Failure probability under parameter uncertainty*

R. Gerrard   A. Tsanakas[†‡]

Cass Business School, City University London

Abstract: In many problems of risk analysis, failure is equivalent to the event of a random risk factor exceeding a given threshold. Failure probabilities can be controlled if a decision maker is able to set the threshold at an appropriate level. This abstract situation applies for example to environmental risks with infrastructure controls; to supply chain risks with inventory controls; and to insurance solvency risks with capital controls. However, uncertainty around the distribution of the risk factor implies that parameter error will be present and the measures taken to control failure probabilities may not be effective. We show that parameter uncertainty increases the probability (understood as expected frequency) of failures. For a large class of loss distributions, arising from increasing transformations of location-scale families (including the Log-Normal, Weibull and Pareto distributions), the paper shows that failure probabilities can be exactly calculated, as they are independent of the true (but unknown) parameters. Hence it is possible to obtain an explicit measure of the effect of parameter uncertainty on failure probability. Failure probability can be controlled in two different ways: (a) by reducing the nominal required failure probability, depending on the size of the available data set and (b) by modifying of the distribution itself that is used to calculate the risk control. Approach (a) corresponds to a frequentist/regulatory view of probability, while approach (b) is consistent with a Bayesian/personalistic view. We furthermore show that the two approaches are consistent in achieving the required failure probability. Finally, we briefly discuss the effects of data pooling and its systemic risk implications.

Keywords: *parameter uncertainty, solvency, Value-at-Risk, insurance, location-scale families.*

---

1   INTRODUCTION

1.1   Problem statement

We consider the following generic situation emerging in probabilistic risk analysis:

- A system is subject to a state corresponding to 'failure'. Failure occurs when a risk factor, modelled as a random variable, exceeds a predetermined threshold value.

- A risk management objective is to limit the probability of failure to a small value. In order to do this, it is possible to take control measures, whose effect is to vary (e.g. increase) the threshold. (This is equivalent to a shift in the risk factor distribution.) Therefore, for a fixed level of acceptable failure probability, the desired threshold can be viewed as a percentile of the risk factor distribution.

- The distribution of the risk factor, in particular the tail probabilities needed to design control measures, is unknown and must be estimated from data. This introduces parameter error and therefore one cannot be fully confident that the target failure probability is achieved.

Below we give three classes of examples, where the above setting may be seen as a reasonable model of failure and control. The third of these examples, focusing around insurance solvency, is the leading example around which arguments in this paper are constructed. The mathematical background developed is relevant for all examples.

Example 1 (Environmental risks and infrastructure controls)**.** *In a simplified model of flood risk, river or coastal flooding ('failure') ensues when when water levels exceed (*overtop*) the height of man-made flood defenses. The acceptable probability for such failure is quite low; for example in the Netherlands, exceedance probabilities between 1/4000 and 1/10000 have been specified for coastal areas* [1]*. This safety requirement, combined with the relative infrequency of great floods, implies that extrapolations to the tail probabilities of water levels are needed* [2]*.*

More broadly, investment in infrastructure and appropriate regulation can reduce the vulnerability to natural hazards. For example building codes can reduce the vulnerability to seismic risk, thus reducing the probability of

damage to a building subjected to a 'design earthquake' with given return period [3].

Example 2 (Supply chain risks and inventory controls). *In inventory problems, sufficient quantities of some good need to be stocked in order to be able to satisfy demand over a predetermined time period. Unknown demand is now the risk factor and the level of stock held is the control. As failure, we consider the scenario where demand exceeds stocked goods. Very different situations, besides classic supply chain problems, can be framed as inventory risks.*

An example in health risk management relates to the possible shortage of antiviral drugs in the case of an influenza epidemic. Health providers such as the National Health Service in Britain stock antiviral drugs [4] *and the maintenance of an insufficient stock would be considered as a precautionary failure*[5]. *Hence the probability of antiviral drugs running out in the event of an outbreak should be very low. However the relative rarity of major epidemics and the complexity of their dynamics, imply that there may be substantial uncertainty around the potential demand on antiviral drugs.*

Moving to a different context, the US government holds a Stategic Petroleum Reserve, currently (July 2010) of 727m barrels of crude oil corresponding to approximately 75 days of import protection, to be used in case of an energy supply crisis [6]. *Emergency drawdowns are rare, the last two having taken place in the aftermath of the 2005 Hurricane Katrina disaster and during the 1990/91 Desert Shield/Storm operation. Setting the size of the reserve to a level that will be able to handle an major crisis in international energy supply is a matter of substantial strategic importance for the US government.*

Example 3 (Insolvency risk and capital controls). *Financial firms such as insurance companies are exposed to random future liabilities, such as insurance claims, drops in asset values, and operational losses. In order that the firms are able to pay their liabilities under adverse scenarios, they hold* risk capital. *Risk capital is calculated according to a regulatory principle and/or the firm's own risk tolerance level. Most insurance and banking regulation adopts the Value-at-Risk principle for risk capital calculation, requiring that the probability of future losses exceeding capital is limited to a fixed low level* [7]). *For example the impending Solvency II regulatory regime for European insurers requires that the probability of insolvency (failure) for an insurance company be at most 0.5% per annum* [8]. *Hence risk capital is typically*

determined as a percentile of a loss distribution.

The calculation of risk capital is subject to substantial parameter uncertainty, since the calculation of an extreme percentile from a data set of limited size leads to potentially inaccurate capital estimates. In particular, for many types of insurance risk such as catastrophe insurance, characterised by low frequency / high severity events, relevant data are necessarily scarce and capital may have to be estimated from just a few tens of loss observations.

1.2 Results

We henceforth use the language of the third example, focusing on the problem of insurance solvency. In this contribution we address the following research questions:

1. Does parameter uncertainty increase the probability of insolvency of insurance companies and by how much?

2. Can the regulatory capital setting regime be adjusted to reflect the effect of parameter uncertainty?

3. What is the potential effect of firms' loss data sharing arrangements, both at the firm and the systemic risk level?

To answer the first question, we propose a conceptualisation of parameter uncertainty via frequentist arguments. The key idea here is to represent the calculated risk capital as a random variable, since it is a function of the random sample from which capital has been calculated. Then, the probability of insolvency is calculated as the probability of a random loss variable exceeding a random capital amount. We argue that this is essentially a regulatory view of parameter uncertainty, whereby the probability of insolvency is understood as an expected frequency of insolvency across risk portfolios. (It is noted that regulators are not the only agents having a stake in the solvency of an insurance operation. Rating agencies often perform a role similar to that of regulators, in awarding a rating that is, at least loosely, associated with a failure probability. Hence this 'regulatory view,' also applies to rating agencies. Moreover reinsurers are often closely involved in actual modelling of the risks that primary insurance companies cede to them. For a reinsurer, the probability of an insurance claim exceeding a given threshold is of interest, as this would trigger a payment to the insurer.)

The probabilities of solvency and insolvency thus calculated generally depend on the values of the true but unknown loss distribution parameters. We show that for a wide range of loss distributions, derived by increasing transforms of location-scale families and including well known distributions such as Normal, Lognormal, Exponential, Pareto and Weibull, this dependence on the true parameters vanishes. Hence, if only the family of loss distributions considered is known, it is possible to calculate the effect of parameter uncertainty on solvency probabilities. Typically, for capital calculated according to a given required confidence level, the probability of solvency will increase with the sample size of the historical data used to calculate tail probabilities, so that a larger data set implies that the holder of a risk portfolio is more secure.

It is also shown that, while the effect of parameter uncertainty on solvency probabilities is affected by the family of loss distributions considered, it is not affected in an obvious way by the distribution's tail behaviour. It is for example shown that the probability of solvency may be the same for a light-tailed (e.g. exponential) and a heavy-tailed distribution (e.g. Pareto).

Addressing the second question, we propose two methods for adjusting the risk capital calculation for financial firms, so that the problem of parameter uncertainty is addressed and the probability of insolvency constrained at an acceptable level. In the first method, capital has to be calculated using a different confidence level (estimated percentile of the loss distribution) for each risk portfolio. The confidence level used for each portfolio depends on the number of loss data points available, so that a portfolio with a long history (and hence low parameter uncertainty) receives more lenient treatment than a portfolio for which few relevant loss observations exist.

While this method is consistent with controlling the probability of insolvency, as proposed in this paper, we argue that it is ill suited for practical application, as it is at odds with principles-based regulatory practice. Instead, we propose setting capital requirements using a predictive distribution, obtained by standard Bayesian methods. So, rather than adjusting the confidence level of the capital setting regime, the loss distribution used for capital calculation is replaced by a more dispersed one. We then show that for loss distributions in transformed location-scale families, the use of a predictive distribution serves its regulatory purpose, by producing a probability of solvency, as viewed by a regulator, at the required level.

5

Finally, we discuss the case of holders of risk portfolios with similar exposures sharing their loss data, in order to reduce parameter uncertainty. We argue that, while this reduces the insolvency probability for individual portfolios, one has to make sure that systemic risk is not increased due to the dependence between portfolios' insolvency events induced by data sharing.

In the next section the relation of this research to the literature is discussed. The effect of parameter uncertainty on the solvency probability is presented in Section 2 and the proposed adjustments to the capital setting regime in Section 3. Data sharing is discussed in Section 4. Conclusions are given in Section 5. In these sections the concepts are presented in a relatively informal manner and illustrated by examples where explicit derivations are possible. All results are formally stated and proved in the Appendix.

1.3 Relation to the literature

Parameter and model uncertainty have long been fundamental issues in probabilistic risk analysis [9]. While consideration of such uncertainties in a Bayesian framework is common [10], there has been a vigorous debate among risk analysts regarding appropriate methods of quantifying such uncertainties; in particular the argument that probability is appropriate only for the modelling of process variability, rather than epistemic uncertainties has been made (see [11] as well as the responses to that article).

In the context of actuarial, insurance and financial risk management, parameter uncertainty and data issues have been discussed [12],[13],[14],[15],[16], though the implications of parameter uncertainty for solvency have been rarely studied in detail [17]. The emergence of risk-sensitive regulation in insurance markets, such as the European Solvency II regime, has produced a renewed interest in parameter uncertainty, in particular among practitioners who are called to estimate extreme percentiles from incomplete samples [18],[19],[20].

The approaches used in the literature for quantifying parameter uncertainty in insurance are generally based on deriving predictive distributions [12],[16]. Our approach is quite different, as it is based on a notion of parameter uncertainty, that draws from ideas of frequentist predictive inference [21]. By viewing the capital required to assure a given solvency probability as itself a random variable (through its dependence on the random sample), we reflect the way that uncertainty permeates insurers' decision processes,

an issue not generally addressed in the literature.

