City Research Online

City, University of London Institutional Repository

Citation: Bignozzi, V. and Tsanakas, A. (2016). Model uncertainty in risk capital measurement. Journal of Risk, 18(3), pp. 1-24.

This is the submitted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/14620/

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

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

Model uncertainty in risk capital measurement

Valeria Bignozzi[1] and Andreas Tsanakas[2]

Abstract

The required solvency capital for a financial portfolio is typically given by a tail risk measure such as Value-at-Risk. Estimating the value of that risk measure from a limited, often small, sample of data gives rise to potential errors in the selection of the statistical model and the estimation of its parameters. We propose to quantify the effectiveness of a capital estimation procedure via the notions of residual estimation risk and estimated capital risk. It is shown that for capital estimation procedures that do not require the specification of a model (eg historical simulation) or for worst-case scenario procedures the impact of model uncertainty is substantial, while capital estimation procedures that allow for multiple candidate models using Bayesian methods, partially eliminate model error. In the same setting, we propose a way of quantifying model error that allows to disentangle the impact of model uncertainty from that of parameter uncertainty. We illustrate these ideas by simulation examples considering standard loss and return distributions used in banking and insurance.

AMS 2000 Subject Classification: 62P05.

Keywords: Model uncertainty, Model error, Historical simulation, Worst-case approach, Bayesian model averaging, Value-at-Risk.

1 Introduction

A risk measure ρ is a functional that assigns to every financial loss Y a real number $\rho(Y)$. Such a number is intended to summarize statistical information related to the expected level and variability of the loss Y. The value of the risk measure $\rho(Y)$ can be taken to represent the minimal amount of capital that added to Y makes it acceptable (Artzner et al., 1999). Risk measures are generally used to compute regulatory capital requirements. For example, according to Basel III (Basel Committee on Banking Supervision, 2013), the measure of market risk is based on the Value-at-Risk (VaR) of a portfolio with confidence level 99% and 10-days time-horizon. The impending European Solvency II regime postulates a Solvency Capital Requirement set using the Value-at-Risk of the net asset position at the 99.5% level (European Commission, 2010), under a 1-year time horizon.

Risk measures used in regulation are typically law-invariant, that is, they depend uniquely on the probability distribution (here also denoted *model*) of the financial loss. Once a model for Y has been fixed, the computation of $\rho(Y)$ is generally straightforward and can be easily solved either analytically or numerically. For instance, the Value-at-Risk at level p for a financial loss Y can be obtained by simply inverting the distribution function of Y. Selecting the most appropriate model for Y can be challenging in practice. To start with, a model is a mathematical tool used to represent, *in a simplified way*, possible outcomes for the financial loss Y – thus model error is an unavoidable feature of modeling. It is not the purpose of this paper to digress on the limits of mathematical models to describe reality; interesting discussions can be

[1]Corresponding author. School of Economics and Management, University of Florence, Italy; valeria.bignozzi@unifi.it.

[2]Cass Business School, City University London; a.tsanakas.1@city.ac.uk.

found in Danielsson et al. (2001), Colander et al. (2009) and Lawson (2009). Furthermore, it is common practice to estimate models from a set of available data, often of limited size, with few observations in the tail. As a consequence, the estimate of the risk measure (and thus the required capital) will depend on observed data and generally differ from its theoretical value. Limited data do not allow the identification of the correct model, for example due to the low power of statistical Goodness-of-Fit tests observed in practical applications (Frankland et al., 2014), an issue made worse by structural changes in the data generating process.

There are three main approaches to estimating a model:

i) A model F_θ is specified and its unknown parameter(s) θ are estimated from available data, eg via Maximum Likelihood. These are generally referred to as *parametric approaches*;

ii) The distribution for Y is estimated empirically from available data; these are called *model-free approaches*;

iii) A set of candidate models is considered and their parameters are estimated using approaches in i), then either one or an average of those models is used.

For none of those approaches it is immediately clear how the estimated risk measure should be adjusted to reflect the potential for model error.

Approaches in *i*) focus on the estimation of the parameters assuming that a model for Y is known; see Gerrard and Tsanakas (2011), Bignozzi and Tsanakas (2015). Method *ii*) has the advantage that few assumptions on the model have to be made, however it requires a large sample of data to give accurate results and hence is often not practical. Finally approaches in *iii*) permit a systematic treatment of model uncertainty, however they also present some issues: a) the set of candidate models considered plays a central role on the capital and becomes itself a source of possible error; b) these approaches can lead to wrong conclusions when the data generating process is not well described by any model in the set; c)they also require the estimation of parameters. Here we will focus on approaches of the type *ii*) and *iii*)

Before continuing it may be useful to distinguish model uncertainty from parameter uncertainty. The former corresponds to uncertainty arising from not knowing the model; the latter arises from uncertainty about the true parameters, assuming that the model has been correctly chosen. Parameter/model uncertainty lead to parameter/model error when the wrong parameters/model are used. The distinction between model and parameter error is arbitrary and often dictated by practical modeling concerns. A typical example is that of nested models: mistaking a t-distribution for a Normal can be viewed as either model error (wrong distribution family) or parameter error (right family but wrong degrees of freedom). In any risk capital estimation procedure, both parameter and model uncertainty will be at work. It is then not obvious how to disentangle the one effect from the other, when considering potential deviations of the estimated capital from its theoretical value.

The impact of model uncertainty on risk measurement has been investigated in several strands of the literature. Cont et al. (2010) show that different risk measures exhibit different sensitivities to different estimation procedures (in particular they focus on approaches *i*) and *ii*)) and emphasize the importance of a risk measure estimate's robustness with respect to changes in the sample. Krätschmer et al. (2014) develop a more general notion of robustness that allows comparisons between convex risk measures. Barrieu and Scandolo (2015) mainly focus on model risk within a class of candidate models (corresponding to approaches in *iii*)). They propose three different measures: the absolute, relative and local measure of model risk. These measures all depend on the highest and/or lowest capital available for the model set considered. Model uncertainty arising from ignorance of the dependence structure between portfolio risks has been investigated by Embrechts et al. (2013). Recently, Boucher et al. (2014) proposed an

approach to empirically adjust risk measure estimates in order to limit the impact of model uncertainty.

In this paper, we propose a frequentist approach to measure the impact of model uncertainty, introducing two specific measures of model risk. The estimated capital is considered a random variable via its dependence on a random sample. Gerrard and Tsanakas (2011) investigate the changes in the probability of future losses exceeding capital, when the estimated capital is random due to parameter uncertainty. Bignozzi and Tsanakas (2015) quantify the impact of parameter uncertainty on risk measures beyond VaR. Here we extend and complement those approaches to the more realistic case where model uncertainty is present. Considering the estimated capital as a random variable, allows us to treat model uncertainty as just another source of risk and to quantify, in monetary units, the additional capital required in respect of model uncertainty. Furthermore, this capital adjustment generally depends on the size of the dataset, distinguishing the approach proposed here from other measures of model risk in the literature.

The first measure of model risk considered, termed *residual estimation risk*, is a direct extension of the measure of parameter risk proposed by Bignozzi and Tsanakas (2015), and consists in computing model risk as the extra capital required to make a financial position acceptable after the random capital has been added to it. The second measure, called *estimated capital risk*, is introduced in this paper and quantifies the riskiness of holding a capital different from the theoretical one.

In Section 2 we review the basic properties of risk measures and introduce the two measures of model risk. In Section 3 we show how these model risk measures can be used to assess the impact of model uncertainty when the model-free historical simulation and worst-case approaches are used to estimate the required capital. In Section 4 Bayesian methods are considered, which allow us to disentangle parameter from model unceratinty.

2 Risk measures and measures of model uncertainty

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space and $\mathcal{X} \subseteq L^0(\Omega, \mathcal{F}, \mathbb{P})$ be a convex cone of measurable random variables. Any random variable $Y \in \mathcal{X}$ represents a financial loss over a fixed time horizon. We write $Y \sim F$ to denote that Y has distribution function (df) F and when necessary indicate with F_Y the df of a random variable Y. F is also called a *model* for Y.

A distribution *family* F_θ is a set of models indexed by a parameter vector $\theta \in \Theta \subseteq \mathbb{R}^d$. For example, a location-scale family is a family of distribution functions characterized by a location and a scale parameter $\mu \in \mathbb{R}, \sigma > 0$, such that for every random variable Y in the family, the variable $a + bY$, $a \in \mathbb{R}, b > 0$ belongs to the same family. We denote with $F_{0,1}$ the df of a random variable with $\mu = 0$ and $\sigma = 1$ and $F_{\mu,\sigma}$ the distribution with parameters (μ, σ). For every $x \in \mathbb{R}$, $F_{\mu,\sigma}(x) = F_{0,1}(\frac{x-\mu}{\sigma})$.

