Parameter uncertainty and residual estimation risk*

Valeria Bignozzi† Andreas Tsanakas‡

September 29, 2014

Abstract

The notion of residual estimation risk is introduced in order to study the impact of parameter uncertainty on capital adequacy, for a given risk measure and capital estimation procedure. Residual estimation risk is derived by applying the risk measure on a portfolio consisting of a random loss and a capital estimator, reflecting the randomness inherent in the data. Residual risk thus equals the additional amount of capital that needs to be added to the portfolio to make it acceptable. We propose modified capital estimation procedures, based on parametric bootstrapping and on predictive distributions, which tend to increase capital requirements, by compensating for parameter uncertainty and leading to a residual risk close to zero. In the particular case of location-scale families of distributions, the analysis simplifies substantially and a capital estimator can always be found that leads to a residual risk of exactly zero.

Keywords: Parameter uncertainty, model uncertainty, bootstrap, predictive distribution, location-scale families, risk measures, solvency.

*This is a preprint of an article accepted for publication in the *Journal of Risk and Insurance*, ©2004 the American Risk and Insurance Association.

†School of Economics and Management, University of Firenze, Italy; email: valeria.bignozzi@unifi.it

‡Cass Business School, City University London; email: a.tsanakas.1@city.ac.uk

1 Introduction

Insurance decisions, such as pricing, reserving and capital setting, are informed by the outputs of statistical risk models. Such models are typically parametric. However, the true values of parameters are in principle unknown and must be estimated from samples of relevant observations. Such samples are often very small and statistical error means that the estimated parameter values can diverge substantially from the true values. The potential for error in estimated parameters, termed *parameter uncertainty*, introduces possible error into insurance decisions based on model outputs. For example, if the estimated frequency used to model claims from a natural hazard is lower than the true one, insurance policies may be under-priced and a portfolio of such policies under-capitalized. It is conventional to view parameter uncertainty in the context of an otherwise correctly specified risk model. If the model is not known with certainty, we talk of *model uncertainty*.

Decisions sensitive to tails of distributions, for which limited information resides in available data, are more sensitive to parameter error. Thus, there is particular focus on applications where the extremes of loss distributions are of interest, for example when setting capital by a tail risk measure like Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR), or when pricing high reinsurance layers. Investigations by insurance practitioners have shown that the impact of parameter uncertainty in realistic modeling applications can indeed be very substantial, see Mata (2000) and Borowicz and Norman (2008). High sensitivity to parameter error has also been demonstrated in the context of credit risk modeling by McNeil et al. (2005) and in a banking context by Jorion (1996), who argues that confidence bands should be reported alongside estimated VaRs. Cont et al. (2010) study risk measurement procedures and their sensitivity to changes in data used for estimation, with reference to the notion of qualitative robustness.

In an early response to parameter uncertainty, Venezian (1983) recommends an explicit increase to insurer capital, with the adjustment reflecting estimation volatility. Allowance for parameter uncertainty is now often made in actuarial risk models, see Cairns et al.

(2006) on stochastic mortality and longevity bond pricing and Verrall and England (2006) on stochastic claims reserving. A broadly applicable response to parameter uncertainty is to work with predictive distributions, arising as weighted averages of distributions with different parameter values. Cairns (2000) argues in favor of a fully Bayesian approach to capture parameter (as well as model) uncertainty. Predictive distributions tend to be more volatile than distributions that are derived via point estimates of parameters and thus can lead to more conservative decisions, see also Landsman and Tsanakas (2012). Gerrard and Tsanakas (2011) show that the increase in VaR that using a predictive distribution implies, is appropriate for restoring a frequentist failure probability to its required nominal level.

In the present contribution we also study risk measurement procedures, but from a perspective that is complementary to that of Cont et al. (2010). We introduce a criterion for assessing, in monetary units, the impact of parameter uncertainty on capital adequacy. We assume that the required capital for an insurance company is calculated by applying to a random loss variable a risk measure that is positive homogeneous, translation invariant, and law invariant. Commonly used risk measures, such as VaR or TVaR, satisfy these properties. A random sample for the loss is available and the estimated capital is a function of that sample; we call this function the *capital estimator*. To assess the effectiveness of a capital setting procedure, the risk measure is applied to the difference between the loss (a random variable representing process variability) minus the capital estimator (a random variable reflecting variability due to estimation). The result of this calculation we call *residual estimation risk*. If the distribution of the loss is known, the residual risk is zero. In general, residual estimation risk is not equal to zero and can be viewed as the (deterministic) amount of capital that needs to be added to the capital estimator such that the total position becomes acceptable. When the sample size grows, the capital estimator typically converges to the unknown required capital and the residual estimation risk thus goes to zero.

This approach can in principle be used to assess the effectiveness of any estimation

procedure, even though our focus is on parametric models. In particular, model uncertainty can also be assessed via residual estimation risk, if model selection is data driven and thus reflected in the capital estimator (Bignozzi and Tsanakas, 2013). A strength of the proposed approach is that it allows to consistently rank different estimation procedures. We show that monotonicity and subadditivity of the risk measure ensure that capital estimators are appropriately ranked with respect to stochastic dominance and convex stochastic order. Moreover, residual risk is expressed in monetary units and derived with reference to a risk measure that explicitly represents preferences.

Apart from quantifying residual estimation risk, in the context of parametric models, we discuss modifications to the capital setting procedure that reduce or even eliminate residual risk. The first approach we discuss is a parametric bootstrap correction to the capital estimator. Since the exact value of residual risk generally depends on unknown parameters, we rely on a simple plug-in estimator of residual risk. Under mild conditions it is shown that this corrected capital estimator leads to an improvement in residual risk and its effectiveness is demonstrated by numerical examples, including the common log-normal and Pareto models. In the second approach we propose, capital is set using a predictive distribution derived by standard Bayesian arguments. The good predictive performance of Bayesian methods under frequentist criteria has been highlighted in the literature (Smith, 1999; Datta et al., 2000; Gerrard and Tsanakas, 2011). Once more, numerical examples show that use of the predictive distribution nearly eliminates residual estimation risk.

In the special case of location-scale models, such as the normal and Student t distributions, several simplifications appear. It is shown that residual risk is independent of the location parameter and proportional to the scale parameter. It follows that a simple modification of the risk measure used to set capital can lead to a residual risk of exactly zero. For example, if a TVaR measure at given confidence level is used to set capital, this may be adjusted by using a different confidence level, depending on the size of the data sample, but not on unknown parameters. Furthermore, for location-scale families,

repeated bootstrap corrections to the capital estimator can be performed without increasing computational cost, as no nested simulations are needed. Finally, it is shown that for location families and for scale families, the use of a predictive distribution respectively eliminates residual estimation risk and a closely related functional.

Throughout the paper, we deal with models such as the Pareto and the log-normal, which can give rise to distributions with infinite means, either by the values taken by parameter estimators or by the mixture involved in deriving predictive distributions. Infinite-mean models do not allow the evaluation of coherent risk measures such as TVaR (Nešlehová et al., 2006). For this reason, we often use in examples a truncated version of TVaR, which we call Range-Value-at-Risk (RVaR). This risk measure, already considered by Cont et al. (2010), considers a larger part of the tail than VaR, but is not coherent and is well defined even when the mean of the loss distribution is infinite.

In Section 2 we introduce the notion of residual estimation risk and present numerical examples of its quantification. Section 3 discusses methods for controlling residual risk applicable to general loss distributions, while Section 4 deals with the specific case of location-scale families. Brief conclusions are given in Section 5. Details of calculations are documented in the Appendix of Section 6.

2 Residual estimation risk

2.1 Risk measures

The financial loss of a portfolio is modeled by a random variable Y, defined on a standard non-atomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Thus, in the event $\{Y > 0\}$ a portfolio loss occurs, while $\{Y \leq 0\}$ corresponds to a gain. The distribution of Y is F. When considering parametric models, we write $F \equiv F(\cdot; \theta)$ where $\theta \in \Theta$ is a vector of parameters. All inequalities between random variables are meant to hold $\mathbb{P}$-a.s.

A risk measure is a functional ρ that assigns to every financial loss Y a real number

$\rho(Y)$. (Whenever this notation is used, we implicitly assume that the distribution of Y is such that $\rho(Y)$ is well defined.) $\rho(Y)$ is expressed in monetary units and may represent a regulatory capital requirement, which is the interpretation we follow here. Following Artzner et al. (1999), a loss is *acceptable* if $\rho(Y) \leq 0$ and *not acceptable* if $\rho(Y) > 0$.

We consider in this paper only risk measures that satisfy the following four standard properties, which are assumed throughout the paper and not explicitly stated further on.

Monotonicity: If $Y_1 \leq Y_2$, $\rho(Y_1) \leq \rho(Y_2)$.

Translation invariance: If $a \in \mathbb{R}$, $\rho(Y + a) = \rho(Y) + a$;

Positive homogeneity: If $\lambda \geq 0$, $\rho(\lambda Y) = \lambda \rho(Y)$;

Law invariance: If $Y_1 \overset{d}{=} Y_2$, $\rho(Y_1) = \rho(Y_2)$,

where $\overset{d}{=}$ denotes equality in distribution.