While we reserve a frequentist setting to reflect the regulatory view of uncertainty, we propose a Bayesian approach, better suited to the view of insurance market participants, for the actual method of risk capital calculation. We show that for the family of loss distributions considered in this paper these two different views of uncertainty are in practice equivalent, which follows from literature on probability matching priors and the frequentist validity of Bayesian procedures [22].

The following limitations apply to the scope of our article:

- While we propose approaches based on complementary statistical methodologies and interpret these from the point of view of different stakeholders, we do not aim at discussing the philosophical underpinnings of parameter uncertainty quantification.

- We limit ourselves to the situation where the distribution family of the modelled risks is known, with only the parameters needing to be estimated. (The precise level of prior knowledge required about the underlying loss distribution will be made clearer in Section 2.3.)

- We do not consider the possibility of structural changes in the data generating process, which would call for a different conception of return periods of extreme loss events [23].

- Asymptotic analysis of the problems we are discussing is possible [21],[24], but is not considered in this study, since our focus is on situations where samples are very small.

2 SOLVENCY UNDER UNCERTAINTY

2.1 Probability of solvency

A portfolio of risks will produce a future loss over a fixed time horizon given by random variable Y. We assume that Y follows a probability distribution $F(\cdot; \theta)$, where $\theta \in S \subset \mathbb{R}^d$ is the unknown vector of parameters. The distribution family F is known.

The parameter θ is estimated from a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ (representing for example the portfolio's loss history or a benchmark data set), where $X_i \sim F_i(\cdot; \theta)$, $i = 1, \ldots, n$. Again the families $F_i(\cdot; \theta)$ are known. We do not require that each data point comes from the same distribution

as the loss Y, but only that it is governed by the same parameter(s). We assume throughout that the random variables $X_1, \ldots, X_n, Y$ are mutually independent and that the distributions $F_1, \ldots, F_n, F$ are continuous, invertible, with densities $f_1, \ldots, f_n, f$. The unknown parameter θ is estimated from the sample $\mathbf{X}$ by an estimator $\hat{\theta} = \hat{\theta}(\mathbf{X})$, typically using maximum likelihood methods.

$X_1, \ldots, X_n$ may represent the company's observed losses from n time periods (aggregates for each period). The difference in the distributions $F_1, \ldots, F_n$ may then represent necessary adjustments, e.g. for risk exposures and business volumes changing over time. Alternatively, we could take i.i.d. $X_1, \ldots, X_n$ to stand for individual observed losses (e.g. insurance claims for a specific year arising from the insurer's risk portfolio). Then the future loss is given by a compound model $Y = \sum_{j=1}^{N} X_j'$, where X_j' has the same distribution as X_1 and N is the annual loss frequency with (known) distribution. Then Y will again have a distribution with parameters θ. We note the distinction between n, the size of the data-set used affecting parameter error, and N, the size of the portfolio driving process error.

A regulator requires that the risk capital $c(p; \theta)$ held in respect of the future loss Y be given by the relationship:

$$P_\theta[Y \leq c(p; \theta)] = p \implies c(p; \theta) = F^{-1}(p; \theta), \tag{1}$$

where $P_\theta[A]$ is the probability of an event A, for parameter value θ, and $F^{-1}(p; \theta)$ is the $100p^{th}$ percentile of the distribution $F(\cdot; \theta)$. However, given that the parameter vector θ is unknown, capital will in practice be set as a percentile of the loss distribution, using the estimated parameter values. Denote the capital thus calculated by:

$$c(p; \mathbf{X}) = F^{-1}(p; \hat{\theta}(\mathbf{X})) \tag{2}$$

The risk capital $c(p; \mathbf{X})$ is thus a function of the confidence level and the data. From the point of view of the portfolio holder, given the particular data observed, the capital $c(p; \mathbf{X})$ is a fixed number, which may be smaller or greater than the theoretically correct capital level $F^{-1}(p; \theta)$. However from the point of view of an experimental designer (or regulator), the capital $c(p; \mathbf{X})$ appears to be a random variable, due to the randomness of the vector $\mathbf{X}$. Following the latter interpretation, we can define the probability of solvency as:

$$\gamma(p; \theta) = P_\theta[Y \leq c(p; \mathbf{X})]. \tag{3}$$

This is the probability, calculated with respect to the true (but unknown) parameters θ, of the future loss Y (a random variable) being lower than the capital held (another random variable). We denote the associated insolvency probability by $\overline{\gamma}(p; \theta) = 1 - \gamma(p; \theta)$.

How are we then to interpret equation (3)? Assume that there is a large number of independent risk portfolios, for all of which capital is calculated according to the same method. Due to the randomness of the respective samples, some of those portfolios will be allocated capital that is higher and some capital that is lower than the theoretically correct value. Then $\gamma(p; \theta)$ corresponds to the expected fraction of portfolios for which the future loss will be lower than the capital held. So, from the perspective of a regulator the quantity $1 - \gamma(p; \theta)$ can be seen as an expected default frequency across portfolios, taking into account the potential parameter error in capital estimates.

Alternatively, one could view (3) in the context of a Monte-Carlo experiment. Assume that a large number m of 'histories' $\left(\mathbf{X}^{(1)}, Y^{(1)}\right), \ldots,$ $\left(\mathbf{X}^{(m)}, Y^{(m)}\right)$ are simulated, under the true parameters θ. For each history, the capital $c(p, \mathbf{X}^{(i)})$ is calculated according to (2). Then $\gamma(p; \theta)$ represents, asymptotically for large m, the fraction of histories for which the capital $c(p, \mathbf{X}^{(i)})$ is higher than the respective loss $Y^{(i)}$.

In this way, we characterise the probability of (in)solvency for a particular portfolio, allowing for parameter uncertainty. If parameter uncertainty has a detrimental effect one would have $\gamma(p; \theta) \leq p \Leftrightarrow \overline{\gamma}(p; \theta) \geq 1 - p$. Note however that the explicit calculation of $\gamma(p; \theta)$ is in reality problematic, as it presupposes knowledge of the true parameter θ, which is unknown. For example, calculation of the solvency probability via the Monte-Carlo scheme outlined above is not possible if one does not know under which distribution to simulate. Nonetheless, the following two examples show that this problem is not fatal, as dependence on the unknown parameter θ can often be eliminated.

Example 4. *Let $X_1, \ldots, X_n, Y$ be i.i.d. exponentially distributed with mean θ, i.e. $F(x; \theta) = 1 - e^{-x/\theta}$, $x > 0$. The maximum likelihood estimator of θ is $\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} X_j \sim Gam(n, \theta/n)$, so that the moment generating function of $\hat{\theta}$ is $M_{\hat{\theta}}(t) = \left(1 - \frac{\theta}{n}t\right)^{-n}$. The capital held is given by $c(p; \mathbf{X}) = -\hat{\theta} \log(1 - p)$.*

Hence the probability of insolvency $\overline{\gamma}(p;\theta)$ is calculated as:

$$
\begin{aligned}
\overline{\gamma}(p;\theta) &= E_\theta\left[P_\theta\left(Y > c(p;\mathbf{X})|\mathbf{X}\right)\right] \\
&= E\left[\exp\left(-\frac{-\hat{\theta}\log(1-p)}{\theta}\right)\right] \\
&= M_{\hat{\theta}}\left(\frac{\log(1-p)}{\theta}\right) \\
&= \left(1 - \frac{\theta}{n}\frac{\log(1-p)}{\theta}\right)^{-n}
\end{aligned}
$$

$$
\implies \overline{\gamma}(p;\theta) = \left[1 + \frac{1}{n}\log\frac{1}{1-p}\right]^{-n} \tag{4}
$$

We can see that the probability of insolvency $\overline{\gamma}(p;\theta)$ is in fact independent of the true parameter θ and is a function only of the confidence level p and the size of the sample n. In figure 1a), the insolvency probability $\overline{\gamma}(p;\theta)$ is plotted against the sample size n for confidence levels $p = 0.99, 0.995, 0.999$. The following can be observed:

- *The probability of insolvency is decreasing in the number of data points n. This makes intuitive sense: the larger the sample, the higher the accuracy of the capital assessment and thus the lower the probability of insolvency.*

- *In particular as the sample size becomes very large, the insolvency probability tends to its required value $1 - p$, as can also be seen by noting that $\lim_{n\to\infty}\left[1 + \frac{1}{n}\log\frac{1}{1-p}\right]^n = \exp\left(\log\frac{1}{1-p}\right) = \frac{1}{1-p}$, so that $\lim_{n\to\infty}\overline{\gamma}(p;\theta) = 1 - p$.*

- *For small data sets, parameter error may cause the probability of insolvency to be substantially higher than the required value; in that sense a high confidence level quoted may be rather misleading in the presence of parameter uncertainty. For example if $n = 10$ and $p = 0.99, 0.995, 0.999$ the ratios of the true to the required probability of insolvency are $\frac{\overline{\gamma}(0.99;\theta)}{1-0.99} = \frac{0.023}{0.01} = 2.26$, $\frac{\overline{\gamma}(0.995;\theta)}{1-0.995} = \frac{0.014}{0.005} = 2.85$, $\frac{\overline{\gamma}(0.999;\theta)}{1-0.999} = \frac{0.005}{0.001} = 5.24$.*

Example 5. *Suppose now that $X_1, \ldots, X_n, Y$ are independent normal $N(\mu,\sigma^2)$ variables, so that $\theta = (\mu,\sigma)$. The maximum likelihood estimators are $\hat{\mu} = \sum_{j=1}^n X_j$ and $\hat{\sigma}^2 = n^{-1}\sum(X_i - \bar{X})^2$, with $\hat{\sigma}$ independent of $\hat{\mu}$. We may write $\hat{\mu} = \mu + \sigma U/\sqrt{n}$, $\hat{\sigma}^2 = \sigma^2 V/n$, where $U \sim N(0,1)$,*
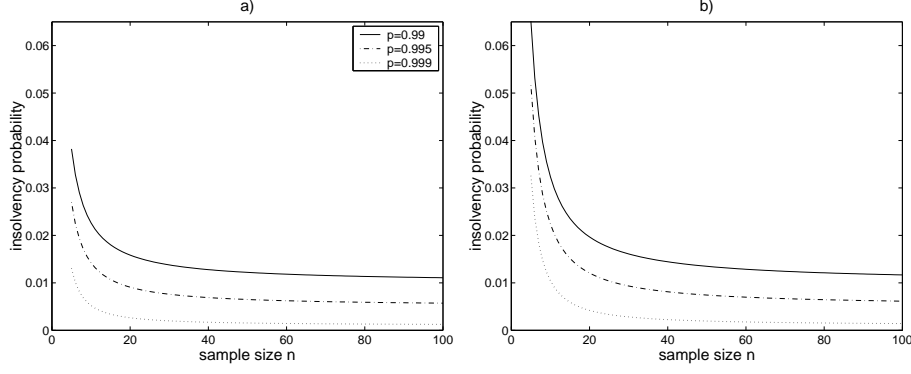
Figure 1: Probability of insolvency $\overline{\gamma}(p; \theta)$ for a) exponentially and b) normally distributed losses, against sample size n, for confidence levels $p = 0.99,\ 0.995,\ 0.999$

