	99.4%
$n = 50$	Normal/Lognormal Distribution	89.2%	94.3%	98.7%	99.3%
	Pareto Distribution	89.6	94.7%	98.8%	99.4%
$n = 100$	Normal/Lognormal Distribution	89.6%	94.7%	98.9%	99.4%
	Pareto Distribution	89.8%	94.8%	98.9%	99.4%

Table 3: $P(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)))$ with parametric bootstrapping

Note that in both cases (for non-parametric and parametric bootstrapping) the probability $P(\boldsymbol{X} \le SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)) < \alpha$ for every α and every n. The examples show that the risk capital obtained using the bootstrapping approach is not sufficient to cover the losses with the required probability of at least α. For example, for $n = 10$ and the quantile $\alpha = 99.5\%$ relevant for Solvency II the failure probability that the risk capital does not cover the losses is 1.9% (resp. 1.7%) for the lognormal distribution using non-parametric resp. parametric bootstrapping. Bootstrapping is neither an appropriate probability distribution in the sense of Definition 2.4 nor is it conservative in the sense of (2).
For low quantiles and the lognormal distribution bootstrapping is even worse than the method without the consideration of parameter uncertainty. For higher quantiles bootstrapping increases the probability $P(\boldsymbol{X} \le SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ but in a non-systematic way not leading to the desired value α.

Remark 3.2. A reasonable guess would be to assume that the difference between $P(\boldsymbol{X} \le SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ and α increases with the standard deviation or the heaviness of the tail of the distribution. However, since the solvency probability $P(\boldsymbol{X} \le SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ does not depend on the parameter (see Appendix B), the results do not depend on the standard deviation for the normal/lognormal distribution or the heaviness of the tail of the distribution. Although Pareto distributions have an heavier tail than normal distributions, the failure probability is smaller. Hence, there seems to be no direct link between the heaviness of the tail of the distribution and the difference between $P(\boldsymbol{X} \le SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M))$ and α.

3.3 The Bayesian approach

Another approach to model parameter uncertainty recommended in the literature [3, 6, 7, 11, 16] is based on Bayesian theory. The Bayesian approach assumes that the unknown true parameter is the realization of a random variable $\boldsymbol{\theta}$ with a prior distribution $\pi(\theta)$. In this setting, Bayes' theorem (cf. [13], Theorem 15.8.) allows to determine the predictive distribution, that is, the distribution for a new observation y given a known sample $x = (x_1, \ldots, x_n)$. For more details on the Bayesian method see [13], Section 15 and [19], Section 3.
Notice that the application of the Bayesian approach within a Monte Carlo based internal model usually demands rather sophisticated mathematical methods resp. integration of special software (cf. [13], Section 15.4).
The a priori distribution required by the Bayesian approach can not be observed in practice (recall d) on p. 4). However, in [11] the authors proved

that for distributions belonging to a transformed location-scale family (for the definition of a transformed location-scale family see Definition A.3) together with the maximum likelihood method for parameter estimation and for the "non-informative prior" $\pi(\mu, \sigma) = \sigma^{-1}$

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{D}_1, \ldots, \boldsymbol{D}_n; M)\right) = \alpha,$$

i.e. the obtained parameter distribution is appropriate in the sense of Definition 2.4 (see [11], Example 9 and Proposition 2 in the Appendix A.3).

3.4 Adjusting the risk capital

Gerrard and Tsanakas propose a further method based on the raise of the confidence level to reflect the parameter uncertainty ([11], Section 3.1).
In our opinion, this method has some drawbacks: Although parameter risk an an economic risk like the risk of random fluctuation, the method of raising the confidence level does not yield a probability distribution of the full risk (including parameter uncertainty). This means a loss of information. Futhermore, it is not clear how to apply the method to models based on Monte Carlo simulation where the subrisks are aggregated scenario wise to an overall risk.
Similar reservations can be expressed concerning the approach suggested in [2] to increase the risk capital by a capital add-on reflecting the residual risk due to parameter uncertainty.

4 An approach based on Fisher's fiducial inference

4.1 Change of perspective

Recall the meaning of the terms "theoretical perspective" and "undertaking's perspective" from the introduction. As already pointed out, the undertaking's perspective is the economically relevant perspective, since the uncertainty about the true parameter refers to the real risk of potential true losses of the undertaking. In contrast, the randomness of the estimate $\hat{\boldsymbol{\theta}}$ in the theoretical perspective does not directly affect the potential real losses of the undertaking, since the true risk of the undertaking does not depend on $\hat{\boldsymbol{\theta}}$. Hence our goal is the measurement of the parameter uncertainty from the undertaking's perspective taking into account the requirements a)-d) mentioned in the introduction on p. 4.

To achieve this goal our idea is to transform the theoretical perspective into the undertaking's perspective. More precisely, we would like to deduce a probability distribution for the true parameter conditional on the observed estimate $\hat{\theta}_0$ using the dependency of the distribution of the estimate $\hat{\boldsymbol{\theta}}$ on the true parameter θ without the need for an a priori distribution.
In the figure below the picture on the left hand side illustrates the "theoretical perspective": The true parameter is fixed but, if we draw N different samples, we get different estimates $\hat{\theta}_1, \ldots, \hat{\theta}_N$ based on the respective sample. The picture on the right hand side represents the undertaking's perspective, where $\hat{\theta}_0$ is estimated from the fixed observed historical data. There is uncertainty about the true parameter illustrated by different possible values, say $\theta_1, \ldots, \theta_N$, of the true parameter:

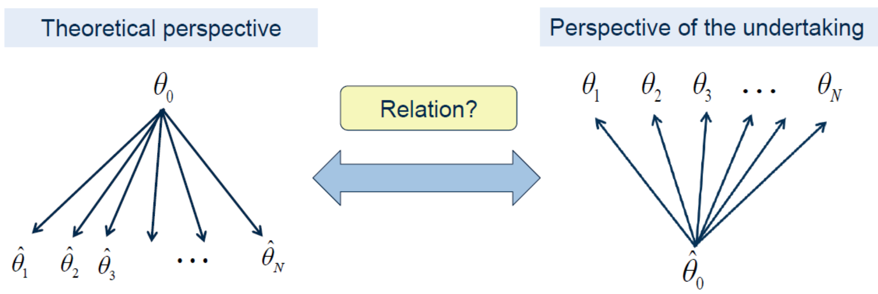

Figure 1: Can one perspective be transformed into the other?

In the sequel we denote the true parameter by θ (not θ_0), since from the undertaking's perspective θ is not known and, therefore, appears to be random.

4.2 Fiducial inference

In the sequel we show that Fisher's fiducial argument ([9], for a short introduction see [10], more details can be found in [20]) yields a concept how to perform the change of perspective described in Section 4.1, without any further assumptions as e.g. the Bayesian assumption of a prior distribution. We give a simple standard example illustrating Fisher's concept of fiducial inference: Let the risk $\boldsymbol{X}$ as well as the independent observations $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ be $N(\mu; 1)$-distributed. The estimated parameter $\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}$ is $N(\mu; 1/n)$-distributed and $\boldsymbol{Z} = \hat{\boldsymbol{\mu}} - \mu$ is $N(0; 1/n)$-distributed, that is, in this theoretical perspective $\hat{\boldsymbol{\mu}}$ is random but μ is fixed. The change into the undertaking's perspective is achieved by solving the equation with respect to $\mu = \hat{\boldsymbol{\mu}} - \boldsymbol{Z}$:

Inserting the observed estimate $\hat{\mu}$ and keeping in mind that $\boldsymbol{Z}$ is $N(0;1)$-distributed implies a distribution for the unknown true parameter μ, i.e., given $\hat{\mu}$, it follows that $\boldsymbol{\mu}$ is $N(\hat{\mu};1)$-distributed from the undertaking's perspective.

In general, let $\boldsymbol{Z} = G(\hat{\boldsymbol{\theta}};\theta)$ be a pivotal quantity, i.e. the distribution of $\boldsymbol{Z}$ is independent of θ. Then we get the fiducial distribution of $\boldsymbol{\theta}$ given an observed estimate $\hat{\theta}_0$ by solving the equation $\boldsymbol{Z} = G(\hat{\theta}_0;\boldsymbol{\theta})$ for $\boldsymbol{\theta}$.

For our problem a pivotal quantity is easy to find:
Let $\hat{F}(\cdot|\theta)$ be the cumulative distribution function of $\hat{\boldsymbol{\theta}}$. Thus $\boldsymbol{\varsigma} := \hat{F}(\hat{\boldsymbol{\theta}}|\theta)$ is uniformly distributed on $[0;1]$ and, therefore, $\hat{F}(\hat{\boldsymbol{\theta}}|\theta)$ is a pivotal quantity. For an observed estimate $\hat{\theta}_0$ the fiducial distribution of $\boldsymbol{\theta}$ is given as a solution of the equation $\hat{F}(\hat{\theta}_0|\boldsymbol{\theta}) = \boldsymbol{\varsigma}$, i.e. $\boldsymbol{\theta} = \hat{F}(\hat{\theta}_0|\cdot)^{-1}(\boldsymbol{\varsigma})$.
Note that $\hat{\boldsymbol{\theta}} \sim \hat{F}(\cdot|\theta)^{-1}(\boldsymbol{\varsigma}) =: \hat{\theta}(\boldsymbol{\varsigma},\theta)$. Since

$$\hat{\theta}(\boldsymbol{\varsigma},\boldsymbol{\theta}) = \hat{F}(\cdot,\boldsymbol{\theta})^{-1}(\boldsymbol{\varsigma}) = \hat{F}(\cdot|\boldsymbol{\theta})^{-1}\left(\hat{F}(\hat{\theta}_0|\boldsymbol{\theta})\right) = \hat{\theta}_0,$$

the fiducial quantity $\boldsymbol{\theta} = \theta(\boldsymbol{\varsigma};\hat{\theta}_0)$ solves the equation

$$\hat{\theta}(\boldsymbol{\varsigma},\cdot) = \hat{\theta}_0. \tag{4}$$

Note that, intuitively, solving Equation (4) is equivalent to the change from the theoretical perspective into the undertaking's perspective: Roughly spoken, solving Equation (4) is equivalent to "inverting the direction of the arrows" in Figure 1 from $\theta \mapsto \hat{\theta}$ into $\hat{\theta} \mapsto \theta$ assuming that the arrows are parametrized by $\boldsymbol{\varsigma}$.