Law invariance requires that two losses with the same distribution have the same capital requirement. Because of this, a risk measure can also be evaluated as a functional of a distribution, such that for $Y \sim F$ we may denote $\rho(Y) \equiv \rho[F]$. With this notation, translation invariance and positive homogeneity can be written as $\rho[F(\cdot - a)] = \rho[F] + a$ and $\rho[F(\frac{\cdot}{\lambda})] = \lambda \rho[F]$ respectively.

Three risk measures satisfying the above properties are

$$\mathrm{VaR}_p(Y) := \inf\{a \in \mathbb{R} \mid \mathbb{P}(Y \leq a) \geq p\} = F^{-1}(p), \tag{1}$$

$$\mathrm{TVaR}_p(Y) := \frac{1}{1-p} \int_p^1 \mathrm{VaR}_u(Y) du = \mathbb{E}(Y | Y > F^{-1}(p)), \tag{2}$$

$$\mathrm{RVaR}_{p_1,p_2}(Y) := \frac{1}{p_2 - p_1} \int_{p_1}^{p_2} \mathrm{VaR}_u(Y) du = \mathbb{E}(Y | F^{-1}(p_1) < Y < F^{-1}(p_2)). \tag{3}$$

The second equality in each of (1), (2), and (3), holds under the additional assumption that the distribution F is invertible. All three risk measures are special cases of the general class of *distortion* risk measures (Wang et al., 1997). The VaR_p measure, used extensively in insurance and banking regulation, is the $100p^{th}$ percentile of the loss distribution. VaR_p is characterized by its insensitivity to the extreme tails of loss distributions, see Dowd

and Blake (2006). TVaR_p corrects for this defect by considering the average of all VaRs above the $100p^{th}$ percentile. However, this introduces sensitivity to extreme percentiles, which may not be reliably estimable from limited data. Furthermore, TVaR is not defined for distributions with infinite means. RVaR_{p_1,p_2}, proposed by Cont et al. (2010), offers a compromise between those two risk measures: while it considers most of the tail, it does not reflect some very extreme losses that a TVaR_{p_1} measure would consider. Thus, for $p_2 < 1$ is always well defined.

An additional property often required is:

Subadditivity: For every Y_1, Y_2, $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$,

Risk measures satisfying monotonicity, translation invariance, positive homogeneity and subadditivity are termed *coherent*, see Artzner et al. (1999). Of the three risk measures discussed above, only TVaR is subadditive and thus coherent.

2.2 Residual estimation risk

For the loss Y, the value of a law invariant risk measure depends on the distribution function F, which is typically unknown and needs to be estimated from data. An i.i.d. random sample of size n from F will be denoted by $\mathbf{X} = \{X_1, \ldots, X_n\}$; with slight abuse of notation we write $\mathbf{X} \sim F$. We assume that Y is independent of $\mathbf{X}$. It follows that the capital that the holder of Y needs to hold, with reference to the risk measure ρ, will also depend on the random sample itself and is denoted by $\eta(\mathbf{X})$. Similar to Cont et al. (2010), risk measurement is viewed as a two-step procedure. First the distribution F needs to be estimated from data; denote the estimator of the distribution by $F_{\mathbf{X}}$. Second, the risk measure ρ is evaluated at $F_{\mathbf{X}}$, yielding the *capital estimator* $\eta(\mathbf{X}) = \rho[F_{\mathbf{X}}]$.

Cont et al. (2010) investigated the robustness of risk measurement procedures with respect to small changes in the data set. They consider the notion of *qualitative robustness*, which is closely related to weak continuity of statistical functionals. In this paper we offer a complementary discussion of risk measurement procedures, considering the impact of

estimation volatility on capital adequacy, rather than robustness.

From translation invariance we have

$$\rho(Y - \rho(Y)) = 0, \tag{4}$$

such that, by monotonicity, $\rho(Y)$ is the minimal amount of capital that needs to be subtracted from the loss Y to make it acceptable. Reflecting the variability in the random sample $\mathbf{X}$, we can consider $Y - \eta(\mathbf{X})$ as the random variable that represents the loss, after the (random) capital estimator has been subtracted from it. We then define as *residual estimation risk* the quantity

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

Equation (5) is analogous to (4), with the theoretical capital value $\rho(Y)$ substituted by the capital estimator $\eta(\mathbf{X})$. A positive residual risk implies that the impact of model uncertainty is such that subtracting the capital estimator $\eta(\mathbf{X})$ from the loss Y does not produce an acceptable loss. Hence more safely invested capital needs to be held. The residual estimation risk thus reflects the extra amount of capital that needs to be added to the estimated capital $\eta(\mathbf{X})$ in order to make Y acceptable, in particular it holds that $\rho(Y - (\eta(\mathbf{X}) + \mathrm{RR}(F, \eta))) = 0$. By its definition, residual risk depends on the risk measure ρ; this dependence is suppressed in the notation.

Of course, alternative measures of model and parameter uncertainty are available. For example, confidence intervals for the estimated risk measure $\rho[F_{\mathbf{X}}]$ can be calculated. However, such an approach does not provide clear guidance about how risk estimates should be plausibly adjusted, e.g. at which confidence level of $\rho[F_{\mathbf{X}}]$ capital should be set.[1] By measuring model risk in monetary units and associating it with the risk measure itself, stated risk preferences provide guidance on designing capital setting procedures, as will be seen in Section 3. Furthermore, the presence of the future loss Y and a related estimator in the same expression distinguish the present approach from uncertainty assessments based

[1] The issue of how risk measures should be calibrated, e.g. what the confidence level of VaR should be remains open – but this is a different problem. Residual risk allows a measurement of model uncertainty that is consistent with such externally given risk preferences.

on e.g. confidence intervals, by placing it in the broader context of *predictive inference* (see e.g. Barndorff-Nielsen and Cox, 1996) and associating it with current considerations of backtesting and model validation, see McNeil et al. (2005, Section 4.4.3) and Ziegel (2014).

Before elaborating on that last point, some elementary properties of residual estimation risk are collected. First, recall the definitions of stochastic dominance and convex order for random variables (see Denuit et al., 2005, Sections 3.3-3.4). The random variable S is smaller than T in stochastic dominance, $S \preceq_{st} T$, if $\mathrm{VaR}_p(S) \leq \mathrm{VaR}_p(T)$ for all $p \in [0,1]$. Intuitively $S \preceq_{st} T$ implies that S is a smaller risk than T; in particular one can always find $S' \stackrel{d}{=} S$, $T' \stackrel{d}{=} T$ such that $\mathbb{P}(S' \leq T') = 1$. The variability of S and T can be compared via the convex order. S is smaller than T in convex order, $S \preceq_{cx} T$, if $\mathbb{E}(v(S)) \leq \mathbb{E}(v(T))$, for all convex functions v such that the expectations exist. Immediate consequences of $S \preceq_{cx} T$ are $\mathbb{E}(S) = \mathbb{E}(T)$ and $\mathrm{Var}(S) \leq \mathrm{Var}(T)$.

Proposition 2.1. *For a risk measure ρ and distribution F, the following hold:*

a) For any distribution F^*, it is $\mathrm{RR}(F, \rho[F^*]) = \rho[F] - \rho[F^*]$.

b) If $\eta_1(\mathbf{X}) \preceq_{st} \eta_2(\mathbf{X})$, then it is $\mathrm{RR}(F, \eta_1) \geq \mathrm{RR}(F, \eta_2)$.

c) If $\eta(\mathbf{X}) \geq (\leq) \rho[F]$, then it is $\mathrm{RR}(F, \eta) \leq (\geq) 0$.

d) If ρ is subadditive and satisfies the Fatou property, then $\eta_1(\mathbf{X}) \preceq_{cx} \eta_2(\mathbf{X})$ implies $\mathrm{RR}(F, \eta_1) \leq \mathrm{RR}(F, \eta_2)$

e) If ρ is subadditive and $\eta(\mathbf{X}) = \lambda\eta_1(\mathbf{X}) + (1-\lambda)\eta_2(\mathbf{X})$ holds for all $\lambda \in [0,1]$, then it is $\mathrm{RR}(F, \eta) \leq \lambda \mathrm{RR}(F, \eta_1) + (1-\lambda)\mathrm{RR}(F, \eta_2)$.

Proof. Part a) follows from translation invariance of ρ. For part b), by Proposition 3.3.17 in Denuit et al. (2005) it is $Y - \eta_2(\mathbf{X}) \preceq_{st} Y - \eta_1(\mathbf{X})$. The stated inequality then follows from monotonicity of ρ and Theorem 4.1 of Bäuerle and Müller (2006). Part c) is similar to b). For part d), by noting that $\eta_1(\mathbf{X}) \preceq_{cx} \eta_2(\mathbf{X}) \Leftrightarrow -\eta_1(\mathbf{X}) \preceq_{cx} -\eta_2(\mathbf{X})$ and applying

Proposition 3.4.25 in Denuit et al. (2005) it is $Y - \eta_1(\mathbf{X}) \preceq_{cx} Y - \eta_2(\mathbf{X})$. The result then follows from subadditivity and positive homogeneity of ρ (implying convexity) and Theorem 4.2 of Bäuerle and Müller (2006). Part e) follows directly from subadditivity and positive homogeneity of ρ. □

Part a) of Proposition 2.2 implies that estimating capital using a fixed distribution F^* (essentially ignoring the data) leads to residual risk equal precisely to the difference between the true and estimated level of capital, which is a simple measure of model error. Part b) shows that choosing a larger capital estimator, in the sense of stochastic dominance, will lead to a reduction in residual risk. Part c) is a special case of b): designing a capital estimator that is always larger than the required capital under the true distribution guarantees negative residual risk. Finally, part d) shows that residual risk penalizes volatile capital estimators, for risk measures that are subadditive and satisfy the Fatou property (a continuity requirement satisfied by all risk measures considered here – see Bäuerle and Müller (2006) for details). Finally, the convexity of RR in η (part e) is desirable as it implies that averaging of two capital estimators will lead to a residual risk that is an improvement on the worst performing of the two.