$V \sim \chi^2_{n-1}$, and V is independent of U. The capital held is given by $c(p, \mathbf{X}) = F^{-1}(p; (\hat{\mu}, \hat{\sigma})) = \hat{\mu} + \hat{\sigma}\Phi^{-1}(p)$, where Φ is the $N(0, 1)$ distribution function. Now

$$P_\theta[Y \leq c(p, \mathbf{X})] = P_\theta[Y - \hat{\mu} \leq \hat{\sigma}\Phi^{-1}(p)],$$

Observe that $Y - \hat{\mu} \sim N\left(0, \frac{n+1}{n}\sigma^2\right)$, independent of V, implying that

$$\sqrt{\frac{n}{n+1}} \cdot \frac{Y - \hat{\mu}}{\sqrt{\sigma^2 V/(n-1)}} \sim t_{n-1}.$$

In consequence,

$$\gamma(p; \theta) = T_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\Phi^{-1}(p)\right), \tag{5}$$

where $T_{n-1}(\cdot)$ denotes the distribution of a t_{n-1} variable. In figure 1b) the insolvency probability is again plotted as a function of n, giving a very similar pattern to that observed for the exponential distribution. For $n \to \infty$ we have $T_{n-1} \to \Phi$ so that again $\lim_{n \to \infty} \gamma(p; \theta) = p$.

2.2   Increasing transformations

Given that the confidence level p is generally chosen to be close to 1, the solvency probability $\gamma(p; \theta)$ relates to the extreme tail of the loss distribution. It may then be assumed that the overall shape and in particular the tail properties of the loss distribution $F(\cdot; \theta)$ have a substantial effect on the solvency probability. For example, is the ratio $\frac{\overline{\gamma}(p; \theta)}{1-p}$ higher for heavy- rather than light-tailed distributions?

11

Such a relationship is not straightforward. To show this consider another portfolio, for which the sample and future loss considered satisfy

$$X_1^* \overset{d}{=} h_1(X_1), \ldots, X_n^* \overset{d}{=} h_n(X_n), Y^* \overset{d}{=} h(Y),$$

where $\overset{d}{=}$ denotes equality in distribution and $h_1, \ldots, h_n, h$ are strictly increasing functions not depending on the parameter θ.

It can then be easily shown (Lemma 2 in the Appendix) that the solvency probabilities for the two portfolios are actually going to be the same, that is:

$$P_\theta[Y \le F^{-1}(p; \hat{\theta}(\mathbf{X}))] = P_\theta[Y^* \le F^{*-1}(p; \hat{\theta}(\mathbf{X}^*))] \qquad (6)$$

Note that for a non-linear function h, the probability distributions of the losses Y and Y^* may have very different shapes, e.g. Y may be light-tailed and/or symmetric, while Y^* is heavy-tailed and/or skewed. Remarkably, the effect of parameter uncertainty on the solvency probability of those two portfolios is the same. This is seen in the following example.

Example 6. *Consider i.i.d. exponential losses $X_1, \ldots, X_n, Y \sim F(x; \theta) = 1 - e^{-\frac{x}{\theta}}$, $x > 0$. Define now $X_1^*, \ldots, X_n^*, Y^* \sim F^*(x; \theta) = 1 - x^{-\frac{1}{\theta}}$, $x > 1$ via the transformation $h(t) = e^t$. Therefore, $X_1^*, \ldots, X_n^*, Y^*$ follow a one-parameter Pareto distribution. The tail-behaviour of the Pareto distribution is very different to that of the exponential distribution; in fact it is known from extreme value theory [25] that the exponential and Pareto distributions form limiting cases of the tails of light- and heavy-tailed distributions respectively. Lemma 2 implies that for both these distributions the probability of solvency $\gamma(p; \theta)$ will be the same, given by formula (4). This implies that for the Pareto distribution the solvency probability is again independent of the unknown parameter θ.*

Similarly, for normally distributed losses $X_1, \ldots, X_n, Y \sim N(\mu, \sigma^2)$, by applying the transformation $h(t) = e^t$ we get a Lognormally distributed loss profile $X_1^*, \ldots, X_n^*, Y^*$. Again the Normal and Lognormal distributions are very different; the former is symmetric, has unbounded support and is light-tailed, while the latter is skewed, has positive support and is heavy tailed (in the sense of sub-exponentiality [25]). The formula (5) will apply to both distributions, so that the solvency probability for the Lognormal distribution is again independent of the unknown parameters (μ, σ).

Hence the shape and heavy-tailedness of the loss distribution does not affect the solvency probabilities in an obvious way. What may, however, affect

these probabilities is the number of unknown parameters in the distribution. In Table I the insolvency probabilities $\overline{\gamma}(p; \theta)$ for the Exponential/Pareto, Normal/Log-normal models, as well as a one-parameter Normal model with known mean, are compared for $p = 0.99, 0.995, 0.999$. It can be seen that in all cases the 2-parameter Normal/Log-normal model produces a higher probability of insolvency than the Exponential/Pareto one. This may be attributed to the fact that the Normal/Log-normal model has an additional location parameter, which exacerbates the effect of parameter uncertainty. On the other hand, a 1-parameter Normal model, with only the variance (scale parameter) unknown, gives results very close to those of the exponential.

Table I: Comparison of insolvency probabilities $\overline{\gamma}(p; \theta)$ for the Exponential/Pareto and Normal/Log-normal models (for $p = 0.99, 0.995, 0.999$).

p = 0.99	**n = 10**	**n = 20**	**n = 50**	**n = 100**	**n = ∞**
Exponential / Pareto	0.0226	0.0159	0.0122	0.0111	0.0100
Normal / Lognormal	0.0323	0.0197	0.0135	0.0117	0.0100
1-param. Normal ($\theta = \sigma$)	0.0212	0.0153	0.0120	0.0110	0.0100
p = 0.995	**n = 10**	**n = 20**	**n = 50**	**n = 100**	**n = ∞**
Exponential / Pareto	0.0142	0.0091	0.0065	0.0057	0.0050
Normal / Lognormal	0.0224	0.0121	0.0080	0.0061	0.0050
1-param. Normal ($\theta = \sigma$)	0.0138	0.0090	0.0065	0.0057	0.0050
p = 0.999	**n = 10**	**n = 20**	**n = 50**	**n = 100**	**n = ∞**
Exponential / Pareto	0.0052	0.0026	0.0015	0.0013	0.0010
Normal / Lognormal	0.0104	0.0042	0.0020	0.0014	0.0010
1-param. Normal ($\theta = \sigma$)	0.0057	0.0029	0.0016	0.0013	0.0010

2.3 Independence of solvency probability from parameters

In the examples of Sections 2.1 and 2.2 it was seen that for particular distributions it is possible to simply calculate the probability of solvency under parameter uncertainty, since it does not depend on the unknown true parameters. In fact, this useful property holds not just for the special cases considered, but for a wide class of probability distributions, including many popular choices in risk modelling, that may be symmetric or asymmetric, with bounded or unbounded support, light- or heavy-tailed.

The families of distributions we consider are 2-parameter *transformed location-scale families* and 1-parameter *transformed scale families*. A random variable follows a distribution belonging to a transformed (location-)

scale family if an increasing function of that random variable belongs to a (location-) scale family. For example, if Y is exponentially distributed (a scale family) then the random variable $Y^* = e^Y$ is Pareto distributed (a transformed scale family). A formal definition of transformed (location-) scale families is given the Appendix. (Transformed location families can be similarly defined and results similar to the ones derived later on for scale and location-scale families hold. However these distributions are less useful for loss modelling purposes and we will not be concerned with them here.)

Examples of location-scale families are the normal distribution, the t-distribution, the logistic distribution, and the Laplace distribution. Such distributions are commonly used in modelling asset log-returns [7],[26]. By letting $h(t) = e^t$, we can derive from the distributions above, the log-normal, log-t, log-logistic (or Fisk) and log-Laplace distributions. Besides their possible use in asset return modelling these are popular models in modelling heavy-tailed insurance or operational risk losses [27],[7]). It is simple to show that other popular loss distributions such as the two-parameter Weibull [24] can also be transformed to a (non-symmetric) location-scale family.

The simpler class of 1-parameter scale families includes the exponential distribution, the Gamma and Weibull distributions (with shape parameter fixed), or indeed any location-scale family with the location parameter fixed. By transforming the exponential distribution via $h(t) = e^t$, we get the 1-parameter Pareto distribution, with cdf $1 - x^{-1/\theta}, x > 1$. It is easy to see that the usual Pareto distribution $1 - b^a(b + x)^{-a}$ also belongs to a transformed scale family if we keep either of the two parameters fixed.

Some of these distributions and their uses in insurance loss modelling are summarised in Table II.

We then have the following result, stated as Proposition 1 in the Appendix. If the probability distributions of the data $F_1(\cdot; \theta), \ldots, F_n(\cdot; \theta)$ and of the future loss $F(\cdot; \theta)$ belong to a transformed (location-) scale family, capital is calculated by an estimated loss percentile as in equation (2), and the parameter(s) θ are estimated via maximum likelihood, the corresponding probability of insolvency $\gamma(p; \theta)$ does not depend on the true parameter(s) θ. Therefore, for a wide range of useful loss models it is possible to determine explicitly the effect of parameter uncertainty on the probability of solvency.

Table II: Distributions in transformed (location-)scale families that are commonly used in insurance loss modelling. Location and scale parameters are denoted by μ, σ respectively; parameters otherwise denoted are considered as known.