A risk measure ρ is a functional $\rho: \mathcal{X} \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ that assigns to every loss $Y \in \mathcal{X}$ a real number $\rho(Y)$. In the paper we always work with risk measures satisfying the following properties without further mention (see for instance Föllmer and Schied, 2011). For every $Y_1, Y_2 \in \mathcal{X}$:

(a) *Law-invariance.* If $Y_1 \overset{d}{=} Y_2$, then $\rho(Y_1) = \rho(Y_2)$;

(b) *Translation invariance.* If $m \in \mathbb{R}$, then $\rho(Y_1 + m) = \rho(Y_1) + m$;

(c) *Monotonicity.* If $Y_1 \geq Y_2$ $\mathbb{P}$-a.s., then $\rho(Y_1) \geq \rho(Y_2)$,

where $\overset{d}{=}$ denotes equality in distribution. Thanks to law-invariance, risk measures may be defined as functionals on sets of distribution functions rather than random variables. Thus, for

a random variable $Y \sim F$, we use interchangeably the notations $\rho(Y)$ and $\rho[F]$. Other properties often proposed in the literature are:

(d) *Positive homogeneity.* If $\lambda \geq 0$, then $\rho(\lambda Y_1) = \lambda \rho(Y_1)$;

(e) *Subadditivity.* $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$.

A risk measure ρ that satisfies properties (b) and (c) can be used to calculate regulatory capital requirements. Its value $\rho(Y)$ identifies the smallest amount of capital that, added to the loss Y makes it acceptable, that is, leads to

$$\rho(Y - \rho(Y)) = 0. \tag{1}$$

A standard example of risk measure that satisfies (a), (b), (c) and (d) is the Value-at-Risk (VaR) at level p, defined as

$$\mathrm{VaR}_p(Y) := \inf\{m \in \mathbb{R} \mid \mathbb{P}(Y \leq m) \geq p\}, \quad p \in (0, 1). \tag{2}$$

Other examples are Expected Shortfall (that satisfies also (e)) as well as the more general class of distortion risk measures (Wang, 1996) and the closely related spectral risk measures (Acerbi, 2002).

When the distribution function F of a random loss Y is unknown, it has to be estimated from a set of available data $\mathbf{x} = (x_1, \ldots, x_n)$. The data are a realization of the random vector $\mathbf{X} = (X_1, \ldots, X_n)$, where each variable X_i, $i = 1, \ldots, n$ has df $F^{(i)}$. We assume throughout that $X_1, \ldots, X_n, Y$ are independent. Furthermore we assume that $Y \sim F$ and $F^{(i)} = F$ for every $i = 1, \ldots, n$, and write $\mathbf{X}, Y \sim F$.[3] An estimator of the risk measure $\rho(Y)$ is a function of the random sample $\mathbf{X}$ and will be denoted by $\eta(\mathbf{X})$. We distinguish between $\rho(Y) = \rho[F]$, the theoretical value of the risk measure, and $\eta(\mathbf{X})$, a random variable reflecting estimation volatility.

For example, the df F may be estimated via the empirical df

$$\hat{F}_{\mathbf{X}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}, \tag{3}$$

where $\mathbb{I}$ denotes the indicator function. Then the estimated capital would be $\eta(\mathbf{X}) = \rho[\hat{F}_{\mathbf{X}}]$. Of course, for each fixed outcome $\mathbf{x} = (x_1, \ldots, x_n)$ of the random vector $\mathbf{X}$, $\eta(\mathbf{x})$ is a real number. In general, $\eta(\mathbf{x}) \neq \rho(Y)$ so that $\eta(\mathbf{x})$ cannot be interpreted anymore as the minimal amount of capital that makes Y acceptable. Thus, estimating the risk measure from a finite sample of data leads to under/over-estimates of the required capital.

The dependence of η on the random vector $\mathbf{X}$ makes it possible to express model uncertainty as a source of risk (ie as a random variable) itself. The first measure of model risk considered is a direct extension of the measure of parameter risk proposed by Bignozzi and Tsanakas (2015):

Definition 2.1. *For $Y, \mathbf{X} \sim F$, we define the Residual Estimation Risk (RR) for the risk measure ρ, as*

$$\mathrm{RR}(F, \eta, \rho) := \rho(Y - \eta(\mathbf{X})). \tag{4}$$

[3]This i.i.d. assumption is not strictly necessary and is made for simplicity of exposition. If the assumption was dropped, the RR and ECR measures subsequently defined would depend on the joint distributions of $(\mathbf{X}, Y)$ and of $\mathbf{X}$ respectively.

RR also depends on the number of data n, but this is suppressed in the notation. Note the formal similarity between (1) and (4). Since $\eta(\mathbf{X})$ is now a random variable and not a fixed number as it was in (1) it will generally be $\mathrm{RR}(F, \eta, \rho) \neq 0$ so that extra capital will be required to make Y acceptable under model uncertainty. In particular, from the translation invariance property of ρ it follows that

$$\rho\left(Y - \eta(\mathbf{X}) - \mathrm{RR}(F, \eta, \rho)\right) = 0. \tag{5}$$

Thus, $\mathrm{RR}(F, \eta, \rho)$ represents the additional amount of cash that, added to the random capital $\eta(\mathbf{X})$ (ie removed from $Y - \eta(\mathbf{X})$), makes Y acceptable, while considering simultaneously the randomness arising from the stochastic nature of Y and the volatility arising from the capital estimation procedure.

The second measure of model risk that we present is the following:

Definition 2.2. *For $Y, \mathbf{X} \sim F$, we define the Estimated Capital Risk (ECR) for the risk measure ρ, as*

$$\mathrm{ECR}(F, \eta, \rho) := \rho(-\eta(\mathbf{X})) + \rho(Y). \tag{6}$$

Once more, the absence of model uncertainty would imply $\eta(\mathbf{X}) = \rho(Y)$ such that $\mathrm{ECR}(F, \eta, \rho) = 0$. Again, a positive value of the ECR indicates an underestimate of the capital, while $\mathrm{ECR} < 0$ implies an overestimate of the capital.

By rewriting ECR as $\rho(\rho(Y) - \eta(\mathbf{X}))$ the difference between RR and ECR becomes clearer. The former assesses the risk of the financial loss $Y - \eta(\mathbf{X})$, while the latter measures the risk of holding an estimated capital different from the theoretical one. ECR is in general more conservative than RR, as the following result shows.

Proposition 2.3. *For a subadditive risk measure ρ it always holds:*

$$\mathrm{RR}(F, \eta, \rho) \leq \mathrm{ECR}(F, \eta, \rho).$$

Proof. Denote $Z = -\eta(\mathbf{X})$. From the subadditivity property of ρ we have

$$\mathrm{RR}(F, \eta, \rho) = \rho(Y + Z) \leq \rho(Y) + \rho(Z) = \mathrm{ECR}(F, \eta, \rho).$$

□

Similar to the measures of model risk proposed by Barrieu and Scandolo (2015), the RR and ECR are functions of the unknown distribution function F. Thus for calculation of those quantities, a *reference measure* is typically specified. We refer to Alexander and Sarabia (2012) for insights on how to choose this reference measure. The dependence of RR and ECR on the true but unknown distribution F is consistent with other frequentist quality criteria of estimators, such as the bias or Mean-Squared-Error. Of course, the question arises of how one may assess model error by a measure that requires knowledge of the model. In the case where only parameter uncertainty is present, the dependence on F can be eliminated using a bootstrapping approach (Bignozzi and Tsanakas, 2015). However, in the more general case of model uncertainty discussed in this paper, such an approach is not possible. A pragmatic way forward is to evaluate RR and ECR on a range of plausible reference measures, as a form of sensitivity analysis to model error. This is the approach pursued in this paper, particularly in Section 4.

Location-scale families, such as the Normal, Student t or Logistic distributions are often used to model asset returns. Let $\mathbf{X}, Y \sim F_{\mu,\sigma}$, where μ, σ are the location and scale parameters. Then, for a positive homogenous and translation invariant risk measure ρ it is $\rho(Y) = \mu + \sigma\rho[F_{0,1}]$.