The properties of residual estimation risk induce quality rankings of capital estimators. This relates, but is distinct, to the discussion of *elicitability* of risk measures, which concerns the potential for assessing the quality of individual risk forecasts via a particular scoring rule – see Gneiting (2011) and Ziegel (2014) for more detail. A full discussion of elicitability is beyond the scope of this paper, but we note two key differences. First, in our context the future loss is compared to an estimator (a random variable) rather than a fixed forecast, allowing the assessment of the quality of risk measurement procedures rather than individual forecasts. Second, the scoring approach penalizes all deviations between forecasts and realized losses, while residual risk is allowed to be negative, thus distinguishing between scenarios of potential under- and over-capitalization.

Links between residual risk and backtesting become apparent by considering the par-

ticular case of $\rho \equiv \mathrm{VaR}_p,\ \eta(\mathbf{X}) = \mathrm{VaR}_p[F_{\mathbf{X}}]$. Then, the equivalence holds,

$$\mathrm{RR}(F, \eta) = \mathrm{VaR}_p(Y - \eta(\mathbf{X})) \geq 0 \Leftrightarrow \mathbb{P}(Y > \eta(\mathbf{X})) \geq 1 - p. \tag{6}$$

The right-hand-side of inequality (6) signifies a probability of failure (future loss exceeding the capital estimator) higher than the acceptable level $1-p$ and was used as a measure of parameter uncertainty by Gerrard and Tsanakas (2011). The quantity $\mathbb{P}(Y > \eta(\mathbf{X}))$ can be interpreted as the expected relative frequency of violations when backtesting a VaR model. Hence, the definition of residual estimation risk (5) can be related to a backtesting criterion for risk measures more general than VaR; see the discussion on backtesting VaR and TVaR in McNeil et al. (2005), Section 4.4.3.

2.3 Residual risk and parameter estimation

In the rest of the paper (with the exception of Example 11), we focus on residual risk due to parameter rather than model uncertainty. Hence from now on we set $F \equiv F(\cdot;\theta)$, where the distribution family $F(\cdot;\cdot)$ is known, but the parameter θ needs to be estimated. Consequently, we simplify notation somewhat and write from now on $\mathrm{RR}(\theta, \eta) \equiv \mathrm{RR}(F(\cdot;\theta), \eta)$. The residual risk depends on unknown but true parameters. This is similar to other standard frequentist quality criteria such as e.g. the mean-squared-error.

By $\hat{\theta}$ we denote an estimator of θ based on the random sample $\mathbf{X}$. Unless otherwise specified, parameters are estimated by likelihood maximization, such that $\eta(\mathbf{X}) = \rho[F(\cdot;\hat{\theta})]$ is the MLE of $\rho[F(\cdot;\theta)]$.

We now present examples to illustrate the impact of parameter uncertainty on different risk measures and distributions. In all numerical examples in this paper, the residual risk $\rho(Y - \eta(\mathbf{X}))$, for any distribution and capital estimator, is calculated numerically via Monte-Carlo simulation. The precise details of the simulation algorithm are given in Section 6.1. To allow comparisons, we will report in examples a normalized version of residual risk:

$$\mathrm{NRR}(F, \eta) = \frac{\mathrm{RR}(F, \eta)}{\rho(Y) - \mathbb{E}(Y)}, \tag{7}$$

In all examples the denominator $\rho(Y) - \mathbb{E}(Y)$ is analytically calculated.

Example 1 (Normal, MLE)**.** First consider a simple normal model, $Y, \mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$. The mean μ and the standard deviation σ are unknown. Hence $F(\cdot; (\mu, \sigma)) \equiv \Phi\left(\frac{\cdot - \mu}{\sigma}\right)$, where Φ is the standard normal distribution. The standard normal density is denoted by ϕ. We can write $Y \stackrel{d}{=} \mu + \sigma Z$, where $Z \sim \mathcal{N}(0, 1)$. The MLE $(\hat{\mu}, \hat{\sigma}^2)$ satisfies

$$(\hat{\mu},\ \hat{\sigma}^2) \stackrel{d}{=} \left(\mu + \frac{\sigma}{\sqrt{n}} U,\ \frac{\sigma^2 V}{n}\right),$$

where $U \sim \mathcal{N}(0, 1)$ and $V \sim \chi^2_{n-1}$, with U, V independent.

By the translation invariance and positive homogeneity of ρ it is $\rho(Y) = \mu + \sigma c$, where $c = \rho(Z)$. The capital estimator becomes

$$\eta(\mathbf{X}) = \rho[F(\cdot; (\hat{\mu}, \hat{\sigma}))] = \mu + \frac{\sigma}{\sqrt{n}} U + \sigma \sqrt{\frac{V}{n}} c.$$

Consequently, the residual risk can be calculated as

$$\mathrm{RR}((\mu, \sigma), \eta) = \sigma \rho \left(Z - \frac{1}{\sqrt{n}} U - \sqrt{\frac{V}{n}} c\right),$$

where Z is independent of U, V. The residual risk thus does not depend on the mean and is proportional to the standard deviation. Furthermore, normalization as in (7) gives the parameter free quantity, $\mathrm{NRR}((\mu, \sigma), \eta) = \frac{\rho\left(Z - \frac{1}{\sqrt{n}} U - \sqrt{\frac{V}{n}} c\right)}{c}$. In Section 4 it will be seen that this is generally the case for location-scale families.

Table 1: Normalized residual estimation risk for a normally distributed risk with risk measure TVaR_p, and the MLE capital estimator.

	n=20	n=50	n=100
p=0.95	0.112	0.046	0.023
p=0.99	0.141	0.059	0.030
p=0.995	0.154	0.065	0.033

Table 1 presents the normalized residual risks for TVaR_p and different values of p and n. For this risk measure, it is $c = \mathrm{TVaR}_q(Z) = \frac{\phi(\Phi^{-1}(q))}{1-q}$ (e.g. McNeil et al., 2005, Example 2.18). The values obtained demonstrate the substantial sensitivity of residual risk on the sample size. Thus, when $n = 20$ the residual risk is approximately between 11% and 15%

of the required capital, while for a moderate sample of $n = 100$, the residual risk takes values around 3% of capital. □

For the commonly used log-normal model, some of the simplicity of Example 1 is lost.

Example 2 (Log-normal, MLE)**.** Let $Y' \sim \mathcal{N}(\mu, \sigma^2)$ such that $Y = e^{Y'}$ is a log-normal random variable. Here we use an $\mathrm{RVaR}_{p_1,p_2}(Y)$ measure, which can be calculated as

$$\begin{aligned}\mathrm{RVaR}_{p_1,p_2}(Y) &= \frac{1}{p_2 - p_1} \int_{\mathrm{VaR}_{p_1}(Y)}^{\mathrm{VaR}_{p_2}(Y)} \frac{1}{t\sqrt{2\pi\sigma^2}} t e^{-\frac{(\ln(t)-\mu)^2}{\sigma}} dt \\ &= \frac{e^{\mu+\frac{1}{2}\sigma^2}}{p_2 - p_1} \left[\Phi(\Phi^{-1}(p_2) - \sigma) - \Phi(\Phi^{-1}(p_1) - \sigma)\right].\end{aligned}$$

Thus the capital estimator is

$$\eta(\mathbf{X}) = \frac{e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}}{p_2 - p_1} \left[\Phi(\Phi^{-1}(p_2) - \hat{\sigma}) - \Phi(\Phi^{-1}(p_1) - \hat{\sigma})\right].$$

The residual risk for a log-normal random variable is given by

$$\mathrm{RR}((\mu,\sigma),\eta) = \mathrm{RVaR}_{p_1,p_2}\left(Y - \frac{e^{\hat{\mu}+\frac{1}{2}\hat{\sigma}^2}}{p_2 - p_1}\left(\Phi\left(\Phi^{-1}(p_2) - \hat{\sigma}\right) - \Phi\left(\Phi^{-1}(p_1) - \hat{\sigma}\right)\right)\right),$$

where $(\hat{\mu}, \hat{\sigma}^2)$ is a random variable with the same distribution as in Example 1.

Values of normalized residual estimation risk for the log-normal distribution are presented in Table 2. For (μ, σ), the parameter choices $(4.6002, 0.0998)$ and $(4.4936, 0.4724)$ are used, corresponding to the same mean $\mathbb{E}(Y) = 100$ and coefficients of variation $\mathrm{CV}(Y) = \sqrt{\mathrm{Var}(Y)}/\mathbb{E}(Y)$ taking values 0.1 and 0.5 respectively. A substantial dependence of the normalized residual risk on the coefficient of variation is observed, with a more volatile distribution leading to higher residual estimation risks. Residual risk is also greater for higher p_1, corresponding to risk measures focusing further in the tail. □

Coherent risk measures, such as TVaR, are not well defined for random losses with infinite means (e.g. Nešlehová et al., 2006). The implication of this for the quantification of residual risk are illustrated in the following example.