Note that there have been numerous criticisms by generations of statisticians of the fiducial approach: Fisher's arguments were essentially based on intuition and he was not able to justify that fiducial probabilities are in fact probabilities in an acceptable sense. Indeed, there are examples where fiducial probabilities do not add or integrate to 1 (for further examples and more details see [10]). As pointed out in [20], Fisher's attempts to generalize his concept to the multiparameter setting were not satisfactory.
In the sequel we introduce a new method based on a slight modification of Equation (4) which generalizes to the multiparameter setting. Despite the criticism of Fisher's concept of fiducial inference we prove the appropriateness of the parameter distribution resulting from our method in the sense of Definition 2.4 under a special uniformity criterion (see Theorem 4.3). This uniformity criterion is proved to hold for a wide class of distributions and for common estimators relevant in practice.

4.3 The method

Consider a random set of data $\boldsymbol{D} = (\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n)$ with distribution function $F_{\boldsymbol{D}} = F_{\boldsymbol{D}(\theta)}$. Recall the notation $\boldsymbol{D} = D(\boldsymbol{\zeta}, \theta)$ with $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)$ uniformly distributed on $[0;1]^n$ introduced in Notation 1 in Section 1 and let $(d_1, \ldots, d_n) = D(\zeta, \theta)$ be a fixed set of data drawn from $\boldsymbol{D}$.
Moreover, we fix an estimation method S which given the sample $(d_1, \ldots, d_n)$ determines an estimate $\hat{\theta}$, that is,

$$S : (d_1, \ldots, d_n) \mapsto \hat{\theta} \in I \subset \mathbb{R}^d.$$

We define a mapping $\hat{\theta} : [0;1]^n \times I \to I$ by

$$\hat{\theta}(\zeta, \theta) := S\left(D(\zeta, \theta)\right). \tag{5}$$

Given the data $\boldsymbol{D} = D(\boldsymbol{\zeta}, \theta)$ we get the random estimate

$$\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{\zeta}, \theta) = S(D(\boldsymbol{\zeta}, \theta)). \tag{6}$$

Figure 2 illustrates the procedure above leading to Equation (6):

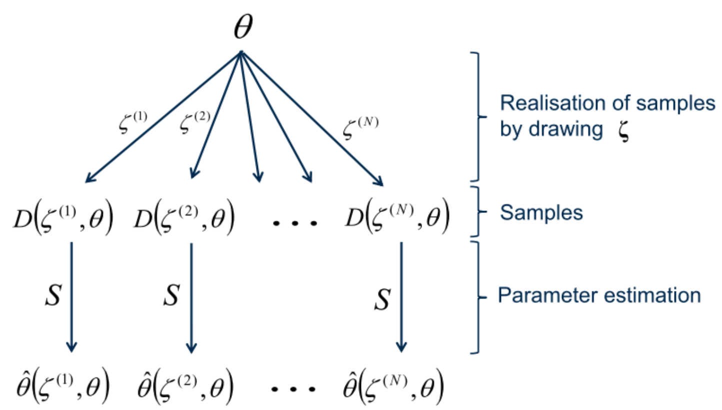

Figure 2: Process of parameter estimation

We start with the unknown parameter θ resp. the random variable $\boldsymbol{X}(\theta)$. A fixed historical observation corresponds to a fixed sample $\zeta = (\zeta_1, \ldots, \zeta_n)$ drawn from $\boldsymbol{\zeta}$. This fixed sample ζ determines the data $D = D(\zeta, \theta)$. Applying the estimation method S leads to the estimated parameter $\hat{\theta} = \hat{\theta}(\zeta, \theta)$.

Note that (6) models the estimate as a random variable given the true parameter θ, i.e. Equation (6) represents the "theoretical perspective" (cf. Section

4.1).
Recall Definition 1.2 and the intuitive arguments of Section 4.1: Our goal is to find a transformation allowing to change from the theoretical perspective into the economically relevant perspective from the undertaking where the estimate $\hat{\theta}_0$ is fixed and there is uncertainty about the true parameter θ.
Note that from the undertaking's perspective the uncertainty about the true parameter θ is due to the fact that the realization of the "historical risk factor" $\zeta \in [0;1]^n$ is unknown. Consider Figure 2 resp. Equation (6) above: Intuitively, if the undertaking knew the realization of the historical risk factor $\zeta \in [0;1]^n$, it would be able to conclude the value of the true parameter θ from the observed estimate $\hat{\theta}_0$. This means that changing into the undertaking's perspective is equivalent to the inversion of $\hat{\theta}(\zeta, \cdot)$ for every fixed ζ.
More precisely: Given the fixed estimate $\hat{\theta}_0$ observed by the undertaking, we are looking for a solution of the equation

$$\hat{\theta}(\zeta, \cdot) = \hat{\theta}_0. \tag{7}$$

Note that Equation (7) is only a slight modification of Equation (4): We just replace the real number $\varsigma \in [0;1]$, representing a "confidence level" or a "single risk factor" of the parameter estimate, by the vector $\zeta \in [0;1]^n$ representing the n "historical risk factors" determining the sample D of size n.
Solving (7) for every fixed $\zeta \in [0;1]^n$ and for every right hand side means that we have to invert the function $\mathfrak{h}_\zeta : I \to I$ given by

$$\mathfrak{h}_\zeta(\theta) := \hat{\theta}(\zeta, \theta).$$

In the sequel we assume that $\mathfrak{h}_\zeta$ is invertible. Then $\mathfrak{h}_\zeta^{-1}(\hat{\theta}_0)$ is the solution of (7).
Thus we define the modelled risk by the following two step procedure:

1. For a $[0;1]^n$ uniformly distributed random vector $\boldsymbol{\zeta}' = (\boldsymbol{\zeta}'_1, \ldots, \boldsymbol{\zeta}'_n)$ with distribution function given by the copula $\mathcal{C}$ and for a given estimate $\hat{\theta} \in I$ we define
$$\boldsymbol{\theta}_{sim} = \boldsymbol{\theta}_{sim}(\boldsymbol{\zeta}', \hat{\theta}) := \mathfrak{h}_{\boldsymbol{\zeta}'}^{-1}(\hat{\theta}). \tag{8}$$

2. Let $\boldsymbol{\xi}'$ be a $[0;1]$ uniformly distributed random variable independent of $\boldsymbol{\zeta}'$. We write the true risk as $\boldsymbol{X} = X(\boldsymbol{\xi}, \theta)$ and the modelled risk as $\boldsymbol{Y} = X(\boldsymbol{\xi}', \boldsymbol{\theta}_{sim})$ or, more precisely, as
$$\boldsymbol{Y}(\hat{\theta}) := X\left(\boldsymbol{\xi}', \mathfrak{h}_{\boldsymbol{\zeta}'}^{-1}\left(\hat{\theta}\right)\right). \tag{9}$$

Note that the risk factors $\boldsymbol{\xi}'$ and $\boldsymbol{\zeta}'$ generated in the model are independent of the true risk factors $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$.
The following simple examples demonstrate the derivation of the parameter distribution of $\boldsymbol{\theta}_{sim}$ (cf. (7)) and emphasize the practicability of the method:

Example 4.1. We illustrate the method for two explicit distributions. We assume that we are in the special case $\boldsymbol{D}_i = \boldsymbol{X}_i$ where $\boldsymbol{X}_i$ are independent variables with $\boldsymbol{X}_i \sim \boldsymbol{X}$ for $i = 1, \ldots, n$ and $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ is independent of $\boldsymbol{X}$.
Let $\zeta = (\zeta_1, \ldots, \zeta_n)$ and $X_i = X_i(\zeta_i; \theta)$, $i = 1 \ldots, n$, be the observed realizations of the random variables $\boldsymbol{X}_i$, $i = 1, \ldots, n$.

1. Let $\boldsymbol{X}$ be lognormally distributed with parameter $\theta = (\mu, \sigma^2)$. For the maximum likelihood method we have

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i = \frac{1}{n}\sum_{i=1}^{n} \ln X_i(\zeta_i, \theta) = \frac{1}{n}\sum_{i=1}^{n} \mu + \sigma Z_i(\zeta_i) \text{ and}$$
$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln X_i(\zeta_i, \theta) - \hat{\mu})^2 = \frac{\sigma^2}{n}\sum_{i=1}^{n}(Z_i(\xi) - \overline{Z})^2 \tag{10}$$

where $Z_i(\zeta_i) := F_{Z_i}^{-1}(\zeta_i)$, $i = 1, \ldots, n$ are realizations of the standard normally distributed random variables $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$. Equation (10) corresponds to the general Equation (6).
The function $\mathfrak{h}_\zeta(\theta) = \mathfrak{h}_\zeta(\mu, \sigma^2)$ is given by the right hand side of (10). Inverting $\mathfrak{h}_\zeta$ means that we have to solve the Equation (10) for the true parameters (μ, σ^2). Hence

$$\mathfrak{h}_\zeta^{-1}(\hat{\mu}, \hat{\sigma}^2) = \left(\hat{\mu} - \frac{\hat{\sigma} \cdot \sqrt{n} \cdot \overline{Z}}{\sqrt{\sum(Z_i - \overline{Z})^2}}, \frac{\hat{\sigma}^2 \cdot n}{\sum(Z_i - \overline{Z})^2}\right)$$

where $Z_i = Z_i(\zeta_i)$ and $\overline{Z} = \frac{\sum Z_i}{n}$.
Thus according to (8):

$$\boldsymbol{\theta}_{sim} = (\boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}^2_{sim}) = \mathfrak{h}_{\zeta'}^{-1}(\hat{\mu}, \hat{\sigma}^2) = \left(\hat{\mu} - \frac{\hat{\sigma} \cdot \sqrt{n} \cdot \overline{\boldsymbol{Z}}}{\sqrt{\boldsymbol{M}}}, n \cdot \frac{\hat{\sigma}^2}{\boldsymbol{M}}\right)$$

where $\overline{\boldsymbol{Z}}$ is normally distributed with mean 0 and standard deviation $\frac{1}{\sqrt{n}}$ and $\boldsymbol{M}$ is χ^2-distributed both with $n-1$ degrees of freedom. Since

$\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\sigma}}^2$ are independent random variables (see [12], p. 214-216), the random variables $\overline{\boldsymbol{Z}}$ and $\boldsymbol{M}$ are independent.