Name	Density	Use
Exponential	$\frac{1}{\sigma}\exp\left(-\frac{x}{\sigma}\right),\ x>0$	Individual claims, but restricted by its simplistic shape.
Pareto (I)	$\frac{u^{1/\sigma}}{\sigma}x^{-1/\sigma-1},\ x>u$	Large individual claims, often for natural catastrophe risks and in reinsurance.
Pareto (II)	$\frac{b^{1/\sigma}}{\sigma}(b+x)^{-1/\sigma-1},\ x>0$	Large individual claims in excess of a threshold.
Gamma	$\frac{x^{a-1}\exp(-x/\sigma)}{\sigma^a\Gamma(a)},\ x>0$	Aggregate loss in an insurance portfolio, when coefficient of variation $1/\sqrt{a}$ is given.
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	Aggregate loss in well diversified portfolios, e.g. motor insurance.
Log-Normal	$\frac{\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma x},\ x>0$	Individual or portfolio loss, allowing a moderate level of extreme losses. Used widely e.g. in marine, aviation or liability insurance.
Weibull	$\frac{\alpha x^{\alpha-1}}{\beta^\alpha}\exp\left\{-(x/\beta)^\alpha\right\},\ x>0$ $\beta=e^\mu,\ \alpha=1/\sigma$	Individual or portfolio loss, with flexibility of allowing a low or moderate level of extreme losses.

2.4 Numerical calculation of solvency probability

Exact calculation of the probability of solvency can be surmised from Proposition 1 in the Appendix. However, such calculation requires integration with respect to the density of the estimator, which is not generally known in closed form. However, the solvency probability can always be calculated numerically, e.g. via a Monte-Carlo scheme such as the one described in

Section 2.1. The simulation of different loss histories can take place for *any* choice of parameters, for example by setting the location and scale parameters to 0 and 1 respectively; Proposition 1 guarantees that the choice of parameters will not affect the result. The following example illustrates the process.

Example 7. *Consider i.i.d. $X_1, \ldots, X_n \sim F(\cdot; \theta)$ where $F(\cdot; \theta)$, $\theta = (\mu, \sigma)$, is a Weibull distribution such that*

$$F(x; \theta) = 1 - \exp\left\{ -\left(e^{-\mu} x\right)^{1/\sigma} \right\}, \ x > 0.$$

The Weibull distribution is a popular choice for modelling insurance losses, as it can represent a range of tail-behaviour, depending on the choice of shape parameter σ (or $\alpha = 1/\sigma$ in a more usual parameterisation [27]) . The increasing transformation $X_i^* = \log X_i, Y^* = \log Y$ yields random variables following a negative Gumbel distribution

$$F^*(x; \theta) = 1 - \exp\left\{ e^{-\frac{x-\mu}{\sigma}} \right\}.$$

The Gumbel distribution is a location-scale family, hence Proposition 1 applies and the probability of solvency $\gamma(p; \theta)$ does not depend on the true parameters $\theta = (\mu, \sigma)$.

There is no explicit formula for Maximum Likelihood Estimator (MLE) $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ of θ, which is given as the solution to the system of equations

$$n\hat{\sigma} + \sum_{j=1}^{n} \log\left(e^{-\hat{\mu}} X_j\right) - \sum_{j=1}^{n} \left(e^{-\hat{\mu}} X_j\right)^{1/\hat{\sigma}} \log\left(e^{-\hat{\mu}} X_j\right) = 0$$
$$\left(\frac{1}{n} \sum_{j=1}^{n} X_j^{1/\hat{\sigma}}\right)^{\hat{\sigma}} - e^{\hat{\mu}} = 0 \tag{7}$$

It has been shown that in estimating parameters for the Gumbel distribution using small samples, the MLE is outperformed (in the Mean-Squared-Error sense) by estimators based on Probability Weighted Moments (PWM)[28],[29]. PWM estimators are also computationally simpler to evaluate, giving formulas

$$\tilde{\sigma} = \left(\frac{2}{n} \sum_{j=1}^{n} \frac{j-1}{n-1} X_{j:n} - \bar{X}\right) / \log 2, \quad \tilde{\mu} = \bar{X} + \gamma \tilde{\sigma}, \tag{8}$$

where $\bar{X}$ is the sample mean, $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ are the order statistics and $\gamma \simeq 0.57721$ is the Euler constant. The estimators (8), similarly to MLE, satisfy an equivariance property (in the sense of Lemma 3 in the Appendix), which means that Proposition 1 still holds and the solvency probability does not depend on the unknown parameters.

16

The required capital is then calculated as

$$c(p; \mathbf{X}) = e^{-m} \left(-\log(1 - p) \right)^s,\qquad(9)$$

by setting either $(m, s) = (\hat{\mu}, \hat{\sigma})$ or $(m, s) = (\tilde{\mu}, \tilde{\sigma})$.

Hence the following algorithm can be used to calculate $\gamma(p; \theta)$ using a Monte-Carlo sample of size r:

- Loop from $i = 1$ to $i = r$

    - Generate i.i.d. observations $(x_{i,1}, \ldots, x_{i,n}, y_i)$ from $F(\cdot; (0, 1))$.

    - Estimate μ, σ by (7) or (8).

    - Calculate capital $c(p; \mathbf{x}_i)$ by (9).

    - Record a 'success' if $c(p; \mathbf{x}_i) \geq y_i$.

- Estimate $\gamma(p; \theta)$ as the number of successes divided by r.

The algorithm was used in order to calculate the insolvency probability $\bar{\gamma}(p; \theta)$ for confidence levels $p = 0.99, 0.995, 0.999$ and sample sizes $n = 10, 20, 50, 100$. The calculation was carried out using a Monte-Carlo sample of size $r = 10^7$ and parameters were estimated with both the MLE and PWM methods. Results are summarised in Table III. Under PWM estimation, the insolvency probabilities for the Weibull distribution are very close to those of the 2-parameter (Log)normal distribution in Table I. However, for small samples, the probability of insolvency is substantially higher when MLE is used. This is explained by the worse performance of MLE compared to PWM for the Gumbel/Weibull distribution, that has been observed [28],[29]. Intuitively, less accurate parameter estimates are more likely to give rise to inappropriate capital estimates and hence exacerbate the effect of parameter uncertainty. The example indicates that weaknesses in the estimation method itself may substantially increase the probability of insolvency.

Table III: Insolvency probabilities $\overline{\gamma}(p;\theta)$ for the Weibull distribution, using Maximum Likelihood Estimation (MLE) and Probability Weighted Moments (PWM). Calculated with Monte-Carlo sample of size $r = 10^7$.

p = 0.990	n = 10	n = 20	n = 50	n = 100	n = ∞
MLE	0.0531	0.0237	0.0146	0.0124	0.0100
PWM	0.0306	0.0199	0.0139	0.0120	0.0100
p = 0.995	**n = 10**	**n = 20**	**n = 50**	**n = 100**	**n = ∞**
MLE	0.0439	0.0159	0.0084	0.0066	0.0050
PWM	0.0227	0.0132	0.0081	0.0065	0.0050
p = 0.999	**n = 10**	**n = 20**	**n = 50**	**n = 100**	**n = ∞**
MLE	0.0324	0.0075	0.0027	0.0017	0.0010
PWM	0.0129	0.0058	0.0026	0.0017	0.0010

3 ADJUSTING THE CAPITAL REQUIREMENT

It was argued in Section 2 that, when capital is held equal to an estimated percentile of the loss distribution, the solvency probability for the loss portfolio is generally lower than the specified confidence level, due to the effect of parameter uncertainty. Formally,

$$\gamma(p;\theta) = P_\theta[Y \leq F^{-1}(p;\hat{\theta})] \leq P_\theta[Y \leq F^{-1}(p;\theta)] = p.$$

Such a situation is of course unsatisfactory, since the true probability of a portfolio being solvent can be substantially lower than the one required by, for example, a regulator or the holder's own risk management.

In this section we explore two methods for adjusting the capital requirement, in order to reflect the effect of parameter uncertainty on solvency and thus increase the solvency probability to its required level. In both methods the capital requirement becomes higher. The first one relies on raising the confidence level of the solvency capital calculation, while the second one, based on Bayesian arguments, adjusts the probability distribution used for capital setting.

3.1 Raising the confidence level

One could address the problem of the solvency probability being lower than its specified level, by setting the capital at a higher confidence level, that is, require that capital is set by:

$$c(p^*; \mathbf{X}) = F^{-1}(p^*; \hat{\theta}), \quad p^* > p$$

The adjusted confidence level p^* could then be selected by requiring that

$$\gamma(p^*; \mathbf{X}) = P_\theta[Y \leq c(p^*; \mathbf{X})] = p \qquad (10)$$

If the loss distribution $F(\cdot; \theta)$ belongs to a transformed (location-) scale family, the probability of solvency does not depend on the unknown parameter θ. Hence we can explicitly solve equation (10) and specify the adjusted confidence level p^*. The process is illustrated in the following example.

Example 8. *Let as before $X_1, \ldots, X_n, Y$ distributed according to an exponential or one-parameter Pareto distribution. Then it was shown in Example 4 that the solvency probability is given by:*

$$\gamma(p; \theta) = 1 - \left[1 + \frac{1}{n} \log \frac{1}{1-p}\right]^{-n}$$

We require

$$\gamma(p^*; \theta) = p \implies p^* = 1 - \exp\left\{-n\left[(1-p)^{-1/n} - 1\right]\right\}$$

Therefore the confidence level according to which the capital has to be calculated becomes a function of the size n of the dataset. This is demonstrated in figure 2, where the value of p^* is plotted against n for different values of the required solvency level $p = 0.99,\ 0.995,\ 0.999$. It can be seen that p^* decreases as the sample size n gets larger, so that $p^* \to p$ as $n \to \infty$. Hence the holder of a portfolio with a long history of relevant data will have to calculate its capital according to a confidence level close to the required solvency probability p. In contrast, the holder of a portfolio for which very few observed data exist will have to set capital according to a confidence level that is much higher than p. For example, for $n = 10$ we have $p = 0.99 \Rightarrow p^* = 0.9971$, $p = 0.995 \Rightarrow p^* = 0.9991$, $p = 0.999 \Rightarrow p^* = 0.99995$.