More generally, if $Y \overset{d}{=} a + b\tilde{Y}$ for $a \in \mathbb{R}, b > 0$, then $\rho(Y) = a + b\rho(\tilde{Y})$. Furthermore, when $\mathbf{X}, Y \sim F_{\mu,\sigma}$ then standard estimators of the location and scale parameters typically also have location-scale distributions; this is for instance the case when Maximum Likelihood Estimators are used (see Gerrard and Tsanakas, 2011, Lemma 4). As consequence, if we have $Y \overset{d}{=} a + b\tilde{Y}$ and $\mathbf{X} \overset{d}{=} a + b\tilde{\mathbf{X}}$, it typically follows that

$$\eta(\mathbf{X}) = a + b\eta(\tilde{\mathbf{X}}). \tag{7}$$

For more details see Gerrard and Tsanakas (2011). Further examples will be provided in Section 3.1.

In the presence of location-scale families, the following properties can be shown for our measures of model risk.

Proposition 2.4. *Let ρ be a positively homogeneous risk measure. Let $Y, \mathbf{X} \sim F$ where F is a location-scale family, $Y \overset{d}{=} a + b\tilde{Y}$, $a \in \mathbb{R}, b > 0$, and (7) holds. Then we have:*

$$\begin{aligned} \mathrm{RR}(F, \eta, \rho) &= b\mathrm{RR}(\tilde{F}, \eta, \rho) \\ \mathrm{ECR}(F, \eta, \rho) &= b\mathrm{ECR}(\tilde{F}, \eta, \rho). \end{aligned}$$

Proof. From the law-invariance, translation invariance and positive homogeneity of ρ it follows that:

$$\begin{aligned} \mathrm{RR}(F, \eta, \rho) &= \rho(Y - \eta(\mathbf{X})) = \rho(a + b\tilde{Y} - a - b\eta(\tilde{\mathbf{X}})) \\ &= b\rho(\tilde{Y} - \eta(\tilde{\mathbf{X}})) = b\mathrm{RR}(\tilde{F}, \eta, \rho). \end{aligned}$$

The proof for ECR is essentially the same. □

3 Non-parametric approaches

In this section we apply the notions of RR and ECR to assess model risk for the historical simulation and worst-case approaches to capital estimation. The risk measure VaR is used throughout.

3.1 Historical simulation

The simplest model-free method for estimating a risk measure such as VaR is Historical Simulation (HS). It corresponds to computing the risk measure directly on the empirical distribution of Y, ie, it corresponds to setting $\eta(\mathbf{X}) = \rho[\hat{F}_{\mathbf{X}}]$, where $\hat{F}_{\mathbf{X}}$ has been defined in (3). For VaR_p, the procedure is extremely simple: given an i.i.d. sample $\mathbf{X} = (X_1, \ldots, X_n)$ having the same distribution F as Y, consider the order statistics $X_{(1)}, \ldots, X_{(n)}$ obtained by rearranging $\mathbf{X}$ in increasing order, such that $X_{(1)} \leq \ldots \leq X_{(n)}$. Then $\mathrm{VaR}_p(Y)$, corresponding to the p^{th}-quantile of Y, can be estimated as $X_{(\lfloor np \rfloor)}$, where $\lfloor np \rfloor$ is the lower integer part of np (an average of the np and $p(n+1)$ order statistics may also be considered, but we do not consider this point here). This approach has the clear advantage that no model has to be specified for Y and that is very easy to understand and implement. However, it requires a larger sample of data than for instance, a parametric model to be accurate (see eg Danielsson and de Vries, 1997). We report here the following useful result (see for instance Arnold et al., 1992, Chapter 2, pag 12).

Proposition 3.1. *Let $X_1, \dots, X_n$ be a sample of i.i.d. random variables with df F. Then the df $G_{(i)}$ of the order statistic $X_{(i)}$ is given by:*

$$G_{(i)}(x) := \sum_{r=i}^{n} \binom{n}{r} F(x)^r (1 - F(x))^{n-r} \quad \text{for all } i = 1, \dots, n, \text{ and } x \in \mathbb{R}. \tag{8}$$

It follows that we can express the VaR capital computed with HS as a function of the data via $\eta(\mathbf{X}) = X_{(\lfloor np \rfloor)} \sim G_{(\lfloor np \rfloor)}$. For VaR the model risk measures can be computed as

$$\mathrm{RR}(F, \eta_{HS}, \mathrm{VaR}) = \mathrm{VaR}(Y - \eta_{HS}(\mathbf{X})), \tag{9}$$

$$\mathrm{ECR}(F, \eta_{HS}, \mathrm{VaR}) = \mathrm{VaR}(-\eta_{HS}(\mathbf{X})) + \mathrm{VaR}(Y). \tag{10}$$

In Tables 1 and 2 we report the RR and ECR respectively when $Y, \mathbf{X}$ are drawn from a LogNormal or an Inverse Gamma distribution and the capital is estimated using HS. Such distributions are used to model insurance claims or operational risk losses. The parameters are chosen such that both distributions have a mean of 100 and a standard deviation of 25. The model risk measures are also reported as percentages of the capital required in absence of model uncertainty, that is, $\mathrm{VaR}_p(Y)$. For example, for the LogNormal distribution considered here, the theoretical capital is $\mathrm{VaR}_{0.95}(Y) = 145.42$, the residual estimation risk with HS and a data sample of size $n = 100$ is 3.04 which is approximately the 2.1% of the theoretical capital.

As expected, both measures of model risk decrease with the size of the data set and increase when we go further into the tail (increasing the value of p). Furthermore, the ECR is much higher than RR, demonstrating that ECR is a more conservative measure of model risk than RR. (Since VaR_p is not a subadditive measure, this relation is not guaranteed to hold by Proposition 2.3.)

More can be said about the model risk measures for HS, when the loss belongs to a location-scale family. From Proposition 3.1 the next result immediately follows.

Proposition 3.2. *Given an i.i.d. random vector $\mathbf{X}$ with df $F_{\mu,\sigma}$ that belongs to a location-scale family, then also the order statistic $X_{(i)}$ belongs to a location-scale family with the same parameters μ, σ.*

Proof. From (8),

$$\begin{aligned} G_{\mu,\sigma(i)}(x) &= \sum_{r=i}^{n} \binom{n}{r} F_{\mu,\sigma}(x)^r \left(1 - F_{\mu,\sigma}(x)\right)^{n-r} \\ &= \sum_{r=i}^{n} \binom{n}{r} F_{0,1}\left(\frac{x-\mu}{\sigma}\right)^r \left(1 - F_{0,1}\left(\frac{x-\mu}{\sigma}\right)\right)^{n-r} \\ &= G_{0,1(i)}\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

□

Corollary 3.3. *Let ρ be positively homogeneous, $Y, \mathbf{X} \sim F_{\mu,\sigma}$ belong to a location-scale family and let $\tilde{Y}, \tilde{\mathbf{X}}$ be such that $Y \overset{d}{=} \mu + \sigma\tilde{Y}$ and $\mathbf{X} \overset{d}{=} \mu + \sigma\tilde{\mathbf{X}}$, for $\mu \in \mathbb{R}, \sigma > 0$. Then,*

$$\mathrm{RR}(F, \eta_{HS}, \mathrm{VaR}) = \sigma \mathrm{RR}(F_{0,1}, \eta_{HS}, \mathrm{VaR}),$$
$$\mathrm{ECR}(F, \eta_{HS}, \mathrm{VaR}) = \sigma \mathrm{ECR}(F_{0,1}, \eta_{HS}, \mathrm{VaR}).$$

Proof. Since $\eta_{HS}(\mathbf{X}) = \mu + \sigma\eta_{HS}(\tilde{\mathbf{X}})$, the result follows from Propositions 2.4 and 3.2. □

Hence, the measures of model risk do not depend on the true value of the location parameter and are proportional to the scale parameter.

Table 1: Values for the Residual Risk in (9) (expressed also as a percentage of $\text{VaR}_p(Y)$) for a LogNormal distribution with parameters (4.574,0.246) and an Inverse Gamma with parameters (18,1700), a data sample of size n, and risk measure VaR_p.

LogNormal(4.574,0.246)	$n = 100$		$n = 200$		$n = 500$	
	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$
$p = 0.95$	3.04	2.09%	1.58	1.09%	0.65	0.44%
$p = 0.99$	11.22	6.52%	6.43	3.74%	2.89	1.68%
InvGamma(18.00,1700.00)	$n = 100$		$n = 200$		$n = 500$	
	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$
$p = 0.95$	3.35	2.29%	1.74	1.19%	0.72	0.49%
$p = 0.99$	13.04	7.37%	7.61	4.30%	3.40	1.92%

Table 2: Values for the Estimated Capital Risk in (10) (expressed also as a percentage of $\text{VaR}_p(Y)$) for a LogNormal distribution with parameters (4.574,0.246) and an Inverse Gamma with parameters (18,1700), a data sample of size n, and risk measure VaR_p.