2. Let $\boldsymbol{X}$ be a Pareto distributed random variable with parameter θ and distribution function $F_{\boldsymbol{\theta}}(x) = 1 - x^{-\theta}$ for $x > 1$. For the maximum likelihood method we have

$$\hat{\theta} = \frac{n}{\sum \ln X_i(\zeta_i, \theta)} = \frac{n}{\sum \ln X_i(\zeta_i, \theta)^{\theta}} \cdot \theta \sim \frac{n}{\sum \ln Z_i(\zeta_i)} \cdot \theta \qquad (11)$$

where $Z_i(\zeta_i) = F_1^{-1}(\zeta_i)$ is a realization of a Pareto distributed random variable with parameter $\theta = 1$. We have to solve the equation $\mathfrak{h}_{\zeta}(\theta) = \hat{\theta}$ where $\mathfrak{h}_{\zeta}(\theta)$ is given by the right hand side of (11). The function $\mathfrak{h}_{\zeta}$ is invertible with

$$\mathfrak{h}_{\zeta}^{-1}(\hat{\theta}) = \frac{\sum \ln F_1^{-1}(\zeta_i)}{n} \cdot \hat{\theta}.$$

Thus

$$\boldsymbol{\theta}_{sim} = \mathfrak{h}_{\boldsymbol{\zeta}'}^{-1}(\theta) = \frac{\sum \ln F_1^{-1}(\boldsymbol{\zeta}'_i)}{n} \cdot \hat{\theta}.$$

Note that both distributions - the lognormal and the Pareto distribution - are special cases of transformed location-scale families (see Appendix A.1).

We now introduce an assumption which enables us to prove that the probability distribution of $\boldsymbol{\theta}_{sim}$ is appropriate in the sense of Definition 2.4.
Note that $F_{\boldsymbol{X}}(\boldsymbol{X})$ is a $[0;1]$ uniformly distributed random variable. In particular, it is independent of the parameter θ. This motivates the question whether $F_{\boldsymbol{Y}}(\boldsymbol{X})$ where $\boldsymbol{X}$ is the true risk and $\boldsymbol{Y}$ is the modelled risk is independent of θ.

Property 4.2. With the notation introduced above we say that the model satisfies the **property of parameter invariance** if and only if for all $\zeta \in [0;1]^n$, $\xi \in [0;1]$ and $\theta \in I \subseteq \mathbb{R}^d$

$$F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(F^{-1}_{\boldsymbol{X}(\theta)}(\xi)\right)$$

does not depend on θ.

This property is satisfied for a wide class of distributions and important estimation methods used in practice: In Appendix A.2 we prove that Property 4.2 is satisfied if $\boldsymbol{X}$ and the data $\boldsymbol{D}_i$ are mutually independent random variables belonging to transformed location-scale families and if we use one of the following estimation methods S

- maximum likelihood estimation,
- percentile matching,
- Bayesian estimation with prior $\pi(\mu, \sigma) = \sigma^{-v}, v \geq 0$

under reasonable restrictions necessary for the application of these methods. Moreover, we prove Property 4.2 for location-scale families together with the method of moments or the unbiased parameter estimation. See Proposition A.9 in Appendix A.2 for details. In all these cases, the parameter distribution $(\boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}_{sim})$ can be given explicitly (cf. Corollary A.7 in Appendix A.2).
The following theorem shows that under the assumption that the model satisfies Property 4.2 given above the method introduced in this section is appropriate in the sense of Definition 2.4:

Satz 4.3. *With the notation introduced above we assume that the model satisfies the property of parameter invariance. Let* $0 < \alpha < 1$ *and for a given estimate* $\hat{\theta} \in I$ *let*

$$SCR(\alpha; \hat{\theta}) := F^{-1}_{\boldsymbol{Y}(\hat{\theta})}(\alpha),$$

where $\boldsymbol{Y}(\hat{\theta})$ *is defined by (9). Then*

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \hat{\boldsymbol{\theta}})\right) = \alpha.$$

where the random estimate $\hat{\boldsymbol{\theta}}$ *is defined by (6).*
In other words, the method introduced above resp. the distribution of $\boldsymbol{\theta}_{sim}$ *is appropriate in the sense of Definition 2.4.*

Proof. Let $\hat{\theta}_{fix} \in I$ be a fixed parameter independent of the observed sample (e.g. take $\hat{\theta} = (\mu_{fix}, \sigma_{fix}) = (0, 1)$ for the normal distribution). Let $0 < \zeta, \xi < 1$ be arbitrary but fixed, i.e. we first consider all random variables conditioned upon the event $\{(\boldsymbol{\zeta}, \boldsymbol{\xi}) = (\zeta, \xi)\}$.
We consider the parameter

$$\tilde{\theta} := \mathfrak{h}_{\zeta}^{-1}(\hat{\theta}_{fix}). \tag{12}$$

We have

$$\hat{\theta}(\zeta, \tilde{\theta}) = \mathfrak{h}_{\zeta}(\mathfrak{h}_{\zeta}^{-1}(\hat{\theta}_{fix})) = \hat{\theta}_{fix}. \tag{13}$$

According to the property of parameter invariance, equations (12) and (13) and the definition of $\boldsymbol{Y}$ (see Equation (9))

$$\begin{aligned} F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(X(\xi, \theta)\right) &= F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\tilde{\theta}))}\left(X(\xi, \tilde{\theta})\right) \\ &= F_{\boldsymbol{Y}(\hat{\theta}_{fix})}\left(X(\xi, \mathfrak{h}_{\zeta}^{-1}(\hat{\theta}_{fix}))\right) \\ &= F_{\boldsymbol{X}(\boldsymbol{\xi}', \mathfrak{h}_{\boldsymbol{\zeta}'}^{-1}(\hat{\theta}_{fix}))}\left(X(\xi, \mathfrak{h}_{\zeta}^{-1}(\hat{\theta}_{fix}))\right). \end{aligned}$$

Substituting ζ und ξ by the uniformly distributed random variable $\boldsymbol{\zeta}$, $\boldsymbol{\xi}$ we get

$$F_{\boldsymbol{Y}(\hat{\boldsymbol{\theta}})}(\boldsymbol{X}) \sim F_{\boldsymbol{X}(\boldsymbol{\xi}', \mathfrak{h}^{-1}_{\boldsymbol{\zeta}'}(\hat{\theta}_{fix}))}\left(X(\boldsymbol{\xi}, \mathfrak{h}^{-1}_{\boldsymbol{\zeta}}(\hat{\theta}_{fix}))\right).$$

Since

$$X(\boldsymbol{\xi}', \mathfrak{h}^{-1}_{\boldsymbol{\zeta}'}(\hat{\theta}_{fix})) \sim X(\boldsymbol{\xi}, \mathfrak{h}^{-1}_{\boldsymbol{\zeta}}(\hat{\theta}_{fix}))$$

the right hand side is uniformly distributed. The assertion follows:

$$P\left(\boldsymbol{X} \leq SCR(\alpha; \hat{\boldsymbol{\theta}})\right) = P(F_{\boldsymbol{Y}(\hat{\boldsymbol{\theta}})}(\boldsymbol{X}) \leq \alpha) = \alpha.$$

This completes the proof. □

We demonstrate our method presenting a simple application to premium risk.

Example 4.4. (Premium risk) Consider the following simple model for the losses of the undertaking due to the premium risk for some specific line of business:

$$X = P \cdot S + K - P.$$

Assume that the amount of premiums $P = 50.000.000$ € and the fixed costs $K = 9.000.000$ € are known but the expense ratio S is lognormally distributed with unknown parameters (μ, σ).
For the sample $(S_1, \ldots, S_{10}) = (71\%, 84\%, 78\%, 67\%, 70\%, 75\%, 89\%, 68\%, 80\%, 72\%)$ of historical expense ratios the undertaking estimates the parameter by the maximum likelihood method and obtains

$$(\hat{\mu}_0, \hat{\sigma}_0) = (-0.2864, 0.0892).$$

We can write

$$\boldsymbol{X} = P \cdot \exp(\mu + \sigma \boldsymbol{Z}) + K - P = h(\mu + \sigma \boldsymbol{Z})$$

with the increasing function $h(z) = P \cdot \exp(\mu + \sigma z) + K - P$ and $\boldsymbol{Z} \sim N(0; 1)$. Hence $\boldsymbol{X}$ belongs to the transformed location-scale family $\{h(\mu + \sigma \boldsymbol{Z}) | \mu \in \mathbb{R}, \sigma > 0\}$. The probability distribution for $\boldsymbol{\theta}_{sim} = (\boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}_{sim})$ is given by (cf. Appendix A.2, Corollary A.7).

$$\begin{aligned}
\boldsymbol{\mu}_{sim} &= \hat{\mu}_0 - \frac{\hat{\mu}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)} \cdot \hat{\sigma}_0, \\
\boldsymbol{\sigma}_{sim} &= \frac{\hat{\sigma}_0}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}.
\end{aligned}$$