Following the approach suggested here, each loss portfolio will have to be assigned a different confidence level p^*, on which the capital requirement is based, where p^* depends on the loss distribution and the sample size for the portfolio. On the face of it this makes sense: long experience and diligent data collection are rewarded with a lower capital requirement; lack of experience and incomplete data are penalised. It is noted that this argument refers to *relevant* experience. It may be that a company has a long history but changes in the risk profile, e.g. due to changes in portfolio mix or perils driven by climate change, render much of that data set irrelevant. We do
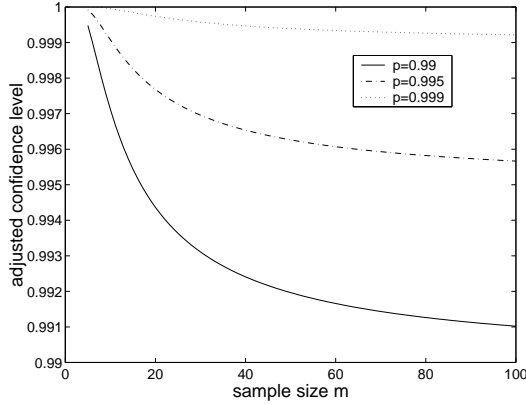
Figure 2: Adjusted confidence level p^* used for capital setting in the exponential/Pareto model, as a function of sample size n, for required solvency probabilities $p = 0.99,\ 0.995,\ 0.999$

not deal with the case of detecting distributional changes over time, and assume throughout that the data used are i.i.d.; discarding irrelevant data of course reduces the sample size and *increases* the capital requirement.

There are other problems with adopting such an approach in practice too. Consider that the capital requirement is being set by a regulator. This means that the regulator would have to set a different confidence level p^* for each regulated risk portfolio. Therefore the regulator would have to be equipped with the knowledge of each portfolio's aggregate loss distribution (assuming this belongs to a transformed location-scale family), so that p^* can be calculated. Gathering such information may be logistically challenging and assumes that regulators would be much closer to insurance company's modelling than they really are. Moreover, such an approach may appear excessively prescriptive and thus at odds with principles-based regulatory regimes such as ICAS and Solvency II.

Furthermore, such capital setting would likely be challenged by the portfolio holders. The adoption of different confidence levels across portfolios would give the impression of inequity and may be politically difficult to support. Implementation of the regime would not be helped by the potential difficulty of explaining it to portfolio holders. The adjustment to the confidence level relies on the rather abstract notion of a random capital level, as in equations (2), (3). As discussed in Section 2.1, this may be best explained as an expected frequency of solvency across portfolios. While such a quantity is certainly meaningful from a regulatory perspective, it is not

necessarily relevant to individual portfolio holders such as insurance firms. After all, the idea of a random capital level may strike portfolio holders, who have already observed past losses and set their capital at a fixed amount, as rather odd. Of course, no regulator would communicate in such abstract terms, but instead state that a higher requirement is due whenever there is insufficient information about the loss distribution. While this may help communication, it would not resolve the underlying epistemic problem.

The next section proposes an adjusted capital requirement that addresses these issues.

3.2   Adjusting the loss distribution

If holders of risk portfolios find the frequentist interpretation of parameter uncertainty used to justify the adjustment in Section 3.1 meaningless, a different conceptualisation of uncertainty is necessary. This can be achieved using standard Bayesian arguments. Unknown parameters can be considered as random quantities and the additional risk to solvency induced by parameter uncertainty can be reflected by constructing a more dispersed predictive distribution.

A prior density $\pi(\theta)$ is defined over the parameter space $\theta \in \mathcal{S}$. The prior π is potentially improper and can be chosen to be uninformative, so that subjective judgement does not enter the calculation. Given the observed data $\mathbf{X} = \mathbf{x}$, the parameters' posterior density is given by:

$$\pi(\theta|\mathbf{x}) = I(\mathbf{x})^{-1}\pi(\theta)\prod_{j=1}^{n} f_j(x_j;\theta), \tag{11}$$

where $I(\mathbf{x}) = \int_{\eta \in \mathcal{S}} \pi(\eta)\prod_{j=1}^{n} f_j(x_j;\eta)d\eta$. The predictive density and distribution of Y are respectively defined as:

$$\hat{f}(y|\mathbf{x}) = \int_{\eta \in \mathcal{S}} f(y;\eta)\pi(\eta|\mathbf{x})d\eta, \quad \hat{F}(y|\mathbf{x}) = \int_{\eta \in \mathcal{S}} F(y;\eta)\pi(\eta|\mathbf{x})d\eta. \tag{12}$$

Probabilities calculated using the predictive density are denoted by $\hat{P}(\cdot|\mathbf{x})$.

An alternative way of raising the capital requirement is then, rather than increase the confidence level, to adjust the loss distribution used for capital setting by using the predictive distribution. This reflects parameter uncertainty by being typically more dispersed than the loss distribution $F(\cdot;\hat{\theta}(\mathbf{x}))$. Thus capital can be set as a percentile of the predictive distribution:

$$\hat{c}(p|\mathbf{x}) = \hat{F}^{-1}(p|\mathbf{x}). \tag{13}$$

Such an approach presents substantial practical advantages compared to the adjustment of the confidence level discussed in Section 3.1. A regulatory regime whereby capital is set as a percentile of the predictive distribution is consistent with principle-based regimes; the regulator has only to specify a confidence level, the same for all companies, and all calculations are carried out by companies themselves. The regulator does not need to have access to loss data or to know the family of loss distributions best describing the portfolio's loss profile. In effect, the change from current regulatory practices would be marginal; rather than the regulator saying 'calculate your capital requirement as the 99.5th percentile of your *estimated* loss distribution,'' he would say 'calculate your capital requirement as the 99.5th percentile of your *predictive* loss distribution.'' Furthermore, from the perspective of portfolio holders, reflecting parameter uncertainty by defining a distribution over the parameters may be easier to interpret than viewing the capital held as a random number.

It is, however, not clear whether setting capital by the predictive distribution has the desired effect, that is, whether it raises the probability of solvency, with respect to the true parameters, to an acceptable level. That is, we would like to know the value of the probability:

$$\delta(p; \theta) = P_\theta[Y \leq \hat{c}(p|\mathbf{X})], \tag{14}$$

where the capital level $\hat{c}(p|\mathbf{X})$ is now viewed as a random variable. Essentially (14) states the expected probability of solvency from the perspective of a regulator, when the capital is set according to (13). Ideally it would be $\delta(p; \theta) = p$, as it would mean that the adjustment to the capital setting regime raised the probability of solvency to its required level. An example where such consistency holds is given next.

Example 9. *Let i.i.d. $X_1, \ldots, X_n, Y$ follow $F(x; \theta) = 1 - e^{-\frac{x}{\theta}}, \ x \geq 0$. Standard arguments show that for prior $\pi(\theta) = \theta^{-\nu}, \ \nu \geq 0$ and observed sample $\mathbf{X} = \mathbf{x}$, the predictive distribution obtained is*

$$\hat{f}(y|\mathbf{x}) = (n + \nu - 1) \frac{(\sum_j x_j)^{n+\nu-1}}{(y + \sum_j x_j)^{n+\nu}} \Rightarrow \hat{P}[Y > y|\mathbf{x}] = \left(\frac{\sum_j x_j}{y + \sum_j x_j} \right)^{n+\nu-1},$$

which is a Pareto distribution; hence the allowance for parameter uncertainty has turned the distribution used for capital assessment from a light- to a heavy tailed one. Capital can then be set as:

$$\hat{c}(p|\mathbf{x}) = \hat{F}^{-1}(p|\mathbf{x}) = \sum_j x_j \left[(1 - p)^{-1/(n+\nu-1)} - 1 \right].$$

The expected frequency of insolvencies under this capital setting method is given by

$$
\begin{aligned}
\overline{\delta}(p;\theta) &= P_\theta\left[Y > \hat{c}(p|\mathbf{X})\right] \\
&= P_\theta\left[Y > \sum_j X_j\left((1-p)^{-1/(n+\nu-1)} - 1\right)\right] \\
&= P_\theta\left[\frac{\sum_j X_j}{\sum_j X_j + Y} < (1-p)^{1/(n+\nu-1)}\right].
\end{aligned}
$$

The quantity $\frac{\sum_j X_j}{\sum_j X_j + Y}$ follows a $Beta(n,1)$ distribution with cumulative distribution $B(x;n,1) = x^n, \ x \in [0,1]$. Therefore

$$
\overline{\delta}(p;\theta) = (1-p)^{n/(n+\nu-1)} \tag{15}
$$

Note that if we let $\nu = 1$, that is, if we use an uninformative prior of the form $\pi(\theta) = \theta^{-1}$ appropriate for scale families [30], we get

$$
\overline{\delta}(p;\theta) = 1 - p \Leftrightarrow \delta(p;\theta) = p,
$$

as required.

The example demonstrated the potential effectiveness of using a predictive distribution for capital setting, in satisfying the regulatory solvency probability requirement. Such consistency between Bayesian and frequentist methods [24],[22] is more generally true for the transformed (location-) scale families discussed in this paper. In particular, as shown in Proposition 2 in the Appendix, for such distributions, when the prior $\pi(\mu,\sigma) = \sigma^{-1}$ is used, it is always the case that $\delta(p;\theta) = p$. More broadly, for priors of the form $\pi(\mu,\sigma) = \sigma^{-\nu}$, we find that the resulting solvency probability again does not depend on the real parameters, so that the probability can be potentially calculated by a regulator.

4   DATA POOLING

In Section 3 it was argued that the effect of parameter uncertainty on solvency can be addressed by a stricter regulatory regime, adopting a higher confidence level or more dispersed loss distribution. These are actions that a regulator may take. Portfolio holders on the other hand can increase their probability of solvency by addressing the key issue discussed in this paper: lack of sufficient data.

We assume that the losses from the portfolios of a number of different companies, follow distributions with the same parameters, after appropriate adjustments. (This simplifying assumption is of course quite strong. We are not concerned here with credibility theory, the optimal combined use of a company's own and benchmark data for pricing purposes [27], which has also been proposed in the context of operational risk capital setting [31]). Then the portfolio holders may decide to pool their data in order that each of them is able to estimate their required capital from a larger data set. As a consequence the solvency probability of each should increase.

We note that the setting here is somewhat contrived, as it is required that companies have independent risk exposures following similar probability distributions, which are estimated using the same benchmark data set. This is a situation not easily envisaged in insurance risk management. However in operational risk management, pooling of loss data between different financial institutions does occur [32], while it is not unreasonable to assume that operational risk losses are independent across institutions.

A side-effect of pooling data is that, even if the losses from different portfolios are independent, the portfolio (in)solvency events become dependent. This is because the capital of each portfolio is calculated on the basis of the same random sample. Hence there is a chance that the capital is collectively over- or under-stated for all portfolios. One has to then examine the effect of data sharing on systemic stability, for example via the joint probability of insolvency. Such an examination is given below, for the simple example of two portfolios with exponentially distributed losses. The process can be carried out for losses in any (location-) scale family, as the joint insolvency probability will again be independent of the true parameter θ.