Example 3 (Pareto, MLE)**.** Let $Y, \mathbf{X}$ follow a one-parameter Pareto with distribution function $F(y;\theta) = 1 - y^{-1/\theta},\ y \geq 1$, which has a finite mean for $\theta < 1$. The MLE of θ is $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} \log(X_i) \sim Gam(n, \theta/n)$, where $Gam(a, b)$ denotes a Gamma distribution with

Table 2: Normalized residual estimation risk for a log-normally distributed risk with risk measure $\text{RVaR}_{p_1,0.997}$, and the MLE capital estimator.

	$\text{CV}(Y) = 0.1$			$\text{CV}(Y) = 0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	0.119	0.049	0.025	0.163	0.071	0.037
$p_1 = 0.99$	0.147	0.062	0.031	0.200	0.091	0.048
$p_1 = 0.995$	0.156	0.066	0.034	0.212	0.098	0.052

Table 3: Normalized residual estimation risk for a Pareto distributed risk, risk measure $\text{RVaR}_{p_1,0.997}$, and MLE capital estimator.

	$\theta = 0.1$			$\theta = 0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	0.130	0.057	0.030	0.207	0.107	0.060
$p_1 = 0.99$	0.156	0.071	0.038	0.227	0.123	0.070
$p_1 = 0.995$	0.165	0.077	0.040	0.237	0.130	0.075

parameters a and b. It is apparent that $\mathbb{P}(\hat{\theta} \geq 1) > 0$. Hence even though $\mathbb{E}(Y) < \infty$, there are probable outcomes of $\hat{\theta}$ such that the capital estimator $\eta(\mathbf{X}) = \rho[F(\cdot;\hat{\theta})]$ is not well defined for coherent risk measures, such as TVaR, that require a finite mean. As a consequence, the residual risk $\text{RR}(\theta,\eta) = \rho(Y - \rho[F(\cdot;\hat{\theta})])$ may also be not well defined.

For the Pareto distribution, simple computations lead to

$$\eta(\mathbf{X}) = \text{RVaR}_{p_1,p_2}[F(\cdot;\hat{\theta})] = \frac{1}{p_2 - p_1}\frac{1}{\hat{\theta} - 1}\left((1-p_2)^{1-\hat{\theta}} - (1-p_1)^{1-\hat{\theta}}\right).$$

The residual estimation risk becomes

$$\text{RR}(\theta,\eta) = \text{RVaR}_{p_1,p_2}\left(Y - \frac{1}{p_2 - p_1}\frac{1}{\hat{\theta} - 1}\left((1-p_2)^{1-\hat{\theta}} - (1-p_1)^{1-\hat{\theta}}\right)\right),$$

where $\hat{\theta} \sim Gam(n, \theta/n)$ is once more considered as a random variable.

Normalized residual estimation risks for the Pareto distribution are presented in Table 3 for parameter values $\theta = 0.1$ and $\theta = 0.5$ (corresponding to the case of an infinite variance). The risk measure RVaR_{p_1,p_2} is used throughout, with $p_1 \in \{0.95, 0.99, 0.995\}$ and $p_2 = 0.997$. Consistently with Example 2, residual risk increases in θ (more heavy tailed distribution) and p_1 (more extreme risk measure). □

3 Controlling residual estimation risk

A further step beyond quantification of residual risk is its control, that is, the design of capital estimators that produce a residual risk (close to) zero. In all cases we have considered, residual risk under an MLE capital estimator is positive. Consequently, the capital estimation procedures we propose here tend to increase the capital in relation to the MLE.

3.1 Parametric bootstrap

In this section, we aim to adjust $\eta(\mathbf{X})$ via estimation of the residual risk $\mathrm{RR}(\theta, \eta)$ that it gives rise to. Since $\mathrm{RR}(\theta, \eta)$ specifically depends on the unknown parameter θ, it is not known to an agent who just observes the random sample $\mathbf{X}$. Thus $\mathrm{RR}(\theta, \eta)$ needs itself to be estimated from the data.

This approach is a form of parametric bootstrapping; the overall principle is as follows. Let the distribution of the parameter estimator $\hat{\theta}$ be $G(\cdot; \theta)$. Then, given $\hat{\theta}$, the bootstrapped estimator $\hat{\theta}^*$ is defined as having distribution $G(\cdot; \hat{\theta})$. The relationship between $\hat{\theta}^*$ and $\hat{\theta}$ mirrors the relationship between $\hat{\theta}$ and θ, which is required for inference. For a rigorous treatment of the bootstrap see Hall (1992).

To make these notions more explicit, first denote by $r_1(\theta) = \mathrm{RR}(\theta, \eta)$ the residual estimation risk as a function of only the true parameter θ. As before, $\eta(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})]$, where $\hat{\theta}$ is the MLE of θ. Since we can interpret $r_1(\theta)$ as the additional capital that needs to be subtracted from $Y - \eta(\mathbf{X})$ in order to make it acceptable, it is reasonable to propose the following *first order bootstrap* capital estimator

$$\eta_{bs1}(\mathbf{X}) = \eta(\mathbf{X}) + r_1(\hat{\theta}) = \rho[F(\cdot; \hat{\theta})] + r_1(\hat{\theta}). \tag{8}$$

In order to calculate $r_1(\hat{\theta}) = \mathrm{RR}(\hat{\theta}, \eta)$, for a given realization of $\hat{\theta}$, it is necessary to simulate from the random variable $\hat{\theta}^* | \hat{\theta} \sim G(\cdot; \hat{\theta})$. The details of this simulation are given in Section 6.2.

The above process can be repeated in order to refine the adjustment to the capital

estimator. Let the residual risk arising from using the capital estimator η_{bs1} be $r_2(\theta) = \mathrm{RR}(\theta, \eta_{bs1})$. Consequently, we can define the *second order bootstrap* capital estimator as

$$\eta_{bs2}(\mathbf{X}) = \eta_{bs1}(\mathbf{X}) + r_2(\hat{\theta}) = \rho[F(\cdot;\hat{\theta})] + r_1(\hat{\theta}) + r_2(\hat{\theta}), \tag{9}$$

and the associated residual risk by $r_3(\theta) = \mathrm{RR}(\theta, \eta_{bs2})$. The process can be further repeated in order to derive bootstrap capital estimators of higher orders.

Proposition 3.1 shows that under weak conditions repeated applications of the bootstrap correction generally produce an improvement in residual risk. In particular, monotonicity of the risk measure ρ ensures that the bootstrap correction operates in the correct direction. Subadditivity of ρ, combined with a weak requirement on the volatility of the estimator of residual risk, ensures the correction does not over- or under-shoot. The proposition is formulated in relation to a bootstrap correction applied to a generic capital estimator $\eta(\mathbf{X})$, which may itself be the product of a previously applied bootstrap correction; we thus drop the subscript from the function $r(\theta)$.

Proposition 3.1. *For $\mathbf{X}, Y \sim F(\cdot;\theta)$, parameter estimator $\hat{\theta}$, risk measure ρ, and capital estimator $\eta(\mathbf{X})$, define $r(\theta) = \mathrm{RR}(F, \eta)$ and $\eta^*(\mathbf{X}) = \eta(\mathbf{X}) + r(\hat{\theta})$.*

a) If $r(\theta) \geq 0$ for any $\theta \in \Theta$, then $\mathrm{RR}(\theta, \eta^*) \leq r(\theta)$.

If in addition ρ is subadditive and $\rho(r(\hat{\theta})) \leq 2r(\theta)$, then it is $\mathrm{RR}(\theta, \eta^*) \geq -r(\theta)$.

b) If $r(\theta) \leq 0$ for any $\theta \in \Theta$, then $\mathrm{RR}(\theta, \eta^*) \geq r(\theta)$.

If in addition ρ is subadditive and $\rho(-r(\hat{\theta})) \leq -2r(\theta)$, then it is $\mathrm{RR}(\theta, \eta^*) \leq -r(\theta)$.

Proof. To prove part *a*), by monotonicity we have

$$\mathrm{RR}(\theta, \eta^*) = \rho(Y - (\eta(\mathbf{X}) + r(\hat{\theta}))) \leq \rho(Y - \eta(\mathbf{X})) = r(\theta).$$

From the subadditivity of ρ it follows that $\rho(Y - (\eta(\mathbf{X}) + r(\hat{\theta}))) \geq \rho(Y - \eta(\mathbf{X})) - \rho(r(\hat{\theta}))$. The assumption $\rho(r(\hat{\theta})) \leq 2r(\theta)$ implies $\mathrm{RR}(\theta, \eta^*) = \rho(Y - \eta(\mathbf{X})) - \rho(r(\hat{\theta})) \geq -r(\theta)$. Part *b*) follows similarly. □

Here we demonstrate the effectiveness of the bootstrap approach for the log-normal and Pareto models.