Since in our example the $\boldsymbol{Z}_i$ are $N(0;1)$-distributed, we get

$$\begin{aligned}\boldsymbol{\mu}_{sim} &\sim -0.2864 - 0.0892 \cdot \frac{\sqrt{10} \cdot \overline{\boldsymbol{Z}}}{\sqrt{\boldsymbol{M}}} \text{ and} \\ \boldsymbol{\sigma}_{sim} &\sim \sqrt{10} \cdot \frac{0.0892}{\sqrt{\boldsymbol{M}}},\end{aligned}$$

where $\overline{\boldsymbol{Z}}$ is normally distributed with mean 0 and standard deviation $\frac{1}{\sqrt{10}}$, $\boldsymbol{M}$ is χ^2-distributed with 9 degrees of freedom and $\overline{\boldsymbol{Z}}$ and $\boldsymbol{M}$ are independent (cf. Example 4.1).
The modelled premium risk $\boldsymbol{Y}$ is defined according to (9) as the predictive distrinution of $\boldsymbol{X}$ with random parameter distribution of $\boldsymbol{\mu}_{sim}$, $\boldsymbol{\sigma}_{sim}$ defined above i.e.

$$\begin{aligned}\boldsymbol{Y} &= P \cdot S(\boldsymbol{\xi}'; \boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}_{sim}) + K - P \\ &= 50.000.000 \cdot S(\boldsymbol{\xi}'; -0.2864 - 0.0892 \cdot \frac{\sqrt{10} \cdot \overline{\boldsymbol{Z}}}{\sqrt{\boldsymbol{M}}}, \frac{\sqrt{10} \cdot 0.0892}{\sqrt{\boldsymbol{M}}}) - 41.000.000\end{aligned}$$

where $S(\alpha; \mu, \sigma) = \exp(\mu + \sigma \cdot F_{\boldsymbol{Z}}^{-1}(\alpha))$ is the α-quantile of $\boldsymbol{S}$ and ξ' is $[0;1]$ uniformly distributed and independent of $\boldsymbol{\mu}_{sim}$ and $\boldsymbol{\sigma}_{sim}$.
By Proposition A.9 in Appendix A.2 the model satisfies Property 4.2. Thus Theorem 4.3 tells us that the parameter distribution defined above is appropriate in the sense of Definition 2.4.
For the given estimate $(\hat{\mu}_0, \hat{\sigma}_0) = (-0.2864, 0.0892)$ we get a modelled SCR of $F_{\boldsymbol{Y}}^{-1}(0.995) = 10.77$ Mio. €. This is significantly higher than the modelled risk capital without parameter risk $F_{\boldsymbol{X}(\hat{\mu},\hat{\sigma})}^{-1}(0.995) = 6.25$ Mio. €.

4.4 Relationship to confidence intervals

The question about parameter uncertainty from the undertaking's perspective closely resembles the question "What can one say about the true parameter given the value of the estimated parameter?". This is the typical question in the context of constructing confidence intervals for an unknown true parameter. Indeed, in the 1-dimensional case the parameter distribution according to the fiducial inference method can equivalently be deduced from the distribution of the bound of a "one-sided left confidence interval" by assuming the confidence level to be $[0;1]$ uniformly distributed.
In the sequel we explain the relationship to confidence intervals using our notation introduced in Subsection 4.3. In particular, we show that in the 1-dimensional case Fisher's original fiducial parameter distribution introduced in Section 4.2 is equal to our parameter distribution introduced in Section

4.3, Equation (8) .
Let θ, $\hat{\theta} \in I \subset \mathbb{R}$, where $I = \mathbb{R}$ or $I = \mathbb{R}_+$. Note that the cumulative distribution function $\hat{F}(\cdot|\theta)$ of $\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{\zeta}, \theta)$ is given by

$$\hat{F}(t|\theta) = P(\hat{\theta}(\boldsymbol{\zeta}, \theta) \leq t) = P(\mathfrak{h}_{\boldsymbol{\zeta}}(\theta) \leq t). \tag{14}$$

We make the reasonable assumption that $\mathfrak{h}_\zeta(\theta) = \hat{\theta}(\zeta, \theta)$ is strictly increasing. According to (14) this is equivalent to $\hat{F}(t|\theta)$ is monotonically decreasing in θ for all $t \in \mathbb{R}$. This is equivalent to $f_\alpha(\cdot) := \hat{F}^{-1}(\alpha|\cdot)$ is strictly increasing in θ for $0 \leq \alpha \leq 1$. A confidence interval for θ given $\hat{\theta}$ with confidence level α is given by $I(\alpha; \hat{\theta}) := [f_\alpha^{-1}(\hat{\theta}); \infty)$ since

$$P(\theta \in I(\alpha, \hat{\boldsymbol{\theta}})) = P(f_\alpha^{-1}(\hat{\boldsymbol{\theta}}) \leq \theta) = P(\hat{\boldsymbol{\theta}} \leq \hat{F}^{-1}(\alpha|\theta)) = \alpha.$$

Thus there is an equivalent, alternative definition of the simulated parameter $\boldsymbol{\theta}_{sim}(\hat{\theta}) = \mathfrak{h}_{\boldsymbol{\zeta}}^{-1}(\hat{\theta})$ using the lower bound of the confidence interval $I(\alpha, \hat{\theta})$:

Proposition 4.5. Let $\tilde{\boldsymbol{\theta}}_{sim}(\hat{\theta}) := f_{\boldsymbol{\varsigma}}^{-1}(\hat{\theta})$ with $\boldsymbol{\varsigma}$ $[0; 1]$ uniformly distributed and let $\mathfrak{h}_{\boldsymbol{\zeta}}$ be strictly increasing. It follows that

$$\tilde{\boldsymbol{\theta}}_{sim}(\hat{\theta}) \sim \boldsymbol{\theta}_{sim}(\hat{\theta}).$$

Moreover, the distribution of $\boldsymbol{\theta}_{sim}$ is equal to the fiducial distribution defined in Section 4.2 based on Fishers fiducial argument.

Proof. For all $x \in I$:

$$\begin{aligned} P(\tilde{\boldsymbol{\theta}}_{sim}(\hat{\theta}) \leq x) &= P(f_{\boldsymbol{\varsigma}}^{-1}(\hat{\theta}) \leq x) \\ &= P(\hat{\theta} \leq \hat{F}^{-1}(\boldsymbol{\varsigma}|x)) = P(\hat{F}(\hat{\theta}|x) \leq \boldsymbol{\varsigma}) \\ &= 1 - \hat{F}(\hat{\theta}|x) = P(\hat{\theta} \leq \hat{\theta}(\boldsymbol{\zeta}, x)) \\ &= P(\mathfrak{h}_{\boldsymbol{\zeta}}^{-1}(\hat{\theta}) \leq x) = P(\boldsymbol{\theta}_{sim}(\hat{\theta}) \leq x). \end{aligned} \tag{15}$$

By the definition of f_α we get $\hat{F}^{-1}(\boldsymbol{\varsigma}|f_{\boldsymbol{\varsigma}}^{-1}(\hat{\theta}_0)) = \hat{\theta}_0$. This implies $\boldsymbol{\varsigma} = \hat{F}(\hat{\theta}_0|f_{\boldsymbol{\varsigma}}^{-1}(\hat{\theta}_0))$ resp. $\tilde{\boldsymbol{\theta}}_{sim} \sim f_{\boldsymbol{\zeta}}^{-1}(\hat{\theta}_0) = \hat{F}^{-1}(\hat{\theta}_0|\cdot)(\boldsymbol{\varsigma})$. Note that $\hat{F}^{-1}(\hat{\theta}_0|\cdot)(\boldsymbol{\varsigma})$ describes Fisher's fiducial distribution (cf. Section 4.2). Since $\boldsymbol{\theta}_{sim} \sim \tilde{\boldsymbol{\theta}}_{sim}$, the proof is complete. □

The implication for practical applications: As soon as we can construct confidence intervals $I(\alpha; \hat{\theta}) = [B(\alpha; \hat{\theta}); \infty)$ with $P\Big(\theta \in I(\alpha, \hat{\boldsymbol{\theta}})\Big) = \alpha$ we can use the lower bound $B(\cdot, \hat{\theta})$ to simulate $\boldsymbol{\theta}_{sim}$ by setting $\boldsymbol{\theta}_{sim}(\hat{\theta}) := B(\boldsymbol{\varsigma}; \hat{\theta})$.

Example 4.6. Let $\boldsymbol{X}$ be exponentially distributed with density function given by $f(x) = \lambda \exp(-\lambda x)$. Then an exact $(1-\alpha)$-confidence interval for λ is given by $\left[\frac{\chi^2_{2n}(\alpha)}{2n\overline{x}}, \infty\right)$ where χ^2_{2n} is the α-quantile of the χ^2-distribution with $2n$ degrees of freedom. Hence by Proposition 4.5

$$\boldsymbol{\lambda}_{sim} \sim \frac{\boldsymbol{S}}{2n\overline{x}}$$

where $\boldsymbol{S}$ follows a χ^2_{2n} distribution.

The following corollary gives an alternative definition for the distribution of the parameter $\boldsymbol{\theta}_{sim}$ introduced in Subsection 4.3.