LogNormal(4.57,0.25)	$n = 100$		$n = 200$		$n = 500$	
	ECR	% $\text{VaR}_p(Y)$	ECR	% $\text{VaR}_p(Y)$	ECR	% $\text{VaR}_p(Y)$
$p = 0.95$	12.84	8.83%	9.07	6.24%	5.70	3.92%
$p = 0.99$	31.01	18.03%	23.29	13.54%	15.47	9.00%
InvGamma(18.00,1700.00)	$n = 100$		$n = 200$		$n = 500$	
	ECR	% $\text{VaR}_p(Y)$	ECR	% $\text{VaR}_p(Y)$	ECR	% $\text{VaR}_p(Y)$
$p = 0.95$	13.98	9.57%	9.92	6.79%	6.27	4.29%
$p = 0.99$	35.61	20.14%	27.02	15.28%	18.18	10.29%

3.2 Worst-case approach

In this section we consider capital estimation procedures that require the *a priori* specification of a set of possible models. We denote such a *model set* as $\mathcal{M}$. $\mathcal{M}$ may or may not be finite. As before, the true model is denoted by F and the data/future loss generated by this model are given by $Y, \mathbf{X} \sim F$.

In general there is no guarantee that F will actually be within the model set $\mathcal{M}$ used for capital estimation. Furthermore, when datasets are small, simple models (eg with small number

of parameters) tend to be used to avoid overfitting; this means that a correct but more complex model F may not be considered in statistical estimation. Thus, the possibility that $F \notin \mathcal{M}$ implies that the use of an estimator based on $\mathcal{M}$ cannot in general completely eliminate model error.

In this section we consider a worst-case approach, while a Bayesian perspective is adopted in Section 4.

The worst-case approach, presented here, is one where the capital is set according to the most conservative model in a set $\mathcal{M}$. This method finds its root in Gilboa and Schmeidler (1989) on robust utility maximization and its use is widespread among practitioners and academics (Barrieu and Scandolo, 2015). The idea behind it is straightforward: in order to be on the safe side we hold capital according to the worst-case.

The set of models $\mathcal{M}$ considered plays a central role. Intuitively, the wider the set, the higher the capital. It is clear that a model set that is too wide can easily lead to trivial results, such as an infinite capital requirement. Moreover, when the true model F does not belong to the set $\mathcal{M}$ specified, this approach loses its interpretation because the true model may be more conservative than any model in $\mathcal{M}$.

In this section we consider the model set discussed in Barrieu and Scandolo (2015). The set of candidate models, that we denote $\mathcal{M}_{\mu,\sigma}$, consists of all the distribution functions with same mean μ and standard deviation σ. The capital is then computed as the *worst-case* (or the highest) capital in $\mathcal{M}_{\hat{\mu}(\mathbf{X}),\hat{\sigma}(\mathbf{X})}$, where $(\hat{\mu}(\mathbf{X}), \hat{\sigma}(\mathbf{X}))$ are non-parametric estimators of the mean and standard deviation. For simplicity, we assume that μ, σ are respectively estimated via the sample mean $\hat{\mu}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$ and sample standard deviation $\hat{\sigma}(\mathbf{X}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}(\mathbf{X}))^2}$. It follows that the sample mean and standard deviation can be written respectively as

$$\hat{\mu}(\mathbf{X}) = \mu + \frac{\sigma}{n}\sum_{i=1}^{n}\tilde{X}_i, \quad \hat{\sigma}(\mathbf{X}) = \sigma\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i - \hat{\mu}(\tilde{\mathbf{X}}))^2}, \tag{11}$$

where $X_i = \mu + \sigma\tilde{X}_i$, for all $i = 1, \ldots, n$.

The worst-case capital estimator is denoted by $\eta_{\hat{\mu},\hat{\sigma}}$ and given by

$$\eta_{\hat{\mu},\hat{\sigma}}(\mathbf{X}) = \sup_{G \in \mathcal{M}_{\hat{\mu}(\mathbf{X}),\hat{\sigma}(\mathbf{X})}} \rho[G], \tag{12}$$

that is, the highest possible value of the risk measure over the set of distributions that have mean and standard deviation equal to the sample mean and sample standard deviation respectively. Note that, given the variability of the estimators of the mean and standard deviation, it is *not* possible to identify the set $\mathcal{M}_{\mu,\sigma}$ with the set $\mathcal{M}_{\hat{\mu}(\mathbf{X}),\hat{\sigma}(\mathbf{X})}$. Therefore it will generally be the case that $F \notin \mathcal{M}_{\hat{\mu}(\mathbf{X}),\hat{\sigma}(\mathbf{X})}$.

For any $V \sim G \in \mathcal{M}_{\mu,\sigma}$, we can write $V \overset{d}{=} \mu + \sigma\tilde{V}$, where $\tilde{V} \sim \tilde{G} \in \mathcal{M}_{0,1}$, with $\mathcal{M}_{0,1}$ the class of distributions with mean 0 and standard deviation 1. Translation invariance and positive homogeneity imply that $\rho[G] = \rho(V) = \mu + \sigma\rho(\tilde{V}) = \mu + \sigma\rho[\tilde{G}]$. Consequently

$$\eta_{\hat{\mu},\hat{\sigma}}(\mathbf{X}) = \hat{\mu}(\mathbf{X}) + \hat{\sigma}(\mathbf{X}) \sup_{\tilde{G} \in \mathcal{M}_{0,1}} \rho[\tilde{G}].$$

Hence it is sufficient to find the worst-case standardized distribution in $\mathcal{M}_{0,1}$, in order to derive the capital estimator $\eta_{\hat{\mu},\hat{\sigma}}$.

Once more, the measures of model risk can be expressed via standardized distributions.

Corollary 3.4. *Let ρ be positively homogeneous and $Y, \mathbf{X} \sim F$, with mean and standard deviation μ and σ respectively. Let $\tilde{Y}, \tilde{\mathbf{X}}$ be such that $Y \overset{d}{=} \mu + \sigma\tilde{Y}$ and $\mathbf{X} \overset{d}{=} \mu + \sigma\tilde{\mathbf{X}}$. Then,*

$$\mathrm{RR}(F, \eta_{\hat{\mu},\hat{\sigma}}, \rho) = \sigma\rho\left(\tilde{Y} - \eta_{\hat{\mu},\hat{\sigma}}\left(\tilde{\mathbf{X}}\right)\right),$$
$$\mathrm{ECR}(F, \eta_{\hat{\mu},\hat{\sigma}}, \rho) = \sigma\left(\rho(\tilde{Y}) - \rho\left(\eta_{\hat{\mu},\hat{\sigma}}\left(\tilde{\mathbf{X}}\right)\right)\right).$$

Proof. By (7) the conditions of Proposition 2.4 apply, wherefrom the result immediately follows. □

When the risk measure considered is VaR the worst-case risk measure takes a particularly simple form; by Cantelli's inequality we have (Barrieu and Scandolo, 2015)

$$\sup_{\tilde{G}\in\mathcal{M}_{0,1}} \mathrm{VaR}_p[\tilde{G}] = \sqrt{\frac{p}{1-p}}.$$

In Tables 3 and 4 we report the RR and ECR for the capital estimator $\eta_{\mathcal{M}_{\hat{\mu},\hat{\sigma}}}$, when the reference model used to simulate $(\mathbf{X}, Y)$ is Normal or Student t ($\nu = 3$) respectively. As expected, the capital required according to this approach is very conservatively calculated, leading to negative RR and ECR. However, the robustness of the worst-case approach is demonstrated by noting that eg the Residual Risk reduces from about 319%-326% of capital in the Normal case, to 121%-143% of capital for the heavier tailed t distribution.

4 Capital estimation using Bayesian methods

4.1 Disentangling parameter and model uncertainty

We now describe the Bayesian approach to model and parameter uncertainty; for a detailed discussion see Cairns (2000). If a model is fixed and only the parameters are unknown, the goal of such a method is to construct a new model that incorporates parameter uncertainty. To do so, the parameters of a distribution are considered as random variables themselves with a *prior distribution*. After a sample of data has been collected, the information from the data and the prior distribution is combined into the *posterior distribution*. Finally a *predictive distribution*, is constructed as the weighted average of the model over the parameter space, weighted by the posterior.