Example 4 (Log-normal, bootstrap)**.** Expressions of the MLE capital estimator $\eta(\mathbf{X})$ and the corresponding residual risk $r_1(\mu,\sigma) = \mathrm{RR}((\mu,\sigma),\eta)$ are given in Example 2. Consequently the bootstrap capital estimator is $\eta_{bs1}(\mathbf{X}) = \eta(\mathbf{X}) + r_1(\hat{\mu},\hat{\sigma})$, where for each estimate $(\hat{\mu},\hat{\sigma})$ it is

$$r_1(\hat{\mu},\hat{\sigma}) = \mathrm{RVaR}_{p_1,p_2}\left(Y^* - \frac{e^{\hat{\mu}^*+\frac{1}{2}\hat{\sigma}^{*2}}}{p_2-p_1}\left(\Phi\left(\Phi^{-1}(p_2)-\hat{\sigma}^*\right) - \Phi\left(\Phi^{-1}(p_1)-\hat{\sigma}^*\right)\right)\right),$$

where, given $(\hat{\mu},\hat{\sigma})$, Y^* follows a log-normal distribution with parameters $(\hat{\mu},\hat{\sigma}^2)$ and $(\hat{\mu}^*,\hat{\sigma}^*) \overset{d}{=} \left(\hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}}U^*,\ \frac{\hat{\sigma}^2 V^*}{n}\right)$, with $U^* \sim \mathcal{N}(0,1)$ and $V^* \sim \chi^2_{n-1}$ independent. See Section 6.2 for a general description of the algorithm for deriving η_{bs1} by simulation.

Normalized residual risk is reported in Table 4. A dramatic improvement is observed in comparison to the unadjusted MLE capital estimator (Table 2), with even a single application of the bootstrap correction leading to a residual risk of nearly zero. □

Table 4: Normalized residual estimation risk for a log-normally distributed risk with risk measure $\mathrm{RVaR}_{p_1,0.997}$, and first-order bootstrap corrected capital estimator η_{bs1}.

	$\mathrm{CV}(Y)=0.1$			$\mathrm{CV}(Y)=0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	0.001	-0.001	0.000	0.002	0.001	0.000
$p_1 = 0.99$	0.001	0.000	0.000	0.002	0.001	0.000
$p_1 = 0.995$	0.002	0.001	0.000	0.001	0.000	0.000

Table 5: Normalized residual estimation risk for a Pareto distributed risk, risk measure $\mathrm{RVaR}_{p_1,0.997}$, and first-order bootstrap corrected capital estimator η_{bs1}.

	$\theta=0.1$			$\theta=0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	0.001	0.001	0.000	0.001	0.000	-0.002
$p_1 = 0.99$	0.001	-0.002	-0.002	0.003	0.001	0.002
$p_1 = 0.995$	0.000	-0.001	0.000	0.002	0.002	0.001

Example 5 (Pareto, bootstrap)**.** Expressions of the MLE capital estimator $\eta(\mathbf{X})$ and the corresponding residual risk $r_1(\theta) = \mathrm{RR}(\theta,\eta)$ are given in Example 3. Once more, the

bootstrap capital estimator is $\eta_{bs1}(\mathbf{X}) = \eta(\mathbf{X}) + r_1(\hat{\theta})$, where for each estimate $\hat{\theta}$ it is

$$r_1(\hat{\theta}) = \mathrm{RVaR}_{p_1,p_2}\left(Y^* - \frac{1}{p_2 - p_1}\frac{1}{\hat{\theta}^* - 1}\left((1-p_2)^{1-\hat{\theta}^*} - (1-p_1)^{1-\hat{\theta}^*}\right)\right),$$

where given $\hat{\theta}$, Y^* follows a Pareto distribution with parameter $\hat{\theta}$ and $\hat{\theta}^*$ a Gamma distribution with parameters $(n, \hat{\theta}/n)$.

Continuing from Example 3, residual estimation risk is calculated for a Pareto distribution and the capital estimator $\eta_{bs1}(\mathbf{X})$. Results are reported in Table 5. Once more, residual risk is essentially eliminated, demonstrating a vast improvement in comparison to the MLE capital estimator (Table 3). □

3.2 Bayesian predictive distribution

The use of a Bayesian predictive distribution is a standard approach to dealing with parameter uncertainty, see Cairns (2000). Under a Bayesian approach, the parameter $\theta \in \Theta$ is considered a random variable itself with *prior* distribution $\pi(\theta)$. Once data $\mathbf{x}$ have been collected, the *posterior* of the parameter, $\pi(\theta|\mathbf{x})$, is obtained by $\pi(\theta|\mathbf{x}) \propto \pi(\theta)\prod_{i=1}^{n} f(x_i;\theta)$. The *predictive* distribution of Y, given the data $\mathbf{x}$, is defined as

$$\hat{F}(\cdot|\mathbf{x}) = \int_{\theta\in\Theta} F(\cdot;\theta)\pi(\theta|\mathbf{x})d\theta. \tag{10}$$

Probabilities and expectations calculated according to the predictive distribution are respectively denoted by $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\hat{\mathbb{E}}(\cdot|\mathbf{x})$.

Parameter uncertainty can be reflected in capital measurement by setting capital according to the predictive distribution. That is, we set

$$\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})]. \tag{11}$$

Note the difference to the bootstrap capital estimators of Section 3.1: there an adjustment to the MLE was produced, while here the probability distribution according to which capital is set is modified.

Predictive distributions tend to be more dispersed in the tail, by their mixture construction seen in (10). Hence, it is plausible that the capital estimator $\eta_{bay}(\mathbf{X})$ will produce

a reduction in the (typically positive) residual risk induced by an MLE capital estimator, such that $\mathrm{RR}(\theta, \rho[\hat{F}(\cdot|\mathbf{X})]) \leq \mathrm{RR}(\theta, \rho[F(\cdot; \hat{\theta})])$. While the residual risk is a frequentist criterion, it has been widely noticed in the literature that Bayesian approaches to prediction tend, at least approximately, to satisfy frequentist quality criteria (see Smith, 1999; Datta et al., 2000). In fact, it follows from results of Gerrard and Tsanakas (2011) that, for a wide set of loss distributions that includes the log-normal and Pareto examples discussed here and the use of a non-informative prior, it is $\mathrm{VaR}_p(Y - \mathrm{VaR}_p[\hat{F}(\cdot|\mathbf{X})]) = 0$, such that the residual risk is completely eliminated for VaR.

The effectiveness of the capital estimator η_{bay} is now demonstrated through examples. It is seen the problems of infinite means emerge for both the log-normal and Pareto models, motivating once more the use of RVaR.

Example 6 (Log-normal, Bayes)**.** Let $Y', \mathbf{X}' \sim \mathcal{N}(\mu, \sigma^2)$ and $Y = \exp(Y')$, $\mathbf{X} = (\exp(X_1'), \ldots, \exp(X_n'))$, such that $Y, \mathbf{X} \sim \mathcal{LN}(\mu, \sigma^2)$. For the log-normal distribution all moments exist, regardless of the value of the parameters, such that for a coherent risk measure like TVaR, the quantity $\rho[F(\cdot; (\hat{\mu}, \hat{\sigma}))]$ will always be well defined.

Consider now capital being set using the predictive distribution of Y, such that $\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})]$. A standard argument (similar to Hogg et al., 2012, Example 11.3.1) shows that, using an uninformative prior $\pi(\mu, \sigma) = 1/\sigma$, the predictive distribution of the normal variable Y' is a Student t distribution. Consequently, (see e.g. Gerrard and Tsanakas, 2011, Lemma 1ii) the predictive distribution of the log-normal variable Y is a "log-t" distribution

$$\hat{F}(y|\mathbf{X}) = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{\log(y) - \hat{\mu}}{\hat{\sigma}}\right),$$

where $\hat{\mu}$, $\hat{\sigma}$ are the MLEs of μ, σ, and t_{n-1} is the distribution function of a standard Student t variable with $n-1$ degrees of freedom.

The expected value associated with $\hat{F}(\cdot|\mathbf{X})$ is $\hat{E}(Y|\mathbf{X}) = \hat{E}\left(\exp(Y)|\mathbf{X}\right)$. However, since the Student t distribution has a regularly varying tail (McNeil et al., 2005, p. 293), its moment generating function is not well defined, implying that $\hat{E}(\exp(Y')|\mathbf{X}) = \infty$. Since the mean associated with the predictive distribution $\hat{F}(\cdot|\mathbf{X})$ is infinite, any capital

estimator of the form $\rho[\hat{F}(\cdot|\mathbf{X})]$ will also be infinite, when a coherent risk measure ρ is used.

Thus the use of the risk measure RVaR is more appropriate. We have

$$\mathrm{VaR}_p[\hat{F}'(y|\mathbf{X}')] = \exp\left(\hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}t_{n-1}^{-1}(p)\right),$$

$$\eta_{bay}(\mathbf{X}) = \mathrm{RVaR}_{p_1,p_2}[\hat{F}'(y|\mathbf{X}')] = \frac{1}{p_2-p_1}\int_{p_1}^{p_2}\exp\left(\hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}t_{n-1}^{-1}(u)\right)du. \quad (12)$$

The integral in (12) needs to be solved numerically.

The normalized residual estimation risk for the log-normal distribution is presented in Table 6. It is seen that using the predictive distribution is highly effective in nearly eliminating residual risk. In particular, comparison to Table 2 reveals the great improvement achieved in relation to MLE, with residual risk for η_{bay} being close to zero. The performance of η_{bay} is thus comparable to that of η_{bs1} reported in Table 4. □

Table 6: Normalized residual estimation risk for a log-normally distributed risk, risk measure $\mathrm{RVaR}_{p_1,0.997}$, and Bayes capital estimator η_{bay}.