Corollary 4.7. Let $\hat{F}(\cdot|\theta)$ be the distribution function of the estimate $\hat{\boldsymbol{\theta}}$ given the true parameter θ. We assume that $f_\alpha(\cdot)$ resp. $\hat{\theta}(\zeta, \cdot)$ is strictly increasing in θ and continuously invertible for all $0 < \alpha < 1$ resp. for all ζ. Furthermore, let $F_{\boldsymbol{\theta}_{sim}}$ be the distribution function of the simulated parameter $\boldsymbol{\theta}_{sim} = \boldsymbol{\theta}_{sim}(\hat{\theta}_0)$. It follows that

$$F_{\boldsymbol{\theta}_{sim}}(x) = 1 - \hat{F}(\hat{\theta}_0|x).$$

Proof. The assertion follows from (15). □

5 Illustration of the impact on risk capital calculation

Consider a sample of size $n = 10$ drawn from a lognormally distributed loss variable $\boldsymbol{X}$:

$$\{150.01, 152.33, 120.47, 131.87, 139.07, 157.97, 128.37, 122.89, 166.47, 133.18\}.$$

The true parameters (μ, σ) are unknown to the undertaking. It estimates the parameters of the distribution using the maximum likelihood method and finds $(\hat{\mu}, \hat{\sigma}) = (4.94, 0.10)$. We derive the following values:

Risk capital			
without parameter risk	with non-parametric bootstrapping	with parametric bootstrapping	with the fiducial inference approach
182.65	184.25	187.24	203.06

Let us increase the volatility of the sample. We now consider the slightly modified sample given by

$$\{150.01, 182.10, 120.47, 211.50, 139.07, 157.97, 199.35, 122.89, 166.47, 133.18\}.$$

In this case the undertaking gets $(\hat{\mu}, \hat{\sigma}) = (5.04, 0.19)$. The results are given as follows:

Risk capital			
without parameter risk	with n non-parametric bootstrapping	with parametric bootstrapping	with the fiducial inference approach
250.93	256.06	263.19	302.90

Note that by Theorem 4.3, the risk capital using the fiducial inference is the solvency capital required to satisfy Definition 1.1 resp. Definition 2.4. Hence, the risk capital without the consideration of parameter risk and the risk capital generated using bootstrapping both underestimate the capital requirement significantly.

6 Summary and Outlook

This contribution deals with parameter uncertainty in risk capital calculation. After introducing a criterion known from the theory of predictive inference which allows to assess the methods used for risk capital calculations, we discussed the approaches previously proposed in the literature. In particular, we showed that the popular and widely used bootstrapping approach is not satisfactory.
We then proposed a new method to model the parameter uncertainty which has the following advantages:

- It measures the parameter uncertainty from the undertaking's perspective, i.e. it models the uncertainty about the true parameter given an estimate based on historical data.
- Our method does not require the specification of an a priori distribution.
- We introduced a sufficient condition (Property 4.2) under which the method appropriately models the parameter risk in the following sense: Taking into account the randomness of the historical data, the (random) Value at Risk (SCR) for the confidence level α, which is modelled based on the random historical data, will not be exceeded by the loss of the next year with a probability exactly equal to α.

- We proved Property 4.2 for transformed location-scale families together with a number of estimation methods including the maximum likelihood method.

- Under these assumptions our method is easy to apply in practice. The parameter distribution can be obtained by a simple Monte Carlo simulation (see Corollary A.7).

- Our method describes the economic risk of the undertaking, in other words the required amount of own funds, taking into account parameter uncertainty resulting from the randomness of historical data.

Our idea is not restricted to non-life insurance but can be applied to any kind of risk capital calculations where parameter uncertainty cannot be neglected. Some questions are left open for future research, e.g.

- If Equation (7) can not be solved analytically, we need an efficient numerical algorithm to solve the equation.

- We require that $\mathfrak{h}_\zeta$ is invertible. Can this assumption be weakened resp. is there a method which works if $\mathfrak{h}_\zeta$ is not invertible?

7 Acknowledgments

We are indebted to the referees whose thorough reviews and many suggestions helped us in significantly improving the paper.

A Appendix

In this section we prove Property 4.2 for a wide class of distributions. More precisely, we prove Property 4.2 for distribution families which are a member of the class of one- or two-parameter transformed location-scale families and for the maximum likelihood estimation method, the Bayesian estimation with prior of the form $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$ and the percentile matching estimation method. For location-scale families we additionally prove Property 4.2 for the method of moments and the unbiased parameter estimation.

A.1 Transformed location-scale families

Let us recall the definition of a transformed location-scale family (cf. [19], Section 5.2).

Definition A.1. A family $\mathcal{F}_{LS} = \left\{\boldsymbol{X}(\theta)|\theta \in I \subset \mathbb{R}^d\right\}$ of univariate probability distributions is called a **location-scale family** if for every $\boldsymbol{X}(\theta_1) \in \mathcal{F}_{LS}$ and $\boldsymbol{X}_2(\theta_2) \in \mathcal{F}_{LS}$ we can write $\boldsymbol{X}_2 \stackrel{d}{=} a\boldsymbol{X}_1 + b$ for some $a, b \in \mathbb{R}$ with $a > 0$.

Remark A.2.

1. Let $\boldsymbol{Z}$ be an arbitrary random variable. Then we can construct a location-scale family by

$$\{\mu + \sigma\boldsymbol{Z} : \mu, \sigma \in \mathbb{R}, \sigma > 0\}.$$

   We can assume that every location-scale family $\mathcal{F}_{LS}$ is given by random variables of the form $\mu + \sigma\boldsymbol{Z}$.

2. If we have given a location-scale family we can choose a particular element $F \in \mathcal{F}_{LS}$ and call this the "standard distribution" of the family $\mathcal{F}_{LS}$. For every $\boldsymbol{X}$ we can write $X \stackrel{d}{=} \mu + \sigma\boldsymbol{Z}$ with $\boldsymbol{Z} \sim F$ and $\sigma > 0$. We write $\boldsymbol{X} \sim F(\cdot;\mu,\sigma)$ with $F(x;\mu,\sigma) = F(\frac{x-\mu}{\sigma})$. Note that $\boldsymbol{Z} \sim F(\cdot;0,1)$.
   We can assume that $\boldsymbol{Z}$ has mean $E\boldsymbol{Z} = 0$ and variance $\mathrm{Var}\boldsymbol{Z} = 1$.

3. Examples of two-parameter location-scale families are the normal distribution or the t-distribution. Both the exponential distribution and the Gamma distribution with known shape parameter are examples of one-parameter scale families.

Definition A.3. Let $\mathcal{F}_{LS}$ be a location-scale family. A set $\mathcal{F}_{LS}^h = \left\{\boldsymbol{X}(\theta)|\theta \in I \subset \mathbb{R}^d\right\}$ of univariate probability distributions is called a **transformed location-scale family** if there exists a strictly increasing function h such that for any $X \in \mathcal{F}_{LS}^h$ we have $h^{-1}(X) \in \mathcal{F}_{LS}$.

Remark A.4.

1. For a random variable $\boldsymbol{X}$ from a transformed location-scale family we have $\boldsymbol{X} \sim F(\cdot;\mu,\sigma;h)$ where $F(x;\mu,\sigma;h) = F\left(\frac{h^{-1}(x)-\mu}{\sigma}\right)$.

2. Examples of transformed location-scale families are the lognormal distribution, the Pareto distribution, the log-t-distribution, the log-logistic distribution, the log-Laplace distribution, the Weilbull distribution and the Gumbel distribution (cf. eg. [11], Table II).

A.2 Property 4.2 for transformed location-scale families

Additionally to the maximum likelihood method (MLE) we consider the following estimation methods:

Percentile matching (PE)

Let $x_1, \dots, x_n$ be a sample of independent realizations drawn from $\boldsymbol{X}(\theta)$ and let $x(1) \leq \dots \leq x(n)$ be the order statistics from the sample. A smoothed empirical p-quantile can be defined by (cf. [13])

$$\hat{\pi}(p) := (1-\lambda)x(j) + \lambda x(j+1),\ j = [(n+1)p],\ \lambda = (n+1)p - j.$$

Choosing quantiles $p_1, \dots, p_d$ a percentile matching estimate $\hat{\theta} \in \mathbb{R}^d$ is a solution of the equations

$$F^{-1}_{\boldsymbol{X}(\hat{\theta})}(p_k) = \hat{\pi}(p_k), \quad k = 1, \dots, d.$$

For the sake of simplicity but without loss of generality, we choose the percentiles in the form $p_k = \frac{m_k}{n+1}, m_k \in \mathbb{N}$ such that $\hat{\pi}(p_k) = x(m_k)$.
Note that the PE does not make sense in the general setting where the data $\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta)$ may have distributions different from the distribution of $\boldsymbol{X}$. Therefore, we assume that $\boldsymbol{D}_1(\theta) = \boldsymbol{X}_1, \dots, \boldsymbol{D}_n(\theta) = \boldsymbol{X}_n$ are independent, identically distributed realizations of $\boldsymbol{X}$ if PE is used as the estimation method.

Bayesian estimation

The parameter is estimated as the mean of the posterior distribution, i.e.

$$\hat{\theta} = \int \theta \cdot \pi(\theta|d) d\theta.$$

The posterior distribution given the sample $d = (d_1, \dots, d_n)$ drawn from $\boldsymbol{D}(\theta) = (\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta))$ is given by

$$\pi(\theta|d) = \frac{f(d|\theta)\pi(\theta)}{\int f(d|\theta)\pi(\theta)d\theta}$$

with the joint probability density function $f(\cdot, \theta)$ of $\boldsymbol{D}(\theta)$ and the prior distribution π.

The method of moments and the unbiased parameter estimation for location-scale families

Consider the independent random variables $\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta)$ belonging to (possible different) location-scale families $\mathfrak{I}_i := \{\mu + \sigma \boldsymbol{Z}_i | \mu \in \mathbb{R}, \sigma > 0\}$, $\boldsymbol{Z}_i$ independent of (μ, σ) but with the same parameter $\theta = (\mu, \sigma)$. Without loss

of generality we assume that $E\boldsymbol{Z}_i = 0$ and $\mathrm{Var}\boldsymbol{Z}_i = 1$.
Let $d = (d_1, \dots, d_n)$ be a sample drawn from $(\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta))$. The method of moments estimates the parameters μ and σ by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} d_i = \overline{d} \text{ and } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2.$$

Additionally, we define the unbiased parameter estimation by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} d_i = \overline{d} \text{ and } \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2.$$

Note that in this case $\hat{\sigma}^2$ is the sample variance.