Example 10. *Consider the simple case where only two portfolios exist, and all losses, within and across portfolios, are i.i.d. Hence we have*

$$\left.\begin{array}{l} \mathbf{X}_1 = (X_{1,1}, \ldots, X_{1,n_1}), Y_1 \\ \mathbf{X}_2 = (X_{2,1}, \ldots, X_{2,n_2}), Y_2 \end{array}\right\} \sim F(\cdot; \theta),$$

where F is an exponential or Pareto distribution. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ and $n = n_1 + n_2$. First consider the case that data are not pooled. Then the probability of insolvency for each portfolio is

$$P_\theta[Y_i > c(p_i; \mathbf{X}_i)] = \left[1 + \frac{1}{n_i} \log \frac{1}{1 - p_i}\right]^{-n_i}, \; i = 1, 2$$

and the joint probability of insolvency is given by:

$$P_\theta[Y_1 > c(p_1; \mathbf{X}_1), Y_2 > c(p_2; \mathbf{X}_2)] = \left[1 + \frac{1}{n_1} \log \frac{1}{1 - p_1}\right]^{-n_1} \left[1 + \frac{1}{n_2} \log \frac{1}{1 - p_2}\right]^{-n_2}.$$
(16)

Let now the two portfolios merge their data sets. Then the probability of insolvency for each portfolio is

$$P_\theta[Y_i > c(p; \mathbf{X})] = \left[1 + \frac{1}{n} \log \frac{1}{1 - p_i}\right]^{-n}, \ i = 1, 2,$$

which is lower than $P_\theta[Y_i > c(p_i; \mathbf{X}_i)]$, demonstrating the beneficial effect of data sharing. Calculation yields the joint probability of insolvency as

$$P_\theta[Y_1 > c(p_1; \mathbf{X}), Y_2 > c(p_2; \mathbf{X})] = \left[1 + \frac{1}{n} \log \left(\frac{1}{(1 - p_1)(1 - p_2)}\right)\right]^{-n}.$$
(17)

So which of the two situations (pooled or un-pooled data) produces a lower joint probability of insolvency? Observe that

$$
\begin{aligned}
P_\theta[Y_1 > c(p_1; \mathbf{X}), Y_2 > c(p_2; \mathbf{X})]^{-1/n} &= 1 + \frac{1}{n} \log \left(\frac{1}{(1 - p_1)(1 - p_2)}\right) \\
&= \frac{1}{n} \left[n_1 \left(1 + \frac{1}{n_1} \log \frac{1}{1 - p_1}\right) + n_2 \left(1 + \frac{1}{n_2} \log \frac{1}{1 - p_2}\right)\right] \\
&\geq \left[\left(1 + \frac{1}{n_1} \log \frac{1}{1 - p_1}\right)^{n_1} \left(1 + \frac{1}{n_2} \log \frac{1}{1 - p_2}\right)^{n_2}\right]^{1/n} \\
&= P_\theta[Y_1 > c(p_1; \mathbf{X}_1), Y_2 > c(p_2; \mathbf{X}_2)]^{-1/n},
\end{aligned}
$$

(where the inequality follows from the fact that the arithmetic mean is greater than the geometric one) implying

$$P_\theta[Y_1 > c(p_1; \mathbf{X}), Y_2 > c(p_2; \mathbf{X})] \leq P_\theta[Y_1 > c(p_1; \mathbf{X}_1), Y_2 > c(p_2; \mathbf{X}_2)]$$

Note that an equality is obtained only for $n_1 \log(1 - p_1) = n_2 \log(1 - p_2)$; in particular when $p_1 = p_2$ we have equality for $n_1 = n_2$. In this simple example, pooling of data is beneficial at the individual portfolio, as well as the systemic level.

For a numerical example we fix $p_1 = p_2 = 0.995$ and let $n_1 = 10, \ 20, \ 50$. In figure 3a) we plot against n_2 the correlation of the indicator functions of insolvency events $\{Y_1 > c(p_1; \mathbf{X})\}$, $\{Y_2 > c(p_2; \mathbf{X})\}$. This 'insolvency event correlation' is similar to the default correlation encountered in credit risk analysis [7]. It can be seen that the correlation is positive, indicating the positive dependence that the pooling of data induces. The correlation
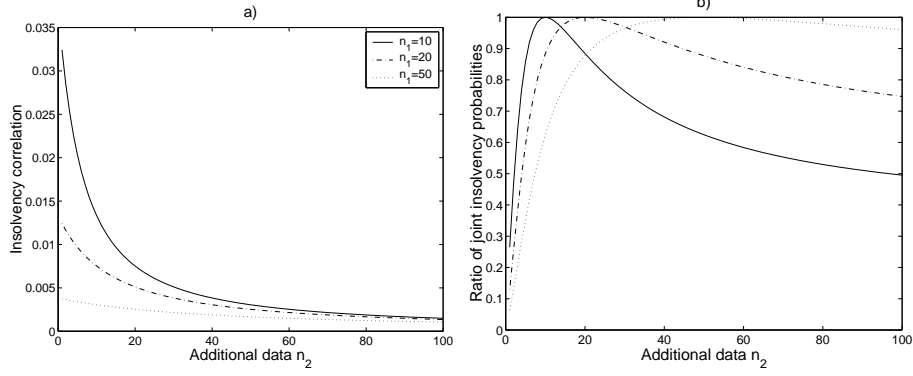
Figure 3: The effect of pooling on joint insolvency, for exponential/Pareto losses, $n_1 = 10, \; ,20, \; 50$ and $p_1 = p_2 = 0.995$: a) insolvency event correlation; b) ratio of the joint insolvency probabilities $P_\theta[Y_1 > c(p_1; \mathbf{X}), Y_2 > c(p_2; \mathbf{X})]/P_\theta[Y_1 > c(p_1; \mathbf{X}_1), Y_2 > c(p_2; \mathbf{X}_2)]$.

is reduced when the number of available data n_1, n_2 increases. In figure 3b), the ratio of the joint insolvency probabilities $P_\theta[Y_1 > c(p_1; \mathbf{X}), Y_2 > c(p_2; \mathbf{X})]/P_\theta[Y_1 > c(p_1; \mathbf{X}_1), Y_2 > c(p_2; \mathbf{X}_2)]$ is plotted. It is seen that the ratio is always < 1, such that data pooling, while introducing insolvency correlation, does not increase systemic risk. The lowest values of the ratio are obtained when n_1 and n_2 are very different. This implies that the greatest benefit of pooling data is obtained when, of the two data-sets pooled, the one is small and the other large.

5   CONCLUSION AND DISCUSSION

In the article it is argued that a frequentist interpretation of parameter uncertainty is appropriate for quantifying its effect on the solvency probability of loss portfolios, when seen through the eyes of a regulator. It is shown that for a rich class of loss distributions, solvency probabilities can be explicitly calculated (even though the real parameters remain unknown) and depend strongly on the size of the data set used for the estimation of risk capital. In particular we show that for small datasets the true probability of insolvency for a portfolio can be much higher than the notional one. When taking the viewpoint of a financial firm, a Bayesian interpretation of uncertainty is more appropriate. Hence we propose an improvement on current capital setting regimes, based on percentiles of predictive distributions, and use a

result from the theory of predictive inference [22] in order to reconcile the two viewpoints.

We also argue that data sharing arrangements should be examined from the point of view of systemic risk, due to the dependence of insolvency events induced by portfolio holders using the same dataset.

The key results for our article are Propositions 1 and 2. These only hold for loss distributions that can be transformed to (location-) scale ones. The assumption is of course rather restrictive; it is however the price we have to pay for deriving explicit formulas for small samples. More general results could be derived by asymptotic arguments[21],[24] that we do not pursue here; such an approach may be of interest, but the case of large samples is possibly less relevant for the study of parameter uncertainty.

Throughout the article we have assumed that the data are independent of each other and of future losses. This assumption is not necessarily realistic; dependency of loss variables, in particular at the tails of distributions, is a recurring theme in financial risk management [7]. Propositions 1 and 2 can easily be extended for data following an arbitrary joint distribution with transformed (location-) scale marginals, as long as the dependence structure is known. The case when the dependence structure itself is unknown and must be estimated from the data is harder and is the subject of future work.

An issue that we have not addressed at all is model uncertainty, since we assumed that the family of distributions that each random variable follows is known. Insofar, our results for insolvency probability may form a best-case scenario, as the inclusion of model uncertainty may paint an even bleaker picture of financial firms' solvency probabilities. Extension of our reasoning to a framework addressing model uncertainty is possible. Bayesian reasoning in the context of model unertainty is well-established [10],[12]. A frequentist view of model uncertainty is also possible, if one formalises model selection according to a prescribed criterion as a data-driven random process. Such an analysis can form the basis of a future investigation.

APPENDIX: FORMAL STATEMENTS AND PROOFS

For the statements of results and derivations in the appendix, all distributions considered are continuous, invertible and have a corresponding density. All random variables are independent. Many of the results stated below are rather simple and can be found in mathematical statistics textbooks; they are restated here for reasons of clarity and completeness.

A   Increasing transformations

For a sample $\mathbf{X} = (X_1, \ldots, X_n)$, such that $X_i \sim F_i(\cdot; \theta)$, $i = 1, \ldots, n$, consider an estimator of the parameter θ, given by $\hat{\theta}(\mathbf{X}; F_1, \cdots, F_n)$. Given functions $h_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, use the notation $h \circ \mathbf{X} = (h_1(X_1), \ldots, h_n(X_n))$. Consider then a transformed sample $\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ such that $X_i^* \sim F_i^*(\cdot; \theta)$, $i = 1, \ldots, n$, so that the estimator of θ from the transformed sample is $\hat{\theta}(\mathbf{X}^*; F_1^*, \cdots, F_n^*)$.

Consider now that a Bayesian approach. A prior $\pi(\cdot)$ is defined over θ, such that the posterior is given by $\pi(\theta | \mathbf{X}) \propto \pi(\theta) \prod_{j}^{n=1} f_j(X_j; \theta)$. Assuming that the same prior π is used, denote by $\pi^*(\theta | \mathbf{X}^*) \propto \pi(\theta) \prod_{j}^{n=1} f_j^*(X_j^*; \theta)$ the posterior of θ from the transformed sample.