In case the model is not known, a set of candidate models $\mathcal{M} = \{F^{(1)}, \ldots, F^{(K)}\}$ is specified. Each model is assigned a prior distribution and a procedure similar to the one just seen for the parameters is applied.

We start by considering a particular model $F^{(k)} \in \mathcal{M}$, and denote $\theta^{(k)} \in \Theta^{(k)} \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$ the parameter vector of the df $F^{(k)}$. The prior $\pi(\theta^{(k)})$ is specified. After collection of data $\mathbf{x}$ (a realization of $\mathbf{X}$), the posterior distribution is calculated according to the Bayes formula:

$$\pi(\theta^{(k)}|\mathbf{x}) = \frac{f^{(k)}(\mathbf{x};\theta^{(k)})\pi(\theta^{(k)})}{\int_{u\in\Theta^{(k)}} f^{(k)}(\mathbf{x};u)\pi(u)du}, \tag{13}$$

where $f^{(k)}(\mathbf{x};\theta^{(k)}) = \prod_{i=1}^n f^{(k)}(x_i;\theta^{(k)})$ is the likelihood of $\mathbf{x} = (x_1, \ldots, x_n)$ under the chosen model. The predictive distribution function, given the data, is then obtained as:

$$\hat{F}^{(k)}(y|\mathbf{x}) = \int_{\theta^{(k)}\in\Theta^{(k)}} F_k(y;\theta^{(k)})\pi(\theta^{(k)}|\mathbf{x})d\theta^{(k)}. \tag{14}$$

Table 3: Values for the Residual Risk and the Estimated Capital Risk with estimator (12) (expressed also as percentages of $\mathrm{VaR}_p(Y)$) for a standard Normal distribution, a data sample of size n, and risk measure VaR_p.

Normal(0,1)	$n = 100$		$n = 200$		$n = 500$	
	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$
$p = 0.95$	-2.67	-162.16%	-2.69	-163.55%	-2.71	-164.42%
$p = 0.99$	-7.43	-319.28%	-7.52	-323.42%	-7.58	-326.00%
Normal(0,1)	$n = 100$		$n = 200$		$n = 500$	
	ECR	$\%\mathrm{VaR}_p(Y)$	ECR	$\%\mathrm{VaR}_p(Y)$	ECR	$\%\mathrm{VaR}_p(Y)$
$p = 0.95$	-2.39	-145.36%	-2.49	-151.39%	-2.58	-156.54%
$p = 0.99$	-6.72	-288.77%	-7.00	-300.69%	-7.23	-310.89%

Table 4: Values for the Residual Risk and the Estimated Capital Risk with estimator (12) (expressed also as percentages of $\mathrm{VaR}_p(Y)$) for a standardized t distribution with 3 degrees of freedom, a data sample of size n, and risk measure VaR_p.

t_3	$n = 100$		$n = 200$		$n = 500$	
	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$
$p = 0.95$	-2.37	-100.56%	-2.55	-108.19%	-2.70	-114.82%
$p = 0.99$	-5.51	-121.33%	-6.01	-132.38%	-6.48	-142.65%
t_3	$n = 100$		$n = 200$		$n = 500$	
	ECR	$\%\mathrm{VaR}_p(Y)$	ECR	$\%\mathrm{VaR}_p(Y)$	ECR	$\%\mathrm{VaR}_p(Y)$
$p = 0.95$	-0.75	-31.86%	-1.00	-42.36%	-1.25	-53.28%
$p = 0.99$	-1.96	-43.23%	-2.62	-57.78%	-3.32	-73.10%

For a fixed sample $\mathbf{X} = \mathbf{x}$, the estimated capital according to the predictive distribution is $\rho\left[\hat{F}^{(k)}(\cdot|\mathbf{x})\right]$. When considering a random sample $\mathbf{X}$, the capital estimator thus becomes

$$\eta^*_{F^{(k)}}(\mathbf{X}) = \rho\left[\hat{F}^{(k)}(\cdot|\mathbf{X})\right]. \tag{15}$$

The following proposition shows how it is possible to disentangle the impact of parameter uncertainty from model uncertainty when the residual estimation risk is used.

Proposition 4.1. *Let $Y, \mathbf{X} \sim F_{\mu,\sigma}$ belong to a location-scale family. Let $\pi(\mu,\sigma) = \frac{1}{\sigma}$ and denote $\mathrm{VaR}_p[\hat{F}_{\mu,\sigma}(\cdot|\boldsymbol{X})]$ the estimated capital according to the predictive distribution. Then,*

$$\mathrm{VaR}_p(Y - \mathrm{VaR}_p[\hat{F}_{\mu,\sigma}(\cdot|\boldsymbol{X})]) = 0,$$

that is, the Bayesian approach eliminates completely the residual estimation risk (4).

Proof. The proof follows directly from the results in Gerrard and Tsanakas (2011), who show that $\mathbb{P}(Y \leq \mathrm{VaR}_p[\hat{F}_{\mu,\sigma}(\cdot|\mathbf{X})]) = p$, wherefrom the stated result is derived. □

Hence, when the *non-informative prior* $\pi(\mu,\sigma) = 1/\sigma$ is used with location-scale families and there is no model uncertainty (the family $F_{\mu,\sigma}$ is known), the RR for VaR is equal to zero. For other law-invariant risk measures, it was shown in Bignozzi and Tsanakas (2015) that the same approach can be very effective in approximately eliminating residual risk.

Consider now the case where also model uncertainty is present. As a consequence of Proposition 4.1, by using the residual estimation risk we can guarantee that in case the model selected is the correct one, the RR would be zero. Hence, when capital estimation is carried out with such a method, RR is only due the potential of model mis-specification. For this reason, in what follows we will consider different approaches where the parameters of the candidate models are always estimated using a Bayesian approach. Furthermore we will focus on the residual estimation risk and will not report results on the estimated capital risk.

4.2 Examples

In Section 3 it was shown how the residual estimation risk and the estimation capital risk can be used to assess the effectiveness of non-parametric estimation procedures in presence of model uncertainty. In the following, we will consider three more approaches to model uncertainty, all based on Bayesian arguments. In a simulation study we show how the residual estimation risk can be used to compare different estimation procedures.

Again a VaR risk measure, $\rho \equiv \mathrm{VaR}_{0.99}$, is considered, and 3 potential loss distributions (reference models) are used to generate $Y, \mathbf{X}$:

1. Weibull distribution (WB);
2. LogNormal distribution (LN);
3. Inverse Gamma distribution (IG).

Densities and related quantities for those distributions are stated in the Appendix. The mean, standard deviation, parameters, and VaR measures at 99% and 99.5% levels are displayed in Table 5.

These distributions provide a variety of shapes. The Weibull distribution (for the given parameters) tends to be light tailed, producing few extremes. The LogNormal distribution has a heavier (subexponential) tail, while the Inverse Gamma distribution has heavy (regularly varying) tails (for details see for example McNeil et al., 2005; Klugman et al., 2008). This can be glanced from the VaR measures reported in Table 5. In the sequel we use the terms "light-" and "heavy-tailed" informally, as reflecting the ordering of the 3 distributions considered apparent in Table 5 (the Weibull having the lightest and the Inverse Gamma having the heaviest tail).

All the methods that we will see in the following sections will require the specification of a model set $\mathcal{M}$. To aid comparison we will use a model set consisting of the following distributions: $F^{(1)} \equiv Gamma(\alpha,\beta)$ (with unknown scale parameter β and fixed shape parameter $\alpha = 16$), $F^{(2)} \equiv LogNormal(\mu,\sigma)$ (with unknown μ and fixed $\sigma = 0.246$), $F^{(3)} \equiv Inverse\ Gamma(\alpha,\delta)$

Table 5: Reference models used in simulation study of model error.

	Mean	**St. Dev.**	**1st par.**	**2nd par.**	$\text{VaR}_{0.99}$	$\text{VaR}_{0.995}$
$\mathbf{WB}(\gamma, \beta)$	100.00	25.00	109.521	4.542	153.29	158.10
$\mathbf{LN}(\mu, \sigma)$	100.00	25.00	4.575	0.246	172.03	182.93
$\mathbf{IG}(\alpha, \beta)$	100.00	25.00	18.000	1700.000	176.78	190.09

(with unknown scale parameter and fixed shape parameter $\alpha = 18$). We thus assume that the shape parameters are known and the scale parameters unknown and write

$$\mathcal{M} = \left\{ F^{(1)}_{16,\beta}, F^{(2)}_{\mu,0.246}, F^{(3)}_{18,\delta} \right\}. \tag{16}$$

The model set used in capital estimation overlaps with, but is not identical to, the set of reference models in Table 5 used to generate random samples.