Lemma A.5. a) The MLE is invariant under differentiable increasing transformations, i.e. let $d_1, \dots, d_n$ be drawn from independent random variables $\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta)$ corresponding to parametric distribution families $\mathfrak{I}_i := \{\boldsymbol{D}_i(\theta) | \theta \in I \subset \mathbb{R}^d\}$, $i = 1 \dots, n$. Let $S(d_1, \dots, d_n)$ be the corresponding MLE. For differentiable increasing functions h_i consider the transformed date $d_i^* = h_i(d_i)$ and the transformed distribution families $\mathfrak{I}_i^* := \{\boldsymbol{h}_i(D(\theta)) | \theta \in I \subset \mathbb{R}^d\}$, $i = 1, \dots, n$. Let $S^*(d_1^*, \dots, d_n^*)$ be the corresponding MLE. Then

$$S^*(d_1^*, \dots, d_n^*) = S(d_1, \dots, d_n).$$

b) Let $D^*(\zeta, \theta) = D^*((\zeta_1, \dots, \zeta_n), \theta) := (h_1(D_1(\zeta_1, \theta)), \dots, h_n(D_n(\zeta_n, \theta)))$. Under the assumption of a) let $\mathfrak{h}_{\boldsymbol{\zeta}}(\theta) := S\left(D(\boldsymbol{\zeta}, \theta)\right)$ resp. $\mathfrak{h}^*_{\boldsymbol{\zeta}}(\theta) := S^*\left(D^*(\boldsymbol{\zeta}, \theta)\right)$. Then the fiducial parameters $\boldsymbol{\theta}_{\boldsymbol{sim}} := \mathfrak{h}_{\boldsymbol{\zeta}}^{-1}(\hat{\theta}_0)$ resp. $\boldsymbol{\theta}^*_{\boldsymbol{sim}} := \mathfrak{h}_{\boldsymbol{\zeta}}^{*-1}(\hat{\theta}_0)$ are equal, that is, $\boldsymbol{\theta}^*_{sim} = \boldsymbol{\theta}_{sim}$.

c) For independent random variables $\boldsymbol{D}_1(\theta), \dots, \boldsymbol{D}_n(\theta)$ belonging to (possible different) location-scale families $\mathfrak{I}_i := \{\mu + \sigma \boldsymbol{Z}_i | \mu \in \mathbb{R}, \sigma > 0\}$, $\boldsymbol{Z}_i$ independent of (μ, σ) but with the same parameter $\theta = (\mu, \sigma)$. Let $d_1, \dots, d_n$ be a sample drawn from $\boldsymbol{D}_1, \dots, \boldsymbol{D}_n$. Let $(\hat{\mu}, \hat{\sigma})$ be the MLE of (μ, σ). It follows that

$$\hat{\mu}(a + b \cdot d_1, \dots, a + b \cdot d_n) = a + b \cdot \hat{\mu}(d_1, \dots, d_n),$$
$$\hat{\sigma}(a + b \cdot d_1, \dots, a + b \cdot d_n) = b \cdot \hat{\sigma}(d_1, \dots, d_n).$$

The assertions a)-b) of the Lemma hold true for the Bayesian estimate. The assertion c) of the Lemma is true for the Bayesian estimate if we choose the prior to be of the form $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$.

Under the additional assumptions that $\boldsymbol{D}_1(\theta) = \boldsymbol{X}_1, \ldots, \boldsymbol{D}_n(\theta) = \boldsymbol{X}_n$ are i.i.d. realizations of $\boldsymbol{X}$ and that $h = h_1 = \ldots = h_n$ the assertions a)-c) of the Lemma hold also if the PE is used as the estimation method.
Moreover, assertion c) is true for the method of moments and the unbiased parameter estimation as described above.

Remark A.6. Note that the differentiability assumption on the transformations in Lemma A.5 can be dropped for PE.

Proof. First note that a) implies $\mathfrak{h}^*_\varsigma(\theta) = \mathfrak{h}_\varsigma(\theta)$. Hence b) follows from a). Therefore, it remains to prove a) and c):
Let us first consider the maximum likelihood method. In this case:

a) is Lemma 1 (i) in [11], Appendix A.1.

c) is Lemma 3 in [11], Appendix A.1.

For the Bayesian estimate:

a) is Lemma 1 (ii) in [11], Appendix A.1.

c) By Lemma 4 in [11], Appendix A.1

$$\pi(\mu, \sigma | a + b \cdot d) = \frac{1}{b^2} \cdot \pi \left(\frac{\mu - a}{b}, \frac{\sigma}{b} | d \right)$$

for all $a, b \in \mathbb{R}$, $b > 0$. Thus

$$\begin{aligned}\hat{\mu}(a + b \cdot d) &= \int \mu \cdot \pi(\mu, \sigma | a + b \cdot d) d(\mu, \sigma) = \int \mu \cdot \frac{1}{b^2} \pi \left(\frac{\mu - a}{b}, \frac{\sigma}{b} | d \right) d(\mu, \sigma) \\ &= \int (a + b \cdot \mu) \cdot \pi(\mu, \sigma | d) d(\mu, \sigma) = a + b \cdot \hat{\mu}(d)\end{aligned}$$

and

$$\begin{aligned}\hat{\sigma}(a + b \cdot d) &= \int \sigma \cdot \pi(\mu, \sigma | a + b \cdot d) d(\mu, \sigma) = \int \sigma \cdot \frac{1}{b^2} \pi \left(\frac{\mu - a}{b}, \frac{\sigma}{b} | d \right) d(\mu, \sigma) \\ &= \int b \cdot \sigma \cdot \pi(\mu, \sigma | d) d(\mu, \sigma) = b \cdot \hat{\sigma}(d).\end{aligned}$$

We now consider the PE:

a) Follows from the equivalence of $F^{-1}_{h(\boldsymbol{X}(\theta))}(p_k) = h(x(m_k))$ and $F^{-1}_{\boldsymbol{X}(\theta)}(p_k) = x(m_k)$.

c) Setting $x_i^* = a + bx_i$ the assertion follows from the equation

$$F^{-1}_{a+b\hat{\mu}(x_1,\ldots,x_n)+b\hat{\sigma}(x_1,\ldots,x_n)\boldsymbol{Z}}(p_k) = a + bF^{-1}_{\hat{\mu}(x_1,\ldots,x_n)+\hat{\sigma}(x_1,\ldots,x_n)\boldsymbol{Z}}(p_k)$$
$$= a + bx(m_k) = x^*(m_k).$$

For the method of moments c) follows from the fact that $\overline{a+bd} = a + b\overline{d}$ where $\overline{d}$ resp. $\overline{a+bd}$ denotes the sample mean of $d_1, \ldots, d_n$ resp. $a+bd_1, \ldots, a+bd_n$ and

$$\hat{\sigma}^2(a+bd) = \frac{1}{n}\sum_{i=1}^{n}\left((a+bd_i) - \overline{a+bd}\right)^2 = \frac{b^2}{n}\sum_{i=1}^{n}\left(d_i - \overline{d}\right)^2 = b^2 \cdot \hat{\sigma}^2.$$

□

The proof for the unbiased parameter estimation is analogous.

Corollary A.7. Let the data $\boldsymbol{D}_1(\theta), \ldots, \boldsymbol{D}_n(\theta)$, $\theta = (\mu, \sigma)$, be independent random variables belonging to transformed location-scale families. Set $\mathfrak{I}^{h_i}_{\boldsymbol{D}_i} := \{h_i(\mu + \sigma\boldsymbol{Z}_i) | \mu \in \mathbb{R}, \sigma > 0\}$ for $i = 1, \ldots, n$ for (sufficiently smooth) increasing functions $h_1, \ldots, h_n$ and $\boldsymbol{Z}_i$ random variables independent of $\theta = (\mu, \sigma)$.
Consider the estimation methods S

- MLE,
- PE,
- Bayesian estimation with prior $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$

under the canonical restrictions mentioned in Lemma A.5 necessary for the application of the respective estimation method.
Then the parameter distribution according to the fiducial approach described in Section 4.3 is explicitly given by

$$\boldsymbol{\mu}_{sim} = \hat{\mu}_0 - \frac{\hat{\mu}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)} \cdot \hat{\sigma}_0,$$
$$\boldsymbol{\sigma}_{sim} = \frac{\hat{\sigma}_0}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}$$

where $\hat{\theta}_0 = (\hat{\mu}_0, \hat{\sigma}_0)$ is the estimate for a given observation $(d_1, \ldots, d_n)$ and $\hat{\mu}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ resp. $\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n))$ are the random estimates for the random vector $(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)$ according the chosen estimation method S.

In particular, the function $\mathfrak{h}_\zeta(\theta) := S(D(\boldsymbol{\zeta}, \theta))$ introduced in Setion 4.3 is invertible.
The assertion of the corollary also holds for location-scale families if we use the method of moments or the unbiased parameter estimation.