Lemma 1. *Consider* $\mathbf{X}$ *and* $\mathbf{X}^*$ *as above with* $h_1, \ldots, h_n$ *be strictly increasing functions. Then:*

   i) The maximum likelihood estimator $\hat{\theta}$ *is invariant to the transformations* $h_1, \ldots, h_n$, *in the sense that*

$$\hat{\theta}(\mathbf{X}; F_1, \cdots, F_n) = \hat{\theta}(\mathbf{X}^*; F_1^*, \cdots, F_n^*).$$

   ii) The posterior of θ *is invariant to* $h_1, \ldots, h_n$, *in the sense that*

$$\pi(\theta; \mathbf{X}) = \pi^*(\theta; \mathbf{X}^*).$$

Proof. Let $v_i(s) = h_i(s)$. The likelihood of the transformed sample $\mathbf{X}$ is

$$\prod_{j=1}^{n} f_j^*(X_j^*; \theta) = \prod_{j=1}^{n} v'(X_j^*) f_j(v(X_j^*); \theta) = \prod_{j=1}^{n} v'(h(X_j)) \prod_{j=1}^{n} f_j(X_j; \theta),$$

Hence the likelihood of the transformed sample $\mathbf{X}^*$ is the same as that of the sample $\mathbf{X}$, multiplied by a term that does not contain θ. From this both i) and ii) immediately follow. $\qquad\square$

Lemma 2. *Let $X_i \sim F_i(\cdot; \theta), i = 1, \ldots, n$, $Y \sim F(\cdot; \theta)$ and transformed variables $X_i^* = h_i(X_i) \sim F_i^*(\cdot; \theta), i = 1, \ldots, n$, $Y^* = h(Y) \sim F^*(\cdot; \theta)$ where the functions $h_1, \ldots, n, h$ are strictly increasing.*

i) Let $\hat{\theta}(\mathbf{X}; F_1, \ldots, F_n)$ be the maximum likelihood estimator. Then

$$P_\theta[Y \leq F^{-1}(p; \hat{\theta}(\mathbf{X}; F_1, \ldots, F_n))] = P_\theta[Y^* \leq F^{*-1}(p; \hat{\theta}(\mathbf{X}^*; F_1^*, \ldots, F_n^*))]$$

ii) Denote the predictive distributions derived from the original and transformed samples by $\hat{F}(y|\mathbf{X}) = \int F(y; \eta) \pi(\eta|\mathbf{X}) d\eta$,
$\hat{F}^*(y|\mathbf{X}^*) = \int F^*(y; \eta) \pi^*(\eta|\mathbf{X}^*) d\eta$ respectively. Then

$$P_\theta[Y \leq \hat{F}^{-1}(p|\mathbf{X})] = P_\theta[Y^* \leq \hat{F}^{*-1}(p|\mathbf{X}^*)].$$

Proof. Part i): By Lemma 1i) and the strict increasingness of h,

$$P_\theta[Y^* \leq F^{*-1}(p; \hat{\theta}(\mathbf{X}^*; F_1^*, \ldots, F_n^*))] = P_\theta[h(Y) \leq h \circ F^{-1}(p; \hat{\theta}(\mathbf{X}; F_1, \ldots, F_n))],$$

which yields the required result.

Part ii): From Lemma 1ii) it follows that $\hat{F}^*(y|\mathbf{X}^*) = \hat{F}(h^{-1}(y)|\mathbf{X}) \Rightarrow$ $\hat{F}^{*-1}(p|\mathbf{X}^*) = h \circ \hat{F}^{-1}(p|\mathbf{X})$. Therefore $P_\theta[Y^* \leq \hat{F}^{*-1}(p|\mathbf{X}^*)] = P_\theta[h(Y) \leq h \circ \hat{F}^{-1}(p|\mathbf{X})]$, which yields the required result. $\square$

B Transformed location-scale families

All definitions and results are given for 2-parameter location-scale families, since results and proofs for location or scale families are very similar.

A set $\mathcal{A}_{LS}$ of univariate probability distributions is called a *location-scale family* if for every $X_1 \sim F_1 \in \mathcal{A}_{LS}, X_2 \sim F_2 \in \mathcal{A}_{LS}$ we can write $X_2 \stackrel{d}{=} aX_1 + b$ for some $a > 0, b \in \mathbb{R}$. We can choose an particular element $F \in \mathcal{A}_{LS}$ and call this the *standard distribution* of the family $\mathcal{A}_{LS}$. Let $Z \sim F$. Then for every X with a distribution in $\mathcal{A}_{LS}$, we can write $X \stackrel{d}{=} \mu + \sigma Z, \sigma > 0$ and use the notation $X \sim F(\cdot; \mu, \sigma)$, where $F(x; \mu; \sigma) = F\left(\frac{x-\mu}{\sigma}\right)$. It obviously is $Z \sim F(\cdot; 0, 1)$.

Let $\mathcal{A}_{LS}$ be a location-scale family. A set $\mathcal{A}_{LS}^h$ of univariate probability distributions is called a *transformed location-scale family,* if there is strictly increasing function h such that for any $X \in \mathcal{A}_{LS}^h$ it is $h^{-1}(X) \in \mathcal{A}_{LS}$. Then we can write $X \sim F(\cdot; \mu, \sigma; h)$, where $F(x; \mu, \sigma; h) = F\left(\frac{h^{-1}(x)-\mu}{\sigma}\right)$, with F the standard distribution of $\mathcal{A}_{LS}$.

Lemma 3. *Consider $X_i \sim F_i(\cdot; \mu, \sigma)$, $i = i, \ldots, n$, belonging to location-scale families. Then the maximum likelihood estimator $\hat{\theta}(\mathbf{X}; F_1, \ldots, F_n) = (\hat{\mu}(\mathbf{X}; F_1, \ldots, F_n), \hat{\sigma}(\mathbf{X}; F_1, \ldots, F_n))$ is location-scale equivariant in the sense that*

$$\hat{\theta}(a\mathbf{X} + b; F_1, \ldots, F_n) = (a\hat{\mu}(\mathbf{X}; F_1, \ldots, F_n) + b, a\hat{\sigma}(\mathbf{X}; F_1, \ldots, F_n)),$$

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

Proof. Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ maximise the likelihood

$$\prod_{j=1}^{n} f_j(X_j; \mu, \sigma) = \prod_{j=1}^{n} \frac{1}{\sigma} f_j\left(\frac{X_j - \mu}{\sigma}\right),$$

where $f_i(\cdot)$ is the standardised density of X_i. For $a > 0$ consider transformed sample $X_i^* = aX_i + b$, $i = 1, \ldots, n$. Its likelihood is:

$$\prod_{j=1}^{n} f_j(X_j^*; \mu, \sigma) = \prod_{j=1}^{n} \frac{1}{\sigma} f_j\left(\frac{aX_j + b - \mu}{\sigma}\right)$$

$$= a^{-n} \prod_{j=1}^{n} \frac{1}{\sigma/a} f_j\left(\frac{X_j - (\mu - b)/a}{\sigma/a}\right)$$

So if $\hat{\theta}^* = (\hat{\mu}^*, \hat{\sigma}^*)$ maximises the likelihood of the transformed sample, it follows that $(\hat{\mu}^* - b)/a = \hat{\mu}$, $\hat{\sigma}^*/a = \hat{\sigma}$ from which the required result follows. $\square$

Consequently, the maximum likelihood estimator $(\hat{\mu}, \hat{\sigma})$ has joint distribution $G(\cdot, \cdot; \mu, \theta)$ and (assuming it is well defined) density $g(\cdot, \cdot; \mu, \theta)$, that can be written as

$$G(m, s; \mu, \sigma) = G\left(\frac{m - \mu}{\sigma}, \frac{s}{\sigma}\right), \qquad g(m, s; \mu, \sigma) = \frac{1}{\sigma^2} g\left(\frac{m - \mu}{\sigma}, \frac{s}{\sigma}\right), \tag{18}$$

where $G(m, s) \equiv G(m, s; 0, 1)$ and $g(m, s) = \frac{\partial^2}{\partial m \partial s} G(m, s)$ do not depend on μ, σ.

Lemma 4. *Consider $X_i \sim F_i(\cdot; \mu, \sigma)$, $i = i, \ldots, n$, belonging to location-scale families. Then, assuming a prior of the form $\pi(\mu, \sigma) = \sigma^{-\nu}$, $\nu \geq 0$, the posterior density is location-scale equivariant in the sense that*

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

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

Proof. Let $X_i^* = aX_i + b$, $i = 1, \ldots, n$, and denote by $\pi^*(m, s; \mathbf{X}^*)$ the posterior density of the parameters under transformed sample, so that:

$$\pi^*(m, s | \mathbf{X}^*) = I(\mathbf{X}^*)^{-1} s^{-\nu} \prod_{j=1}^n f_j(X_j^*; m, s),$$

where $I(\mathbf{X}^*)$ is the normalising constant. It is:

$$s^{-\nu} \prod_{j=1}^n f_j(X_j^*; m, s) = s^{-\nu} \prod_{j=1}^n \frac{1}{s} f_j\left(\frac{aX_j + b - m}{s}\right)$$

$$= a^{-n-\nu} \left(\frac{s}{a}\right)^{-\nu} \prod_{j=1}^n \frac{1}{s/a} f_j\left(X_j; \frac{m-b}{a}, \frac{s}{a}\right)$$

On the other hand:

$$I(\mathbf{X}^*) = \int_{m \in \mathbb{R}} \int_{s \in \mathbb{R}_{++}} s^{-\nu} a^{-n} \prod_{j=1}^n \frac{1}{s/a} f_j\left(X_j; \frac{m-b}{a}, \frac{s}{a}\right) ds\, dm$$

Making the substitutions $\zeta = \frac{m-b}{a}$, $\eta = \frac{s}{a}$, we get

$$I(\mathbf{X}^*) = \int_{\zeta \in \mathbb{R}} \int_{\eta \in \mathbb{R}_{++}} \eta^{-\nu} a^{-n-\nu} \prod_{j=1}^n \frac{1}{\eta} f_j(X_j; \zeta, \eta) \, a^2 d\eta\, d\zeta = a^{2-\nu-n} I(\mathbf{X}),$$

from which the result follows. $\qquad\square$

C   Main results

Proposition 1. *Consider $X_i \sim F_i(\cdot; \mu, \sigma; h_i)$, $i = i, \ldots, n$, $Y \sim F(\cdot; \mu, \sigma; h)$, belonging to transformed location-scale families. Let $(\hat{\mu}(\mathbf{X}), \hat{\sigma}(\mathbf{X}))$ be the maximum likelihood estimator. Then the probability $\gamma(p; \mu, \sigma) = P_{\mu,\sigma}[Y \leq F^{-1}(p; \hat{\mu}, \hat{\sigma}; h)]$ is given by*

$$\gamma(p; \mu, \sigma) = \int_{\zeta \in \mathbb{R}} \int_{\eta \in \mathbb{R}_{++}} F\left(\zeta + \eta F^{-1}(p)\right) g(\zeta, \eta) \, d\eta\, d\zeta, \qquad (19)$$

where F is the distribution of $\frac{h^{-1}(Y) - \mu}{\sigma}$, and $g(\cdot, \cdot)$ is a bivariate density function as in eq. (18). Therefore $\gamma(p; \mu, \sigma)$ does not depend on the parameters (μ, σ).