4.2.1 Bayesian worst-case approach (BWC)

The first method that we consider is the Bayesian worst-case approach (BWC). The idea is similar to the one presented in Section 3.2, but we consider a finite model set. For each model $F^{(k)} \in \mathcal{M} \equiv \{F^{(1)}, \ldots, F^{(K)}\}$, we calculate the risk measure according to the predictive distribution $\hat{F}^{(k)}(\cdot|\mathbf{X})$, that is, $\text{VaR}_p[\hat{F}^{(k)}(\cdot|\mathbf{X})]$. Then capital is set according to the most conservative model

$$\eta_{\mathcal{M},BWC}(\mathbf{X}) = \max_{k \in K} \rho[\hat{F}^{(k)}(\cdot|\mathbf{X})]. \tag{17}$$

Table 6: Values for Residual Risk (expressed also as a percentage of $\text{VaR}_p(Y)$) for the Bayesian Worst Case method, with model set (16). Reference models are given in Table 5. The data sample is of size n and the risk measure VaR_p is used.

	$n = 10$		$n = 50$		$n = 150$	
Reference model	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$
Weibull(109.52, 454)	-15.08	-8.77%	-19.79	-11.51%	-20.77	-12.08%
LogNormal(4.57,0.25)	-3.29	-1.91%	-3.90	-2.27%	-4.09	-2.38%
InvGamma(18.0,1700.00)	-0.16	-0.09%	0.00	0.00%	0.00	0.00%

In Table 6 the residual estimation risk for the Bayesian worst-case capital estimator is displayed. It is seen that the numbers are all negative. This is consistent with the implicit conservativeness of the approach. However, such conservativeness can be easily seen as excessive. Especially for light-tailed distributions such as the Weibull, the residual risk is substantially

different to zero. Moreover, the residual risk does not generally decrease with the sample size, since increasing the sample size does not induce the selection of the correct model.

For the Inverse Gamma reference model, which is in the model set $\mathcal{M}$ and has the heaviest tail of all models considered, the residual risk is nearly zero – this is expected as the Bayesian worst-case approach typically selects capital calculated according to that distribution. The residual estimation risk is generally *not* guaranteed to be negative, since it is easy to envisage a situation whereby the true model is more heavy tailed than any of the distributions in the model set.

4.2.2 Highest posterior approach (HP)

The second approach we consider is one of choosing the model out of $\mathcal{M}$ that fits the data best. While alternative statistical criteria can be used for this (eg the Akaike or Bayes Information Criteria), in this example, to be consistent with Bayesian arguments, we select the model with the highest *posterior weight*. Details for this approach can be found in Draper (1995), Bernardo and Smith (2000) and Cairns (2000).

This technique requires specification for each model in the set $\mathcal{M}$ of a prior probability $p(F^{(k)})$ and a prior distribution on its parameter vector $\pi(\theta^{(k)}|F^{(k)})$. If there is no prior information on the models, one may set $p(F^{(i)}) = 1/K$ for each model. For each model, the prior on the parameters can be chosen according to the arguments of Section 4.1.

The posterior probability of the model $F^{(k)}$, given data $\mathbf{X} = \mathbf{x}$, is:

$$p(F^{(k)}|\mathbf{x}) = \frac{p(\mathbf{x}|F^{(k)})p(F^{(k)})}{\sum_{i=1}^{K} p(\mathbf{x}|F^{(i)})p(F^{(i)})}, \tag{18}$$

where

$$p(\mathbf{x}|F^{(k)}) = \int_{\Theta^{(k)}} f^{(k)}(\mathbf{x};\theta^{(k)})\pi(\theta^{(k)}|F^{(k)})d\theta^{(k)} \tag{19}$$

is called the *marginal distribution* or *prior predictive distribution* and represents the likelihood of $\mathbf{x}$ given $F^{(k)}$.

Formula (18) can be rewritten as:

$$p(F^{(k)}|\mathbf{x}) = \Big(\sum_{i=1}^{K} \frac{p(F^{(i)})}{p(F^{(k)})} \cdot B_{ik}\Big)^{-1}, \tag{20}$$

where B_{ik} is called *Bayes factor* of $F^{(i)}$ on $F^{(k)}$ and is defined by:

$$B_{ik} = \frac{p(\mathbf{x}|F^{(i)})}{p(\mathbf{x}|F^{(k)})}. \tag{21}$$

Values of B_{ik} greater than 1 suggest that $F^{(i)}$ has a higher chance than $F^{(k)}$ to be the correct model given the data sample. Once the Bayes factor is computed, the model that has the highest posterior weight is chosen.

The corresponding capital estimator thus is:

$$\eta_{\mathcal{M},HP}(\mathbf{X}) = \rho[\hat{F}^{*}(\cdot|\mathbf{X})], \tag{22}$$

where

$$F^{*} \in \mathcal{M} \quad \text{and} \quad p(F^{*}|\mathbf{x}) \geq p(F^{(k)}|\mathbf{x}) \qquad \text{for all } k \in \{1,\ldots,K\}$$

and $\hat{F}^{*}$ is the predictive distribution associated with F^{*}.

Thus, the HP approach gives a transparent model selection criterion, according to which capital may be calculated. Moreover, when the true model that generates the data belongs to $\mathcal{M}$, this approach recognizes it asymptotically, in the sense that the true model's posterior weight tends to 1 as the sample size increases (Bernardo and Smith, 2000). When the true model is not in $\mathcal{M}$, the effectiveness of this approach reduces as posterior probability is concentrated on an incorrect model. In that case, an increase in the sample size may lead to an increase, rather than reduction, in the residual estimation risk.

Table 7: Posterior model weights (PW) used in Highest Posterior and Bayesian Capital Averaging methods, for $n = 10$ and $n = 140$.

	PW Gamma		PW LogNormal		PW InvGamma	
Reference model	$n = 10$	$n = 140$	$n = 10$	$n = 140$	$n = 10$	$n = 140$
Weibull(109.52,454.00)	0.459	0.985	0.303	0.014	0.238	0.001
LogNormal(4.57,0.25)	0.332	0.297	0.334	0.406	0.334	0.297
InvGamma(18.00,1700.00)	0.301	0.088	0.334	0.298	0.356	0.614

In Table 7, the posterior model weights of all three models in $\mathcal{M}$ are reported, averaged across simulated scenarios, for $n = 10$ and $n = 140$. Each row of the table corresponds to a different model generating the data and each cell contains average model weights for both data sizes. It can be seen that, typically, for $n = 10$ the posterior weights remain close to uniform, while for $n = 140$ the differences between the weights placed on different models is distinct.

However, it is noted that convergence of the weights can be slow. For example, in the case of LogNormal data (2nd row) for $n = 140$ the LogNormal model is assigned a higher weight of 0.406, but the Gamma and Inverse Gamma models still have substantial weights of 0.297 each. This may be explained by the shape of the LogNormal distribution which is somewhere between the Gamma and Inverse Gamma and thus is not always unequivocally recognized as such by the posterior weights. On the other hand, in the case of Weibull data (1st row), even a small size of $n = 10$ assigns a probability of 0.459 to a Gamma model, becoming 0.985 for $n = 140$. The Weibull is so different to the LogNormal and Inverse Gamma distributions that all the posterior weight becomes quickly focused on the closest model to the data, the Gamma distribution.

In Table 8, the results for the Highest Posterior approach are presented. Here capital is calculated in each simulated scenario according to the model in $\mathcal{M}$ that has the highest posterior weight. The table shows that the HP approach is effective in reducing the residual estimation risk to values close to zero, when the reference model is included in the model set $\mathcal{M}$. For example, for the LogNormal distribution the residual estimation risk is substantially lower than 1% for all data sizes n. However note that the residual risk does not tend to zero in a straightforward way as n increases – this is explained by the slow convergence of posterior weights apparent in the second row of Table 7.

On the other hand, for the Weibull distribution, the most light-tailed reference model, which does also not belong to $\mathcal{M}$, the HP approach is still too conservative.

Table 8: Values for residual risk (expressed also as a percentage of $\text{VaR}_p(Y)$) for the Highest Posterior method, with model set (16). Reference models are given in Table 5. The data sample is of size n and the risk measure VaR_p is used.