Proof. Since by Lemma A.5 a), b) the transformations h_i do neither affect the estimate, nor the function $\mathfrak{h}_\zeta$, nor the fiducial parameter $(\boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}_{sim})$, it is sufficient to prove the assertion for location-scale families.
We use the notation $D_i(\zeta_i, \theta) = \mu + \sigma \boldsymbol{Z}_i$, $\boldsymbol{Z}_i = Z_i(\boldsymbol{\zeta}_i) = F_{Z_i}^{-1}(\boldsymbol{\zeta}_i)$ for $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)$ uniformly distributed on $[0;1]^n$ with independent components. It follows from Lemma A.5 c) that

$$h_\zeta(\mu, \sigma) := S\left(D(\zeta), (\mu, \sigma)\right)) = (\mu + \sigma\hat{\mu}\left(Z_1(\zeta_1), \ldots, Z_n(\zeta_n)\right), \sigma\hat{\sigma}(Z_1(\zeta_1), \ldots, Z_n(\zeta_n)))$$

for fixed $\zeta = (\zeta_1, \ldots, \zeta_n) \in [0;1]^n$.
Its inverse is given by

$$\mathfrak{h}_{\boldsymbol{\zeta}}^{-1}\left((\hat{\mu}_0, \hat{\sigma}_0)\right) = \left(\hat{\mu}_0 - \frac{\hat{\mu}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)} \cdot \hat{\sigma}_0, \frac{\hat{\sigma}_0}{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}\right).$$

This yields the assertion since by definition $(\boldsymbol{\mu}_{sim}, \boldsymbol{\sigma}_{sim}) = \mathfrak{h}_{\boldsymbol{\zeta}}^{-1}\left((\hat{\mu}_0, \hat{\sigma}_0)\right)$ for $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_n)$ uniformly distributed on $[0;1]^n$. □

Proposition A.8. The model introduced in Subsection 4.3 is parameter invariant in the sense of Property 4.2 for location-scale families if the MLE , the Bayesian estimation with prior $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$ or the method of moment is used for parameter estimation.
More precisely, let $\boldsymbol{X}(\theta)$ and the data $\boldsymbol{D}_1(\theta), \ldots, \boldsymbol{D}_n(\theta)$, $\theta = (\mu, \sigma)$, be independent random variables belonging to location-scale families $\mathfrak{I}_{\boldsymbol{X}}$ resp. $\mathfrak{I}_{\boldsymbol{D}_1}, \ldots, \mathfrak{I}_{\boldsymbol{D}_n}$ and use MLE, the Bayesian estimate with prior $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$, the method of moments or the unbiased parameter estimation. Let $\boldsymbol{Y}(\hat{\theta}(\zeta, \theta))$ be the modelled risk according to the fiducial approach (cf. Equation 9). Then for all $\xi \in [0;1]$, $\zeta \in [0;1]^n$

$$F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(F_{\boldsymbol{X}(\theta)}^{-1}(\xi)\right) \text{ is independent of } \theta. \tag{16}$$

Under the additional assumption that $\boldsymbol{D}_1(\theta) = \boldsymbol{X}_1, \ldots, \boldsymbol{D}_n(\theta) = \boldsymbol{X}_n$ are i.i.d. realizations of $\boldsymbol{X}$ and $h = h_1 = \ldots = h_n$ the assertion holds also for the estimation method PE.

Proof. Let $\boldsymbol{X}(\mu, \sigma) = \mu + \sigma \boldsymbol{Z}$ and $\boldsymbol{D}_i(\mu, \sigma) = \mu + \sigma \boldsymbol{Z}_i$ with independent random variables $\boldsymbol{Z} \sim F_{\boldsymbol{Z}}$, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ not depending on (μ, σ).

We get

$$
\begin{aligned}
F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\mu,\sigma))}\left(F^{-1}_{\boldsymbol{X}(\mu,\sigma)}(\xi)\right) &= E_{\boldsymbol{\mu}_{sim},\boldsymbol{\sigma}_{sim}}\left[F_{X(\boldsymbol{\mu}_{sim},\boldsymbol{\sigma}_{sim})}\left(F^{-1}_{\boldsymbol{X}(\mu,\sigma)}(\xi)\right)\right] \\
&= E_{\boldsymbol{\mu}_{sim},\boldsymbol{\sigma}_{sim}}\left[F_{\boldsymbol{Z}}\left(\frac{F^{-1}_{\boldsymbol{X}(\mu,\sigma)}(\xi)-\boldsymbol{\mu}_{sim}}{\boldsymbol{\sigma}_{sim}}\right)\right] \\
&= E_{\boldsymbol{\mu}_{sim},\boldsymbol{\sigma}_{sim}}\left[F_{\boldsymbol{Z}}\left(\frac{\mu-\boldsymbol{\mu}_{sim}}{\boldsymbol{\sigma}_{sim}}+\frac{\sigma}{\boldsymbol{\sigma}_{sim}}F^{-1}_{\boldsymbol{Z}}(\xi)\right)\right].
\end{aligned}
\tag{17}
$$

(18)

Let $\hat{\mu}_0 := \hat{\mu}(d_1, \ldots, d_n)$ and $\hat{\sigma}_0 := \hat{\sigma}(d_1, \ldots, d_n)$ be the estimates of μ resp. σ based on the observed data $d_i = h_i(\mu + \sigma z_i)$, $i = 1, \ldots, n$. From Corollary A.7 and Lemma A.5 it follows that

$$
\frac{\sigma}{\boldsymbol{\sigma}_{sim}} = \sigma \cdot \frac{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\hat{\sigma}_0} = \sigma \cdot \frac{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\sigma\hat{\sigma}(z_1, \ldots, z_n)} = \frac{\hat{\sigma}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)}{\hat{\sigma}(z_1, \ldots, z_n)}
$$

and

$$
\begin{aligned}
\frac{\mu-\boldsymbol{\mu}_{sim}}{\boldsymbol{\sigma}_{sim}} &= \frac{\mu-\hat{\mu}_0}{\sigma}\cdot\frac{\sigma}{\boldsymbol{\sigma}_{sim}} + \frac{\hat{\mu}_0-\boldsymbol{\mu}_{sim}}{\boldsymbol{\sigma}_{sim}} \\
&= \hat{\mu}(z_1, \ldots, z_n)\cdot\frac{\sigma}{\boldsymbol{\sigma}_{sim}} + \hat{\mu}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)
\end{aligned}
$$

is independent of (μ, σ). This completes the proof. □

Proposition A.9. The model introduced in Subsection 4.3 is parameter invariant in the sense of Property 4.2 for increasing transformations of location-scale families if the MLE or the Bayesian estimation with prior $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$ is used for parameter estimation.
More precisely: Let the random variable $\boldsymbol{X}(\theta)$ and the data $\boldsymbol{D}_1(\theta), \ldots, \boldsymbol{D}_n(\theta)$, $\theta = (\mu, \sigma)$, be independent random variables belonging to location-scale families $\mathfrak{I}_{\boldsymbol{X}}$ resp. $\mathfrak{I}_{\boldsymbol{D}_1}, \ldots, \mathfrak{I}_{\boldsymbol{D}_n}$.
For an invertible function h and differentiable, increasing functions $h_1, \ldots, h_n$ consider the transformed random variables $\boldsymbol{X}^*(\theta) = h(\boldsymbol{X}(\theta))$ and $\boldsymbol{D}^*_1(\theta) = h_1(\boldsymbol{D}_1(\theta))$,$\ldots$, $\boldsymbol{D}^*_n(\theta) = h_n(\boldsymbol{D}_n(\theta))$.
Let $\boldsymbol{Y}^*(\hat{\theta}^*(\zeta, \theta))$ be the modelled risk according to the fiducial approach (cf. Equation 9) corresponding to both the respective transformed location-scale families of $\boldsymbol{X}^*, \boldsymbol{D}^*_1, \ldots, \boldsymbol{D}^*_n$ and the MLE or the Bayesian estimate with prior $\pi(\mu, \sigma) = \sigma^{-v}$, $v \geq 0$ as estimation method.
Then for all $\xi \in [0; 1]$, $\zeta \in [0; 1]^n$

$$
F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(F^{-1}_{\boldsymbol{X}(\theta)}(\xi)\right) \text{ is independent of } \theta. \tag{19}
$$

Under the additional assumption that $\boldsymbol{D}_1(\theta) = \boldsymbol{X}_1, \ldots, \boldsymbol{D}_n(\theta) = \boldsymbol{X}_n$ are i.i.d. realizations of $\boldsymbol{X}$ and $h = h_1 = \ldots = h_n$ the assertion holds also for the estimation method PE.

Proof. Lemma A.5 b) yields that

$$\boldsymbol{Y}^* = X^*(\boldsymbol{\xi}, \boldsymbol{\theta}^*_{sim}) = X^*(\boldsymbol{\xi}, \boldsymbol{\theta}_{sim}) = h(X(\boldsymbol{\xi}, \boldsymbol{\theta}_{sim})) = h(\boldsymbol{Y})$$

for $\xi \in [0;1]$ uniformly distributed. Note that by Lemma A.5 a) $\hat{\theta}^*(\zeta, \theta) = \hat{\theta}(\zeta, \theta)$ for fixed $\xi \in [0;1]$ and $\zeta \in [0;1]^n$.
Thus

$$\begin{aligned} F_{\boldsymbol{Y}^*(\hat{\theta}^*(\zeta,\theta))}\left(X^*(\xi,\theta)\right) &= F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(h^{-1} \circ X^*(\xi,\theta)\right) \\ &= F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}\left(h^{-1} \circ h \circ X(\xi,\theta)\right) = F_{\boldsymbol{Y}(\hat{\theta}(\zeta,\theta))}(X(\xi,\theta)). \end{aligned}$$

The assertion follows from Proposition A.8. □

B Independence of the bootstrapping failure probability for transformed location-scale families

In this section we prove that, given a random variable $\boldsymbol{X}$ which belongs to a transformed location-scale family, the probability $P\left(\boldsymbol{X} \leq SCR(\alpha; \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; M)\right)$ does not depend on (μ, σ) if the estimation method S is the maximum likelihood method and M is either non-parametric or parametric bootstrapping.
Throughout this section let $\{\mu + \sigma \boldsymbol{Z} : \mu, \sigma \in \mathbb{R}, \sigma > 0\}$ be a location-scale family and let S be the maximum likelihood method. Let $\{\mu + \sigma Z_1, \ldots, \mu + \sigma Z_n\}$ be a fixed sample drawn from $\mu + \sigma \boldsymbol{Z}$. Bootstrapping generates a random sample $\boldsymbol{X}^*_1 = \mu + \sigma \boldsymbol{Z}^*_1, \ldots, \boldsymbol{X}^*_n = \mu + \sigma \boldsymbol{Z}^*_n$ from $\{X_1 = \mu + \sigma Z_1, \ldots, X_n = \mu + \sigma Z_n\}$. This sample leads to parameters $(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$.