	CV$(Y) = 0.1$			CV$(Y) = 0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	-0.005	-0.001	-0.001	-0.008	-0.003	-0.001
$p_1 = 0.99$	-0.001	0.000	0.000	-0.001	0.000	0.000
$p_1 = 0.995$	0.000	0.000	0.000	0.000	0.000	0.000

Table 7: Normalized residual estimation risk for a Pareto distributed risk, risk measure $\mathrm{RVaR}_{p_1,0.997}$, and Bayes capital estimator η_{bay}.

	$\theta = 0.1$			$\theta = 0.5$		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	-0.005	-0.002	-0.001	0.018	0.012	0.008
$p_1 = 0.99$	0.000	0.000	0.000	0.007	0.006	0.004
$p_1 = 0.995$	0.000	0.000	0.000	0.002	0.002	0.002

Example 7 (Pareto, Bayes)**.** When Y follows a Pareto distribution, the prior $\pi(\theta) = 1/\theta$

leads to a predictive distribution for Y of the form

$$\hat{F}(y|\mathbf{X}) = 1 - \left(\frac{n\hat{\theta}}{\log(y) + n\hat{\theta}} \right)^n, \tag{13}$$

where $\hat{\theta}$ is the MLE of θ. This is a "log-Pareto" distribution, again with infinite mean. For the VaR and RVaR measures of Y we now have,

$$\begin{aligned} \mathrm{VaR}_p[\hat{F}(y|\mathbf{X})] &= \exp\left(\hat{\theta} n((1-p)^{-1/n} - 1)\right), \\ \eta_{bay}(\mathbf{X}) = \mathrm{RVaR}_{p_1,p_2}[\hat{F}(y|\mathbf{X})] &= \frac{1}{p_2 - p_1} \int_{p_1}^{p_2} \exp\left(\hat{\theta} n((1-u)^{-1/n} - 1)\right) du. \end{aligned} \tag{14}$$

Again, the integral in (14) can be solved numerically.

The normalized residual estimation risk for the Pareto distribution is presented in Table 7. Once more, the use of the predictive distribution is highly effective, leading to residual risk levels very close to zero, thus improving on the MLE capital estimators (Table 3). □

4 Quantifying and controlling residual estimation risk for location-scale families

In the current section we focus on residual estimation risk for distribution functions that belong to location-scale families. Such distributions, like the normal, Student t, and Laplace (double-exponential) families are commonly used in modeling asset returns. It will be seen that the case of location-scale families allows substantial simplifications in the quantification and control of residual risk. In particular, exact elimination of residual risk is possible.

4.1 Residual estimation risk for location-scale families

Two random variables Y and Z belong to the same *location-scale family*, if there exist $a \in \mathbb{R}$ and $b > 0$, such that $Y \stackrel{d}{=} bZ + a$. Denote the parameter vector $\theta = (\mu, \sigma)$, such that any random variable in the location-scale family follows $F(\cdot; (\mu, \sigma))$. We say that

$Z \sim F(\cdot;(0,1))$ has a standardized distribution and simply denote it by $F \equiv F(\cdot;(0,1))$. Hence, we can write $Y \sim F(\cdot;(\mu,\sigma)) = F\left(\frac{\cdot-\mu}{\sigma}\right)$.

Estimators of location and scale parameters generally also belong to location-scale families. Specifically, if the parameter vector $\theta = (\mu, \sigma)$ is estimated via Maximum Likelihood, then a standard argument (e.g. Gerrard and Tsanakas, 2011, Lemma 4) shows

$$\hat{\mu} \stackrel{d}{=} \mu + \sigma U, \quad \hat{\sigma} \stackrel{d}{=} \sigma V, \tag{15}$$

where U and V are random variables whose distribution depends on the sample size n, but not on θ.

From the translation invariance, positive homogeneity, and law invariance properties of the risk measure, it follows that for $Y \sim F(\cdot;(\mu,\sigma)),\ Z \sim F$, it is

$$\rho(Y) = \rho(\mu + \sigma Z) = \mu + \sigma\rho[F].$$

Let the capital estimator be based on the MLE, such that $\eta(\mathbf{X}) = \rho[F(\cdot;\hat{\theta})]$, where $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$. Thus, it is $\eta(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\rho[F] = \mu + \sigma U + \sigma V\rho[F]$.

Consequently, the residual estimation risk can be calculated as

$$\mathrm{RR}(\theta,\eta) = \rho(\mu + \sigma Z - \mu - \sigma U - \sigma V\rho[F]) = \sigma\rho(Z - U - V\rho[F]). \tag{16}$$

Hence, while in general the residual estimation risk remains unknown, for location-scale families it does not depend on the location parameter μ and is directly proportional to the scale one σ. In particular, the amount $\rho(Z - U - V\rho[F])$ does not depend on the unknown parameters. This effect was demonstrated in Example 1, when dealing with normally distributed losses.

4.2 Adjustment to the risk measure

In the case of location-scale families it is possible to modify the risk measure in a way that compensates for parameter uncertainty and brings the residual estimation risk down to zero. Consider another risk measure ρ_{adj}, that may be used to set capital. Under this

risk measure, the capital estimator, using again MLE, will be

$$\eta_{adj}(\mathbf{X}) = \rho_{adj}[F(\cdot;\hat{\theta})].$$

Analogously with (16), we can write

$$\mathrm{RR}(\theta, \eta_{adj}) = \rho(Y - \rho_{adj}[F(\cdot;\hat{\theta})]) = \sigma\rho(Z - U - V\rho_{adj}[F]). \tag{17}$$

Noting that the quantity $\rho(Z-U-V\rho_{adj}[F])$ does not depend on the true but unknown parameter θ, it becomes apparent that we can choose the risk measure $\rho_{adj}[F]$ specifically so as to set the residual risk of (17) to zero. For example, if $\rho = \mathrm{TVaR}_p$, we can let $\rho_{adj} = \mathrm{TVaR}_q$ for some $q \neq p$. The process is illustrated by the following example.

Example 8 (Normal, adjusted TVaR)**.** Consider a normal distribution and let $\rho = \mathrm{TVaR}_p$, $\rho_{adj} = \mathrm{TVaR}_q$. The the random variables U, V in (15) become

$$U = \frac{1}{\sqrt{n}}U', \quad V = \sqrt{\frac{V'}{n}},$$

where $U' \sim \mathcal{N}(0,1)$ and $V' \sim \chi^2_{n-1}$, with U', V' independent. Noting that $\mathrm{TVaR}_q(Z) = \frac{\phi(\Phi^{-1}(q))}{1-q}$, the residual risk under the adjusted capital estimator is

$$\mathrm{RR}((\mu,\sigma), \eta_{adj}) = \sigma\mathrm{TVaR}_p\left(Z - \frac{1}{\sqrt{n}}U' - \sqrt{\frac{V'}{n}}\frac{\phi(\Phi^{-1}(q))}{1-q}\right),$$

and a level q that sets the above expression to zero can be found.

To simplify exposition, assume now that the scale parameter σ is known. Then

$$\begin{aligned}\mathrm{RR}(\mu, \eta_{adj}) &= \sigma\mathrm{TVaR}_p\left(Z - \frac{1}{\sqrt{n}}U' - \frac{\phi(\Phi^{-1}(q))}{1-q}\right)\\ &= \sigma\sqrt{1+\frac{1}{n}}\frac{\phi(\Phi^{-1}(p))}{1-p} - \sigma\frac{\phi(\Phi^{-1}(q))}{1-q}.\end{aligned}$$

Therefore, to achieve $\mathrm{RR}(\mu, \eta_{adj}) = 0$, one needs to solve for q the equation

$$\sqrt{1+\frac{1}{n}}\frac{\phi(\Phi^{-1}(p))}{1-p} = \frac{\phi(\Phi^{-1}(q))}{1-q},$$

which is easily done numerically. The resulting adjusted capital estimator thus is

$$\eta_{adj}(\mathbf{X}) = \hat{\mu} + \sigma\mathrm{TVaR}_q(Z) = \hat{\mu} + \sigma\sqrt{1+\frac{1}{n}}\mathrm{TVaR}_p(Z).$$

The required level of q is plotted in Figure 1, against the sample size n, for $p \in \{0.95, 0.99, 0.995\}$. It can be seen that in each case $q > p$ and as the sample size increases

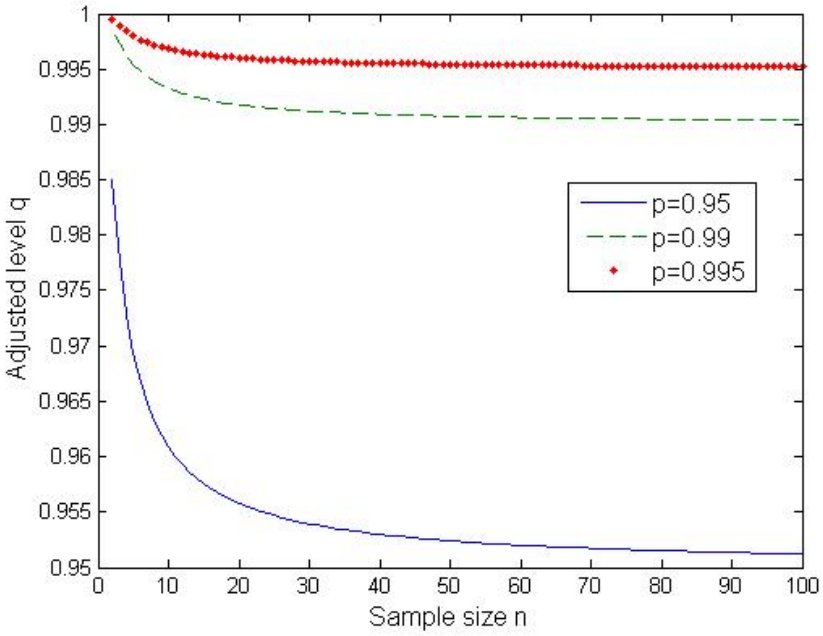

Figure 1: Confidence level q required to eliminate the residual estimation risk for a normal random variable with known scale parameter and risk measure TVaR_p.

the adjusted confidence level q decays to the nominal level p. The difference $q - p$ is more pronounced for very small sample sizes, such that, if η_{adj} were adopted, portfolios with a longer history would be subject to a lower capital requirement. □

4.3 Bootstrap procedure for location-scale families

In Section 3.1 it was demonstrated that repeated bootstrap corrections to the capital estimator produce improvements in residual risk. However, these improvements come at a cost, since each iteration induces a nested simulation. Here it is shown that, for location-scale families, higher order bootstrap capital estimators can be derived exactly, avoiding the need for nested simulations.