	$n = 10$		$n = 50$		$n = 150$	
Reference model	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$	RR	$\%\text{VaR}_p(Y)$
Weibull(109.52,454.00)	-11.57	-7.55%	-13.24	-8.64%	-13.63	-8.89%
LogNormal(4.57,0.25)	-0.86	0.50%	0.05	0.03%	0.21	0.12%
InvGamma(18.00,1700.00)	1.95	1.10%	2.00	1.13%	1.26	0.71%

4.2.3 Bayesian Capital Averaging

The Bayesian Capital Averaging (BC) approach that we consider in this section also requires the specification of a set of candidate models $\mathcal{M} = \{F^{(1)}, \ldots, F^{(k)}\}$. The estimated capital is then obtained as a weighted average of the estimated capital under each model $\rho[\hat{F}^{(k)}(\cdot|\mathbf{X})]$ where the averaging weights are given by the the posterior probabilities of each model $F^{(k)} \in \mathcal{M}$. The corresponding capital estimator then is:

$$\eta_{\mathcal{M},BC}(\mathbf{X}) = \sum_{k=1}^{K} p(F^{(k)}|\mathbf{X})\rho\left[\hat{F}^{(k)}(\cdot|\mathbf{X})\right]. \tag{23}$$

If the predictive distribution for each model can be computed analytically, calculating $\rho[\hat{F}^{(k)}(\cdot|\mathbf{X})]$ is straightforward, such that (23) can be easily evaluated.When the risk measure used is VaR_p, the method produces capital as a weighted average of percentiles across models. This is effectively the *severity blending* method sometimes used by the users of multiple catastrophe models in insurance (Calder et al., 2012). Related ideas appear in Robert and Therond (2014). We note that the BC method described here is distinct from Bayesian Model Averaging (BMA), as proposed by Hoeting et al. (1999) and Cairns (2000). Under BMA, the quantities being averaged are the predictive distributions for each model, rather than the capital amounts. While this approach can be easily implemented in practice, the numerical inversion of a weighted average of predictive distributions that is necessary in the simulation study made application of BMA in the present paper computationally expensive. However, other simulation experiments we performed indicated that the performance of the BC approach presented here is similar to the more standard BMA approach.

Table 9 displays results for the Bayesian Capital Averaging 1 (BC) approach. Here capital calculated for all three models in $\mathcal{M}$ is averaged according to posterior weights. The overall picture is fairly similar to that of Table 8, reporting the results of the HP approach. It is notable that the BC approach does not generally lead to a lower residual estimation risk than the HP approach, such that, at least in the present example, the simpler HP approach appears more effective. Given that with increased sample sizes posterior model weights concentrate on one distribution, this is not surprising.

Table 9: Values for residual risk (expressed also as a percentage of $\mathrm{VaR}_p(Y)$) for the Bayesian Capital Averaging method, with model set (16). Reference models are given in Table 5. The data sample is of size n and the risk measure VaR_p is used.

	$n = 10$		$n = 50$		$n = 150$	
Reference model	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$	RR	$\%\mathrm{VaR}_p(Y)$
Weibull(109.52,454.00)	-11.79	-7.69%	-13.89	-9.06%	-13.64	-8.90%
LogNormal(4.57,0.25)	0.07	0.04%	0.16	0.09%	0.26	0.02%
InvGamma(18.00,1700.00)	3.59	2.03%	3.36	1.90%	2.10	1.19%

4.2.4 Computational issues in obtaining model posteriors

The use of model posteriors in methods HP and BC requires calculation of the posterior weight $p(F^{(k)}|\mathbf{X})$ for each model in $\mathcal{M}$. These can be difficult to compute. Indeed, the marginal distributions $p(\mathbf{x}|F^{(k)})$ are often not available in a closed form and require numerical calculation. Moreover, if $p(\mathbf{x}|F^{(k)})$ is calculated using an improper prior (as is the case in this paper), it will be defined only up to a constant. This constant will also appear in the Bayes factor B_{ik} as c_i/c_j.

Several solutions have been proposed to overcome this issue, including the *Intrinsic Bayes factor* method proposed by Berger and Pericchi (1996). This method consists in using part of the data to estimate the constant c_i and thus make the prior proper. The rest of the data are used to compute the Bayes factor according to this new proper prior. While selecting the correct set of training data is generally computationally demanding, it has been proved (Berger and Pericchi, 1996) that for location-scale families, or scale families, the constant c_i/c_j is always equal to 1, which greatly simplifies calculations. The models considered in this section were all scale models.

5 Conclusions

Using the concepts of residual estimation risk and estimated capital risk, we propose a way of measuring model error under different reference models. This is demonstrated by assessing the performance of non-parametric estimation methods, based on historical simulation and worst-case scenarios under moment constraints. It was shown how these approaches lead to a substantial model risk, with historical simulation understating and the worst-case approach overstating the estimated capital.

For capital estimation procedures that make use of a set of candidate statistical models, it was shown how model uncertainty, as distinct from parameter uncertainty, may be assessed. Through simulation studies it was demonstrated how capital setting based on Bayesian predictive distributions can (a) eliminate the impact of parameter uncertainty and (b) greatly moderate the impact of model uncertainty, as measured by residual estimation risk. For the measure of model uncertainty employed here and the distributions considered, it appears that *capital averaging* across models does not confer an obvious advantage in relation to *model selection*.

The choice of the set of models used in estimation is of course crucial. If the true model is included in that set or can be well approximated by distributions in the set, the Bayesian capital

estimation procedures can be very effective in reducing residual estimation risk. However, if the true model is not included in the model set used in estimation and has properties dissimilar from all elements in that set, the effectiveness of such capital estimation procedures is reduced.

Acknowledgment

Valeria Bignozzi gratefully acknowledges the financial support of the Institute and Faculty of Actuaries, under the research grant *The impact of model error on capital adequacy.* The authors thank an anonymous referee for helpful comments.

Appendix

Here we report the calculations for the predictive density and cumulative distribution for each model $F^{(k)}$ in the model set $\mathcal{M}$ used in the simulation studies as well as the respective capital $\mathrm{VaR}_p\left[\hat{F}^{(k)}(\cdot|\mathbf{X})\right]$, consistently with Section 4.1.

Gamma distribution

Let $Y, X_1, \dots, X_n \sim \Gamma(\alpha, \theta)$ with fixed shape parameter α and unknown scale parameter θ. The probability density function is

$$f(x;\theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}};$$

The MLE for the parameter θ is given by

$$\hat{\theta} = \frac{\sum_i x_i}{n\alpha}.$$

Given the non-informative prior $\pi(\theta) = \frac{1}{\theta}$, the parameter posterior is obtained as

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \frac{\prod_{i=1}^n x_i^{\alpha-1}}{\Gamma(\alpha)^n \theta^{n\alpha+1}} e^{-\frac{\sum_{i=1}^n x_i}{\theta}} \\ &= \frac{\prod_{i=1}^n x_i^{\alpha-1}}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha)}{(\sum_{i=1}^n x_i)^{n\alpha}} \cdot \left[\frac{(\sum_{i=1}^n x_i)^{n\alpha}}{\Gamma(n\alpha)} \theta^{-n\alpha-1} e^{-\frac{\sum_{i=1}^n x_i}{\theta}}\right].\end{aligned}$$

The quantity in the squared bracket is the density function of an Inverse Gamma distribution with parameters $n\alpha$ and $(\sum_{i=1}^n x_i)$. It follows that for it to integrate to 1, the marginal distribution $m(\mathbf{x})$ is the normalizing constant:

$$m(\mathbf{x}) = \frac{\prod_{i=1}^n x_i^{\alpha-1}}{\Gamma(\alpha)^n} \frac{\Gamma(n\alpha)}{(\sum_{i=1}^n x_i)^{n\alpha}}.$$