Lemma B.1. Let $(\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$ be the (random) parameters obtained by bootstrapping. Then $\frac{\mu - \boldsymbol{\mu}^*}{\boldsymbol{\sigma}^*}$ and $\frac{\sigma}{\boldsymbol{\sigma}^*}$ are independent of (μ, σ).

Proof. 1. For non-parametric bootstrapping we get

$$\begin{aligned} \boldsymbol{\mu}^* &= \frac{\sum \boldsymbol{X}^*_i}{n} = \frac{\sum(\mu + \sigma \boldsymbol{Z}^*_i)}{n} = \mu + \sigma \cdot \frac{\sum \boldsymbol{Z}^*_i}{n} \\ \boldsymbol{\sigma}^{*2} &= \frac{1}{n} \sum (\boldsymbol{X}^*_i - \boldsymbol{\mu}^*)^2 = \frac{1}{n} \sum \left(\mu - \boldsymbol{\mu}^* + \sigma \boldsymbol{Z}^*_i\right)^2 \\ &= \frac{\sigma}{n} \sum \left(\frac{\sum \boldsymbol{Z}^*_i}{n} - \boldsymbol{Z}^*_i\right)^2. \end{aligned}$$

It follows that

$$\frac{\sigma^2}{\boldsymbol{\sigma}^{*2}} = \frac{n}{\sum\left(\frac{\sum \boldsymbol{Z}_i^*}{n} - \boldsymbol{Z}_i^*\right)^2}$$

and

$$\frac{\boldsymbol{\mu}^* - \mu}{\boldsymbol{\sigma}^*} = \frac{\boldsymbol{\mu}^* - \mu}{\sigma} \cdot \frac{\sigma}{\boldsymbol{\sigma}^*} = \frac{\sum \boldsymbol{Z}_i^*}{n} \cdot \frac{\sigma}{\boldsymbol{\sigma}^*}$$

are independent of (μ, σ).

2. For parametric bootstrapping $\boldsymbol{X}_i^*$ are independent random variables with distribution equal to $\boldsymbol{X}(\hat{\mu}, \hat{\sigma})$. We get that

$$\begin{aligned}
\frac{\sigma^2}{\boldsymbol{\sigma}^{*2}} &= \frac{\sigma^2}{\hat{\sigma}^2} \cdot \frac{\hat{\sigma}^2}{\boldsymbol{\sigma}^{*2}} = \frac{\sigma^2}{\frac{1}{n}\sum(X_i - \overline{X})^2} \cdot \frac{\hat{\sigma}^2}{\frac{1}{n}\sum(\boldsymbol{X}_i^* - \overline{\boldsymbol{X}_i^*})^2} \\
&= \frac{\sigma^2}{\frac{\sigma^2}{n}\sum(Z_i - \overline{Z})^2} \cdot \frac{\hat{\sigma}^2}{\frac{1}{n}\sum\left(\hat{\mu} + \hat{\sigma}\boldsymbol{Z}_i^* - (\hat{\mu} + \hat{\sigma}\frac{\sum Z_i}{n})\right)^2} \\
&= \frac{n}{\sum(Z_i - \overline{Z})^2} \cdot \frac{n}{\sum\left(\boldsymbol{Z}_i^* - \frac{\sum Z_i}{n})\right)^2}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\boldsymbol{\mu}^* - \mu}{\boldsymbol{\sigma}^*} &= \frac{\boldsymbol{\mu}^* - \hat{\mu}}{\hat{\sigma}} \cdot \frac{\hat{\sigma}}{\boldsymbol{\sigma}^*} + \frac{\hat{\mu} - \mu}{\sigma} \cdot \frac{\sigma}{\boldsymbol{\sigma}^*} \\
&= \frac{\frac{1}{n}\sum(\hat{\mu} + \hat{\sigma}\boldsymbol{Z}_i^*) - \hat{\mu}}{\hat{\sigma}} \cdot \frac{\hat{\sigma}}{\boldsymbol{\sigma}^*} + \frac{\frac{1}{n}\sum(\mu + \sigma Z_i) - \mu}{\sigma} \cdot \frac{\sigma}{\boldsymbol{\sigma}^*} \\
&= \frac{1}{n}\sum \boldsymbol{Z}_i^* \frac{\hat{\sigma}}{\boldsymbol{\sigma}^*} + \frac{1}{n}\frac{\sigma}{\boldsymbol{\sigma}^*}\sum Z_i,
\end{aligned}$$

are both independent of (μ, σ).

□

Proposition B.2. Let $\{\boldsymbol{X}(\theta) : \theta = (\mu, \sigma) \in \mathbb{R}^2, \sigma > 0\}$ be a transformed location-scale family and let S be the maximum likelihood method. For a fixed parameter θ, consider a fixed set of data $X_1(\zeta_1, \theta), \ldots, X_n(\zeta_n, \theta)$ and let $\boldsymbol{Y}(\zeta, \theta)$ be the random variable defined by the bootstrapping approach, that is, $\boldsymbol{Y}(\zeta, \theta) = \boldsymbol{X}(\boldsymbol{\theta}^*)$ where $\boldsymbol{\theta}^* = (\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*)$ is obtained using either parametric or non-parametric bootstrapping.
Then

$$F_{\boldsymbol{Y}(\zeta,\theta)}\left(\boldsymbol{X}(\xi, \theta)\right)$$

does not depend on θ.

Proof. The proof is analogous to the proof of Proposition A.7 and Proposition A.8 above. In this case we consider

$$F_{\boldsymbol{Y}(\zeta,\theta)}\left(F^{-1}_{\boldsymbol{X}(\theta)}(\xi)\right) = E_{\boldsymbol{\mu}^*,\boldsymbol{\sigma}^*}\left[F_{\boldsymbol{Z}}\left(\frac{\mu-\boldsymbol{\mu}^*}{\boldsymbol{\sigma}^*}+\frac{\sigma}{\boldsymbol{\sigma}^*}F^{-1}_{\boldsymbol{Z}}(\xi)\right)\right]$$

and apply Lemma B.1. □

Corollary B.3. **??** With the assumptions of the Proposition B.2 we have

$$P\left(\boldsymbol{X}(\theta) \le SCR(\alpha;\boldsymbol{X}_1(\theta),\ldots,\boldsymbol{X}_n(\theta);M)\right)$$

does not depend on θ if $\boldsymbol{M}$ is either parametric or non-parametric bootstrapping.

Proof. Using Proposition B.2 the assertion follows from

$$\begin{aligned}
&P\left(\boldsymbol{X}(\theta) \le SCR(\alpha;\boldsymbol{X}_1(\theta),\ldots,\boldsymbol{X}_n(\theta);M)\right) = P\left(X(\boldsymbol{\xi},\theta) \le F^{-1}_{Y(\boldsymbol{\zeta},\theta)}(\alpha)\right)\\
&= P\left(F_{Y(\boldsymbol{\zeta},\theta)}(X(\boldsymbol{\xi},\theta)) \le \alpha\right)\\
&= \int_0^1 \int_{[0;1]^n} 1_{\{F_{\boldsymbol{Y}(\zeta,\theta)}(\boldsymbol{X}(\xi,\theta))\le\alpha\}} d\zeta d\xi.
\end{aligned}$$

□

References

[1] Barndorff-Nielsen OE, Cox CR (1996) Prediction and asymptotics. Bernoulli 2(4):319-340

[2] Bignozzi V, Tsanakas A (2014) Parameter Uncertainty and Residual Estimation Risk. Preprint, available at `http://ssrn.com/abstract=2158779`

[3] Borowicz J, Norman J (2006) The effects of parameter uncertainty in the extreme event frequency-severity model. International Congress of Actuaries

[4] Boucher CM, Danielsson J, Kouontchu PS, Maillet BB (2014) Risk models-at-risk. Journal of Banking & Finance 44:77-92

[5] Datta GS, Mukerjee R, Ghosh M, Sweeting TJ (2000) Bayesian prediction with approximate frequentist validity. Annals of Statistics 28(5):1414-1426

[6] Diers D (2007) Das Parameterrisiko - Ein Vorschlag zur Modellierung. Preprint, Universität Ulm

[7] Diers D, Eling M, Linde M (2013) Modeling parameter risk in premium risk in multi-year internal models. The Journal of Risk Finance 14(3):234 - 250

[8] Efron B, Tibshirani RJ (1994) An Introduction to the Bootstrap, Chapman & Hall

[9] Fisher RA (1930) Inverse Probability. Proceedings of the Cambridge Philosophical Society 26:528 - 535

[10] Fraser DA (1968) Fiducial inference. International Encyclopedia Social Sciences, The Macmillan Company and The Free Press 403 - 406, available under `http://fisher.utstat.toronto.edu/dfraser/documents/44.pdf`, retrieved on 28 July 2014

[11] Gerrard R, Tsanakas A (2011) Failure probability under parameter uncertainty. Risk Analysis 8(5):727 - 744

[12] Hogg RV, Craig AT (1995) Introduction to Mathematical Statistics. Prentice Hall, Upper Saddle River, NJ.

[13] Klugman S, Panjer H, Willmot G (2012) Loss Models - >From Data to Decisions. Wiley Series in Probability and Statistics, 4th edition

[14] Mata A (2000) Parameter Uncertainty for Extreme Value Distributions. GIRO Convention Papers

[15] McNeil AJ, Frey R, Embrechts P (2005) Quantitative Risk Management. Princeton University Press

[16] Sauler K (2009) Das Prämienrisiko in der Schadenversicherung unter Solvency II. Gesellschaft f. Finanz- u. Aktuarwiss. Ulm

[17] Sklar A (1959) Fonctions de rèpartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8:229 - 231

[18] Solvency II directive 2009/138/EC (2009) available under `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:335:0001:0155:en:PDF`, retrieved on 28 July 2014

[19] Young GA, Smith RL (2005) Essentials of statistical inference, Cambridge University Press

[20] Zabell SL (1992) R.A. Fisher and Fiducial Argument. Statistical Science 7(3):369 - 387