Proof. We prove the result only for the case of location-scale families. The extension to transformed location scale families is achieved by Lemma 2i). Applying Lemma 3, we have

$$\gamma(p; \mu, \sigma) = E_{\mu,\sigma} P_{\mu,\sigma}[Y \leq F^{-1}(p; \hat{\mu}, \hat{\sigma}; h) | (\hat{\mu}, \hat{\sigma})]$$

$$= \int_{m\in\mathbb{R}} \int_{s\in\mathbb{R}_{++}} F\left(F^{-1}(p;m,s);\mu,\sigma\right) g(m,s;\mu,\sigma)dsdm$$

$$= \int_{m\in\mathbb{R}} \int_{s\in\mathbb{R}_{++}} F\left(\frac{m-\mu}{\sigma} + \frac{s}{\sigma}F^{-1}(p)\right) \frac{1}{\sigma^2}g\left(\frac{m-\mu}{\sigma}, \frac{s}{\sigma}\right) dsdm$$

Carrying out the change of variables $\frac{m-\mu}{\sigma} = \zeta, \frac{s}{\sigma} = \eta$ yields the stated result. $\qquad\square$

Proposition 2. *Consider* $X_i \sim F_i\left(\cdot;\mu,\sigma;h_i\right), \ i = i,\ldots,n, \ Y \sim F\left(\cdot;\mu,\sigma;h\right),$ *belonging to transformed location-scale families. Let* $\hat{F}(y;h|\mathbf{X})$ *be the predictive distribution of* Y *given the data* $\mathbf{X}$, *obtained using prior* $\pi(\mu,\sigma) = \sigma^{-\nu}, \ \nu \geq 0$. *Then the probability* $\delta(p;\mu,\sigma) = P_{\mu,\sigma}[Y \leq \hat{F}^{-1}(p;h|\mathbf{X})]$ *is given by*

$$\delta(p;\mu,\sigma) = E_{\mu,\sigma}\left[F\left(\hat{F}^{-1}(p;h|\mathbf{Z})\right)\right], \tag{20}$$

where F *is the distribution of* $\frac{h^{-1}(Y)-\mu}{\sigma}$ *and* $Z_i = \frac{h_i^{-1}(X_i)-\mu}{\sigma}, \ i = 1,\ldots,n,$ *so that* $\delta(p;\mu,\sigma)$ *does not depend on the parameters* (μ,σ). *In particular, for* $\nu = 1$ *it is:*

$$\delta(p;\mu,\sigma) = p. \tag{21}$$

Proof. We prove the result only for the case of location-scale families. The extension to transformed location scale families is achieved by Lemma 2ii). Let $X_i^* = aX_i + b, a > 0, i = 1,\ldots,n$. By Lemma 4 we have

$$\hat{F}(y|\mathbf{X}^*) = \int_{m\in\mathbb{R}} \int_{s\in\mathbb{R}_{++}} F(y;m,s)\pi(m,s|\mathbf{X}^*)dsdm$$

$$= \int_{m\in\mathbb{R}} \int_{s\in\mathbb{R}_{++}} F\left(\frac{y-m}{s}\right) \frac{1}{a^2}\pi\left(\frac{m-b}{a}, \frac{s}{a}|\mathbf{X}\right) dsdm$$

Making the substitutions $\zeta = \frac{m-b}{a}, \ \eta = \frac{s}{a}$ it follows that

$$\hat{F}(y|\mathbf{X}^*) = \hat{F}\left(\frac{y-b}{a}|\mathbf{X}\right) \Rightarrow \hat{F}^{-1}(p|\mathbf{X}^*) = a\hat{F}^{-1}(p|\mathbf{X}) + b$$

Thus $\hat{F}^{-1}(p|\mathbf{X}) = \sigma\hat{F}^{-1}(p|\mathbf{Z}) + \mu$, where $Z_i = \frac{X_i-\mu}{\sigma}, \ i = 1,\ldots,n$. Consequently

$$\delta(p;\mu,\sigma) = E_{\mu,\sigma}\left[F\left(\hat{F}^{-1}(p|\mathbf{X});\mu,\sigma\right)\right],$$

from which equation 20 directly follows. Equation 21 is obtained as a direct application of Proposition 1 and Example 1 in [22]. A necessary condition of the Proposition is the 'invariance of the predictive region,' which in our notation means the requirement that $Y \leq \hat{F}^{-1}(p|\mathbf{X}) \Leftrightarrow aY + b \leq \hat{F}^{-1}(p|a\mathbf{X} + b), \ a > 0$. This follows directly from the statement $\hat{F}^{-1}(p|a\mathbf{X} + b) = a\hat{F}^{-1}(p|\mathbf{X}) + b$ which has been shown earlier in the proof.

$\qquad\square$

References

[1] Jonkman SN, Kok M, Vrijling JK. Flood risk assessment in the Netherlands: a case study for Dike Ring South Holland. Risk Analysis. 2008;28(5):1357-73.

[2] de Haan L. Fighting the arch-enemy with mathematics. Statistica Neerlandica. 1990;44:45-68.

[3] Ellingwood BR. Earthquake risk assessment of building structures. Reliability Engineering and System Safety. 2001;74:251-62.

[4] Anderson RM. How well are we managing the influenza A/H1N1 pandemic in the UK? British Medical Journal. 2009;339:181-82

[5] Basili M, Franzini M. Understanding the risk of an avian flu pandemic: rational waiting or precautionary failure? Risk Analysis. 2006;26(3):617-30.

[6] US Department of Energy. Strategic Petroleum Reserve. Available at: http://fossil.energy.gov/programs/reserves/spr/, Accessed: July 28, 2010.

[7] McNeil AJ, Frey R, Embrechts P. Quantitative Risk Management: Concepts, Techniques, Tools.NJ: Princeton University Press, 2005.

[8] European Commission. Solvency II Framework Directive (European Parliament legislative resolution of 22 April 2009 on the amended proposal for a directive of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance). Available at: http://ec.europa.eu/internal_market/insurance/solvency/index_en.htm, Accessed: August 12, 2009.

[9] Bedford T, Cooke R. Probabilistic Risk Analysis: Foundations and Methods. Cambridge: Cambridge University Press, 2001.

[10] López Droguett E, Mosleh A. Bayesian methodology for model uncertainty using model performance data. Risk Analysis. 2008;28(5):1457-76.

[11] Aven T. On the need for restricting the pprobabilistic analysis in risk assessments to variability. Risk Analysis. 2010;30(3):354-60.

[12] Cairns AJG. A discussion of parameter and model uncertainty in insurance. Insurance: Mathematics and Economics. 2000;27:31330.

[13] Grossi P, Kunreuther H (eds). Catastrophe Modeling: A New Approach to Managing Risk. NY: Springer, 2005.

[14] Cummins JD, Lewis CM. Catastrophic events, parameter uncertainty and the breakdown of implicit long-term contracting: The case of terrorism insurance. J of Risk and Uncertainty. 2003; 26(2):153-77.

[15] Cummins JD, Lewis CM, Wei R. The market value impact of operational loss events for US banks and insurers. J of Banking and Finance. 2006; 30(10):2605-34.

[16] Verrall RJ, England, PD. Predictive distributions of outstanding liabilities in general insurance. Annals of Actuarial Science. 2006;1(2):221-70.

[17] Powers MR, Venezian EC, Jucá IB. Of happy and hapless regulators: the asymptotics of ruin. Insurance: Mathematics and Economics. 2003;32:31730.

[18] Financial Services Authority. Insurance Sector Briefing: ICAS – one year on. November 2005. Available at: www.fsa.gov.uk/pubs/other/icas_isb.pdf, Accessed August 12, 2009.

[19] Borowicz JM, Norman JP. The effects of parameter uncertainty in the extreme event fequency-severity model. Presented at the 28th International Congress of Actuaries, Paris, 28 May-2 June 2008. Available at: www.ica2006.com/Papiers/3020/3020.pdf, Accessed August 12, 2009.

[20] Richards S. Investigating parameter risk for Solvency II and ICAS. Life and Pensions. 2009 Jun: 36-40.

[21] Barndorff-Nielsen OE, Cox DR. Prediction and Asymptotics. Bernoulli. 1996;2(4):319-40.

[22] Severini TA, Mukerjee R, Ghosh M. On an exact probability matching property of right-invariant priors. Biometrika. 2002;89(4):952-57.

[23] Olsen JR, Lambert JH, Haimes YY. Risk of Extreme Events Under Nonstationary Conditions. Risk Analysis. 1998;18(4):497-510.

[24] Datta GS, Mukerjee R, Ghosh M, Sweeting TJ. Bayesian prediction with approximate frequentist validity. Annals of Statistics. 2000;28(5):1414-26.

[25] Embrechts P, Klüppelberg C, Mikosch T.Modelling Extremal Events for Insurance and Finance. Berlin: Springer, 1997.

[26] Kou SG, Wang H. Option pricing under a double exponential jump diffusion model. Management Science. 2004;50(9):1178-1192.

[27] Klugman SA, Panjer HH, Willmott GE. Loss Models: From Data to Decisions. NJ: Wiley, 2004.

[28] Landwehr JM, Matalas, NC, Wallis, JR. Probability weighted momentscompared with some traditional techniques in estimating Gumbel parameters and quantiles. Water Resources Research. 1979;15:1055-64.

[29] Hosking JRM, Wallis JR, Wood EF. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics. 1985;27:251-61.

[30] Jeffreys H. Theory of Probabiilty. London: Oxford University Press, 1961.

[31] Lambrigger DD, Shevchenko PV, Wüthrich MV. The quantification of operational risk using internal data, relevant external data and expert opinion. J Operational Risk 2. 2007;3:3-27.

[32] Basel Committee on Banking Supervision. Results of the 2008 loss data collection exercise for operational risk. Bank of International Settlements, 2009.