The predictive density function is then

$$\begin{aligned}\hat{f}(y|\mathbf{x}) &= \int_0^\infty f(y;\theta)\cdot\pi(\theta|\mathbf{x})d\theta \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\theta^\alpha} y^{\alpha-1} e^{-\frac{y}{\theta}} \frac{(\sum_{i=1}^n x_i)^{n\alpha}}{\Gamma(n\alpha)} \theta^{-n\alpha-1} e^{-\frac{\sum_{i=1}^n x_i}{\theta}} d\theta \\ &= \frac{y^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^\infty \theta^{-\alpha(n+1)-1} e^{-\frac{y+\sum_{i=1}^n x_i}{\theta}} d\theta \\ &= \frac{y^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha}\Gamma(\alpha(n+1))(y+\sum_{i=1}^n x_i)^{\alpha(n+1)}}{\Gamma(\alpha)\Gamma(n\alpha)(y+\sum_{i=1}^n x_i)^{\alpha(n+1)}\Gamma(\alpha(n+1))} \int_0^\infty \theta^{-\alpha(n+1)-1} e^{-\frac{y+\sum_{i=1}^n x_i}{\theta}} d\theta\end{aligned}$$

$$\implies \hat{f}(y|\mathbf{x}) = \frac{\Gamma(\alpha(n+1))(\sum_{i=1}^n x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \frac{y^{\alpha-1}}{(y+\sum_{i=1}^n x_i)^{\alpha(n+1)}}.$$

and the predictive distribution is:

$$\hat{F}(y|\mathbf{x}) = \int_0^y \hat{f}(t|\mathbf{x})dt = \frac{\Gamma(\alpha(n+1))(\sum_{i=1}^n x_i)^{n\alpha}}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^y \frac{t^{\alpha-1}}{(t+\sum_{i=1}^n x_i)^{\alpha(n+1)}}dt.$$

Considering the change of variable

$$z = \frac{t}{(t+\sum_{i=1}^n x_i)}$$

with

$$dz = \frac{\sum_{i=1}^n x_i}{(t+\sum_{i=1}^n x_i)^2}dt,$$

the above integral becomes

$$\begin{aligned}
\hat{F}(y|\mathbf{x}) &= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha}}{(t+\sum_{i=1}^n x_i)^{\alpha(n+1)}} \frac{(t+\sum_{i=1}^n x_i)^2}{\sum_{i=1}^n x_i} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}(\sum_{i=1}^n x_i)^{n\alpha-1}}{(t+\sum_{i=1}^n x_i)^{\alpha(n+1)-2}} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} \frac{t^{\alpha-1}}{(t+\sum_{i=1}^n x_i)^{\alpha-1}} \cdot \frac{(\sum_{i=1}^n x_i)^{n\alpha-1}}{(t+\sum_{i=1}^n x_i)^{n\alpha-1}} dz \\
&= \frac{\Gamma(\alpha(n+1))}{\Gamma(\alpha)\Gamma(n\alpha)} \int_0^{\frac{y}{(y+\sum_{i=1}^n x_i)}} z^{\alpha-1} \cdot z^{n\alpha-1} dz = \mathrm{I}\left(\frac{y}{(y+\sum_{i=1}^n x_i)}; \alpha, n\alpha\right),
\end{aligned}$$

where $I(\cdot;\alpha,n\alpha)$ is the cumulative distribution function of a Beta random variable. To compute $\mathrm{VaR}_p[\hat{F}(y|\mathbf{x},G)]$, we need to invert that function. Hence we have

$$\frac{y}{(y+\sum_{i=1}^n x_i)} = \mathrm{I}^{-1}(p,\alpha,n\alpha)$$

and

$$\mathrm{VaR}_p[\hat{F}_{GM}(|\mathbf{x})] = \sum_{i=1}^n x_i \frac{\mathrm{I}^{-1}(p,\alpha,n\alpha)}{1-\mathrm{I}^{-1}(p,\alpha,n\alpha)}.$$

LogNormal distribution

Consider a LogNormal distribution with fixed shape parameter σ and unknown scale parameter $\gamma = e^{\mu}$. Its probability density function is

$$f(x,\gamma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \frac{(\log(x)-\log(\gamma))^2}{2\sigma^2}.$$

Given a sample $(x_1,\ldots,x_n)$, the MLE for the parameter γ is

$$\hat{\gamma} = \exp\left(\frac{\sum_i \ln(x_i)}{n}\right).$$

The predictive cumulative function can be easily computed from the predictive cumulative function of a Normal distribution with known parameter σ. It follows that the LogNormal predictive cumulative function is again LogNormal with scale parameter $e^{\frac{\sum_{i=1}^n \ln(x_i)}{n}}$ and shape parameter $\sqrt{1+\frac{1}{n}}\sigma$. The capital $\mathrm{VaR}_p[\hat{F}_{LN}(\cdot|\mathbf{X})]$ is obtained by inverting that function.

Inverse Gamma distribution

If $Z \sim GM(\alpha, 1/\beta)$, then $X = \frac{1}{Z}$ has an inverse gamma distribution with parameters (α, β). Its density function is given by:

$$f(x;\beta) = \frac{\beta^\alpha \exp\left(-\frac{\beta}{x}\right)}{\Gamma(\alpha) x^{\alpha+1}},$$

where β is a scale parameter and α is a shape parameter.

Assuming α is known, the MLE for of β is obtained by

$$\hat{\beta} = \alpha \left(\frac{1}{n} \sum_i \frac{1}{x_i} \right)^{-1}.$$

The predictive cumulative function for the Inverse Gamma, for prior $\pi(\beta) = 1/\beta$ is obtained analogously to that of Gamma distribution, so detailed calculations are omitted. The corresponding capital is obtained as

$$\mathrm{VaR}_p[\hat{F}_{IG}(\cdot|\mathbf{x})] = \frac{1 - \mathrm{I}^{-1}(1-p, \alpha, n\alpha)}{\mathrm{I}^{-1}(1-p, \alpha, n\alpha) \sum_{i=1}^n (1/x_i)}. \tag{24}$$

References

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance 26*(1), 1505–1518.

Alexander, C. and J. M. Sarabia (2012). Quantile uncertainty and value-at-risk model risk. *Risk Analysis 32*(8), 1293–1308.

Arnold, B. C., N. Balakrishnan, and H. N. Nagaraja (1992). *A first course in order statistics*, Volume 54. SIAM, Philadelphia, PA.

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

Barrieu, P. and G. Scandolo (2015). Assessing financial model risk. *European Journal of Operational Research 242*(2), 546–556.

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

Berger, J. and L. R. Pericchi (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association 91*(433), 109–122.

Bernardo, J. and A. Smith (2000). *Bayesian Theory.* Wiley.

Bignozzi, V. and A. Tsanakas (2015). Residual estimation risk. Forthcoming in *Journal of Risk and Insurance.*

Boucher, C. M., J. Danielsson, P. S. Kouontchou, and B. B. Maillet (2014). Risk models-at-risk. *Journal of Banking and Finance 44*, 72–92.

Cairns, A. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics 27*(3), 313–330.

Calder, A., A. Couper, and J. Lo (2012). Catastrophe model blending techniques and governance. `www.actuaries.org.uk/sites/all/files/documents/pdf/catastrophe-model-blending.pdf`.

Colander, D., M. Goldberg, A. Haas, A. Juselius, A. Kirman, T. Lux, and B. Sloth (2009). The financial crisis and the systemic failure of academic economics. *Critical Review 21*(2-3), 249–267.

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

Danielsson, J. and C. G. de Vries (1997). *Extreme returns, tail estimation, and value-at-risk.* Institute of Economic Studies, University of Iceland.

Danielsson, J., P. Embrechts, C. Goodhart, C. Keating, F. Muennich, O. Renault, H. S. Shin, et al. (2001). An academic response to basel II.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society, Series B 57*(1), 45–97.

Embrechts, P., G. Puccetti, and L. Rüschendorf (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance 37*(8), 2750–2764.

European Commission (2010). *QIS 5 Technical Specifications, Annex to Call for Advice from CEIOPS on QIS5.* Brussels.

Föllmer, H. and A. Schied (2011). *Stochastic Finance: An Introduction in Discrete Time* (Third ed.). de Gruyter.

Frankland, R., S. Eshun, L. Hewitt, P. Jakhria, S. Jarvis, A. Rowe, A. Smith, A. Sharp, J. Sharpe, and T. Wilkins (2014). Difficult risks and capital models. *British Actuarial Journal 19*(3), 556–616.

Gerrard, R. and A. Tsanakas (2011). Failure probability under parameter uncertainty. *Risk Analysis 31*(5), 727–744.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–401.

Klugman, S. A., H. H. Panjer, and G. E. Willmot (2008). *Loss Models: From Data to Decisions* (Third ed.). Wiley.

Krätschmer, V., A. Schied, and H. Zähle (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics 18*, 271–295.

Lawson, T. (2009). The current economic crisis: its nature and the course of academic economics. *Cambridge Journal of Economics 33*(4), 759–777.

McNeil, A., R. Frey, and P. Embrechts. (2005). *Quantitative risk management: Concepts, techniques and tools.* Princeton University Press.

Robert, C. Y. and P.-E. Therond (2014). Distortion risk measures, ambiguity aversion and optimal effort. *Astin Bulletin 44*(02), 277–302.

Wang, S. S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin 26*(1), 71–92.