Using again the notation of Section 3.1, for a location-scale family the residual risk of

the MLE and the first-order bootstrap capital estimator are

$$r_1(\theta) = \sigma\rho(Z - U - V\rho(Z)),$$
$$\eta_{bs1}(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\rho(Z) + r_1(\hat{\theta}) = \mu + \sigma U + \sigma V(\rho(Z) + \rho(Z - U - V\rho(Z))). \tag{18}$$

It follows that

$$r_2(\theta) = \rho(Y - \eta_{bs1}(\mathbf{X})) = \sigma\rho(Z - U - V(\rho(Z) + \rho(Z - U - V\rho(Z)))),$$
$$\eta_{bs2}(\mathbf{X}) = \eta_{bs1}(\mathbf{X}) + r_2(\hat{\theta})$$
$$= \hat{\mu} + \hat{\sigma}\left[\rho(Z) + \rho(Z - U - V\rho(Z)) + \rho(Z - U - V(\rho(Z) + \rho(Z - U - V\rho(Z))))\right]. \tag{19}$$

Since the distribution of the random variables Z, U, V does not depend on the true parameters (μ, σ), formulas (18) and (19) can be evaluated from a single set of simulated values from Z, U, V. The above argument can be extended to an arbitrary number of bootstrap iterations.

It is also noted that for the case of location families, where the scale parameter is known, the first-order bootstrap corrected capital estimator gives an exact elimination of residual risk. To see that, one may follow the same steps as above, setting without loss of generality $V = 1$. Then, $r_2(\mu) = \sigma\rho(Z - U - (\rho(Z) + \rho(Z - U - \rho(Z)))) = 0$.

Table 8: Normalized residual estimation risk for a normally distributed risk with sample size n, risk measure TVaR_p, and the bootstrap capital estimators $\eta_{bs1},\ \eta_{bs2}$.

	η_{bs1}			η_{bs2}		
	n=20	n=50	n=100	n=20	n=50	n=100
$p_1 = 0.95$	0.012	0.002	0.000	0.002	0.000	0.000
$p_1 = 0.99$	0.023	0.003	0.001	0.004	0.000	0.000
$p_1 = 0.995$	0.030	0.005	0.001	0.006	0.000	0.000

Example 9 (Normal, bootstrap)**.** Residual risk is now calculated for a normally distributed risk, a TVaR_p risk measure, and the first- and second-order bootstrap capital estimators $\eta_{bs1}(\mathbf{X}), \eta_{bs2}(\mathbf{X})$. Results are reported in Table 8. The first-order bootstrap capital estimator η_{bs1} reduces residual risk compared to the case of the MLE capital

estimator, while the second-order bootstrap estimator reduces the capital even further.

In the particular case of a known standard deviation, we have

$$\begin{aligned} r_1(\mu) &= \mathrm{TVaR}_p(\mu + \sigma Z - \hat{\mu} - \sigma \mathrm{TVaR}_p(Z)) \\ &= \sigma \mathrm{TVaR}_p\left(Z - \frac{1}{\sqrt{n}}U'\right) - \sigma \mathrm{TVaR}_p(Z) \\ &= \sigma\sqrt{1+\frac{1}{n}}\mathrm{TVaR}_p\left(Z\right) - \sigma \mathrm{TVaR}_p(Z), \\ \eta_{bs1}(\mathbf{X}) &= \hat{\mu} + \sigma\sqrt{1+\frac{1}{n}}\mathrm{TVaR}_p\left(Z\right). \end{aligned}$$

Hence, η_{bs1} is exactly the same capital estimator as η_{adj} considered in Example 8, satisfying $\mathrm{RR}(\mu, \eta_{adj}) = 0$. □

4.4 Bayesian predictive distribution for location-scale families

In this section we show that for location-scale families, (a) when the scale parameter is known, residual risk is completely eliminated, and (b) when the location parameter is known, a quantity similar to residual risk equals zero.

Before stating the results we reformulate without proof the content of Proposition 1 in Severini et al. (2002), which is used in the present section. For the sake of simplicity, details about the technical conditions are omitted, but Example 1 in Severini et al. (2002), implies that location-scale families satisfy all the necessary conditions to apply the proposition.

Proposition 4.1. *Severini et al. (2002). For $Y, \mathbf{X} \sim F(\cdot;\theta)$ belonging to a location-scale family with $\theta = (\mu, \sigma) \in \Theta$, let $H(\mathbf{X})$ be a region such that $\hat{\mathbb{P}}(Y \in H(\mathbf{X})|\mathbf{x}) = 1 - \alpha$. Assume that H satisfies the following conditions:*

(i) For each $\theta = (\mu, \sigma) \in \Theta$, $y \in H(\mathbf{x})$ if and only if $y + \mu \in H(\mu + \mathbf{x})$ (for location models) and $\sigma y \in H(\sigma \mathbf{x})$ (for scale models).

(ii) Let $C(\mathbf{x}, y) = 1$ if $y \in H(\mathbf{x})$ and 0 otherwise. There exists $0 < \alpha < 1$, such that $\hat{E}[C(\mathbf{X}, Y)|\mathbf{x}] = 1 - \alpha$.

It follows that $E_\theta[C(\mathbf{X}, Y)] = 1 - \alpha$.

Consider first a location family with parameter $\theta = \mu$. The prior $\pi(\theta) = 1$ is used. It is known that (e.g. see Gerrard and Tsanakas, 2011), if $\mathbf{X} = \mathbf{Z} + b$, where $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $\mathbf{Z} \sim F$, then $\hat{F}(y|\mathbf{z} + b) = \hat{F}(y - b|\mathbf{z})$. Therefore,

$$\rho[\hat{F}(\cdot|\mathbf{x})] = \rho[\hat{F}(\cdot - b|\mathbf{z})] = \rho[\hat{F}(\cdot|\mathbf{z})] + b,$$

due to the translation invariance property of ρ.

Proposition 4.2 shows that using the predictive distribution eliminates residual risk for location families.

Proposition 4.2. *For location families, using the capital estimator* $\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})]$ *yields*

$$\rho(Y - \eta_{bay}(\mathbf{X})) = 0.$$

Proof. The proof follows from an application of Prop. 4.1. Consider the predictive region

$$H_c(\mathbf{X}) = (-\infty,\ \rho[\hat{F}(\cdot|\mathbf{X})] + c]$$

for any constant $c \in \mathbb{R}$. This region is invariant as required, indeed:

$$Y + b \in H_c(\mathbf{X} + b) \Leftrightarrow Y + b \ \leq\ \rho[\hat{F}(\cdot|\mathbf{X} + b)] + c \ =\ \rho[\hat{F}(\cdot|\mathbf{X})] + b + c \Leftrightarrow$$

$$Y \ \leq\ \rho[\hat{F}(\cdot|\mathbf{X})] + c \Leftrightarrow Y \in H_c(\mathbf{X}).$$

It follows that $\hat{\mathbb{P}}(Y - \rho[\hat{F}(\cdot|\mathbf{X})] \leq c|\mathbf{x}) = \mathbb{P}(Y - \rho[\hat{F}(\cdot|\mathbf{X})] \leq c)\ \forall c \in \mathbb{R}$. As this holds for every $c \in \mathbb{R}$, it is implied that the random variable $W = Y - \rho[\hat{F}(\cdot|\mathbf{X})]$ has the same distribution under $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\mathbb{P}(\cdot)$. Thus if $G(w) = \mathbb{P}(W \leq w)$ and $\hat{G}(w|\mathbf{x}) = \hat{\mathbb{P}}(W \leq w|\mathbf{x})$ it is $G(w) = \hat{G}(w|\mathbf{x})$ for all w. By law invariance of ρ it then is $\rho[G(\cdot)] = \rho[\hat{G}(\cdot|\mathbf{x})]$. However, by the construction of the random variable W it is $\rho[\hat{G}(\cdot|\mathbf{x})] = 0$. Hence $\rho[G(\cdot)] = \rho(Y - \rho[\hat{F}(\cdot|\mathbf{X})]) = 0$. □

Suppose now that Y belongs to a scale family, with parameter $\theta = \sigma$. We use the prior $\pi(\theta) = 1/\theta$. If $\mathbf{X} = b\mathbf{Z}$, where $b > 0$, $\mathbf{Z} = (Z_1, \ldots, Z_n)$ and $\mathbf{Z} \sim F$, then $\hat{F}(y|b\mathbf{z}) = \hat{F}(y/b|\mathbf{z})$. Therefore,

$$\rho[\hat{F}(\cdot|\mathbf{x})] = \rho[\hat{F}(\cdot/b|\mathbf{z})] = b\rho[\hat{F}(\cdot|\mathbf{z})],$$

due to the positive homogeneity property of ρ.

Proposition 4.3 shows that for scale-families the capital estimator $\eta_{bay}(\mathbf{X})$ leads to elimination of a scaled version of the residual risk.

Proposition 4.3. *For scale families, using the capital estimator* $\eta_{bay}(\mathbf{X}) = \rho[\hat{F}(\cdot|\mathbf{X})]$ *yields*

$$\rho\left(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]} - 1\right) = 0.$$

Proof. The same procedure as in the proof of Proposition 4.2 is followed. The predictive region is

$$H_c(\mathbf{X}) = (-\infty, c\rho[\hat{F}(\cdot|\mathbf{X})]]$$

for any constant $c \in \mathbb{R}$. This region is invariant as required in Prop 4.1, since

$$bY \in H_c(b\mathbf{X}) \Leftrightarrow bY \;\leq\; c\hat{\rho}[\hat{F}(\cdot|b\mathbf{X})] \;=\; cb\hat{\rho}[\hat{F}(\cdot|\mathbf{X})] \Leftrightarrow$$

$$Y \;\leq\; c\rho[\hat{F}(\cdot|\mathbf{X})] \Leftrightarrow Y \in H_c(\mathbf{X}).$$

It follows that:

$$\hat{\mathbb{P}}\left(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]} \leq c|\mathbf{x}\right) = \mathbb{P}\left(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]} \leq c\right) \quad \forall c \in \mathbb{R}.$$

As this holds for every $c \in \mathbb{R}$, it is implied that the random variable $W = Y/\rho[\hat{F}(\cdot|\mathbf{X})]$ has the same distribution under $\hat{\mathbb{P}}(\cdot|\mathbf{x})$ and $\mathbb{P}(\cdot)$. Thus if $G(w) = \mathbb{P}(W \leq w)$ and $\hat{G}(w|\mathbf{x}) = \hat{\mathbb{P}}(W \leq w|\mathbf{x})$ it is $G(w) = \hat{G}(w|\mathbf{x})$ for all w. By law invariance of ρ it then is $\rho[G(\cdot)] = \rho[\hat{G}(\cdot|\mathbf{x})]$. However, by the construction of the random variable W it is $\rho[\hat{G}(\cdot|\mathbf{x})] = 1$. Hence $\rho[G(\cdot)] = \rho\left(\frac{Y}{\rho[\hat{F}(\cdot|\mathbf{X})]}\right) = 1$. □

For the more general location-scale case, the effectiveness of using a predictive distribution is demonstrated via the following example.

Example 10 (Normal, Bayes)**.** For a normal distribution and prior $\pi(\mu, \sigma) = 1/\sigma$, a standard argument similar to Hogg et al. (2012, Example 11.3.1) shows that the predictive distribution is a Student t distribution

$$\hat{F}(y|\mathbf{X}) = t_{n-1}\left(\sqrt{\frac{n-1}{n+1}}\frac{y-\hat{\mu}}{\hat{\sigma}}\right), \tag{20}$$

where $\hat{\mu}$, $\hat{\sigma}$ are the MLEs of μ, σ, and t_{n-1} is the distribution function of a standard t variable with $n-1$ degrees of freedom. The corresponding value of TVaR is (McNeil et al., 2005, Example 2.19)

$$\eta_{bay}(\mathbf{X}) = \mathrm{TVaR}_p[\hat{F}(y|\mathbf{X})] = \hat{\mu} + \hat{\sigma}\sqrt{\frac{n+1}{n-1}}\left(\frac{g_{n-1}(t_{n-1}^{-1}(p))}{1-p}\right)\left(\frac{n-1+(t_{n-1}^{-1}(p))^2}{n-2}\right), \tag{21}$$

where g_{n-1} is the density of a standard t variable with $n-1$ degrees of freedom. The Student t predictive distribution is heavy-tailed, which generally leads to higher estimated capital levels than the normal.

In Table 9 the corresponding normalized residual risks are reported. Once more, the effectiveness of using η_{bay} is apparent, with a near elimination of residual risk observed. The performance is comparable to the second-order bootstrap estimator seen in Table 8, though η_{bay} appears to slightly overcompensate in increasing capital estimates, leading to slightly negative residual risks. □

Table 9: Normalized residual estimation risk for a normally distributed risk with sample size n, risk measure TVaR_p, and the Bayes capital estimator η_{bay}.

	n=20	n=50	n=100
p=0.95	-0.007	-0.003	-0.001
p=0.99	-0.005	-0.002	-0.001
p=0.995	-0.005	-0.002	-0.001

4.5 The presence of a shape parameter

There are location-scale families that have additional shape parameters, such as the Student t distribution. For such distributions, the computational savings present for location-scale families cannot be fully achieved, but their parametric structure can still be exploited.

Consider a Student t random variable Y with parameters $\theta = (\mu, \sigma, \nu)$, such that $Y \overset{d}{=} \mu + \sigma Z_\nu$, where Z_ν has a standard t distribution with ν degrees of freedom. Let

$\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\nu})$ be an estimator of θ. For many classes of estimators, such as MLEs and the simple estimator described in Section 6.3, the random variables $\hat{\mu}, \hat{\sigma}$ will still follow a (location-)scale distribution, such that we may write, $\hat{\mu} = \mu + \sigma U_\nu,\ \hat{\sigma} = \sigma V_\nu,\ \hat{\nu} = W_\nu$, where the distribution of (U_ν, V_ν, W_ν) does not depend on μ or σ but depends on ν.

For parameter estimator $\hat{\theta}$, the unadjusted capital estimator is $\eta(\mathbf{X}) = \rho[F(\cdot; \hat{\theta})] = \hat{\mu} + \hat{\sigma}\rho[t_{\hat{\nu}}]$, where t_ν is the distribution of a standard Student t variable with ν degrees of freedom. Consequently, the residual estimation risk can be written as

$$\begin{aligned} &\mathrm{RR}(\theta, \eta) = \rho\left(\mu + \sigma Z_\nu - \hat{\mu} - \hat{\sigma}\rho[t_{\hat{\nu}}]\right) = \sigma\zeta(\nu), \\ &\text{where}\ \ \zeta(\nu) = \rho(Z_\nu - U_\nu - V_\nu\rho[t_{\hat{\nu}}]). \end{aligned} \tag{22}$$

Hence we can define the bootstrap estimator as

$$\eta_{bs1}(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\left(\rho[t_{\hat{\nu}}] + \zeta(\hat{\nu})\right). \tag{23}$$

To implement this estimator, numerical evaluation of the function $\zeta(\nu)$ is required. But, as ζ does not depend on the distribution parameters, nested simulations are avoided.

These ideas are demonstrated via the following example, where the issue of model error is also briefly discussed.

Example 11 (Student t, bootstrap, empirical, model error)**.** In this example we work with an RVaR_{p_1,p_2} risk measure, with $p_1 = 0.95, p_2 = 0.997$ and a Student t distribution with $\theta = (\mu, \sigma, \nu) = (0, 1, 5)$, with all three parameters considered unknown in the capital estimation. The RVaR measure of a standard t variable is given in Section 6.5). Sample sizes $n = 50, 10, 200, 500, 1000$ are considered.

Subsequently we calculate the residual estimation risk for different capital estimators. Details about how each of those capital estimators (and in particular the numerical approximation of ζ) simulated is given in Section 6.2; the assessment of residual risk for each estimator is as described in Section 6.1.

Student t: This is an unadjusted estimator of RVaR_{p_1,p_2} obtained by estimating $(\hat{\mu}, \hat{\sigma}, \hat{\nu})$ from $\mathbf{X}$ and setting $\eta(\mathbf{X}) = \hat{\mu} + \hat{\sigma}\mathrm{RVaR}[t_{\hat{\nu}}]$.

Student t (bootstrap): This is a bootstrap corrected estimator of RVaR_{p_1,p_2} obtained as $\eta_{bs1}(\mathbf{X}) = \hat{\mu} + \hat{\sigma}(\mathrm{RVaR}[t_{\hat{\nu}}] + \zeta(\hat{\nu}))$, with $\zeta(\nu)$ as in (22).

Empirical: The RVaR_{p_1,p_2} measure is directly applied in a model-free way, by the empirical distribution of the sample $\mathbf{X}$.

Normal: The possibility of model error is considered, by assuming in capital estimation that the data are actually from a normal distribution and applying the standard normal MLE of RVaR.

Table 10: Normalized residual estimation risk for a Student t_5 distributed risk with sample size n, risk measure $\mathrm{RVaR}_{0.95,0.997}$, and various capital estimators.

Estimation method	n=50	n=100	n=200	n=500	n=1000
Student t	0.054	0.026	0.017	0.009	0.004
Student t (bootstrap)	0.009	0.005	0.004	0.001	0.000
Normal	0.100	0.075	0.061	0.051	0.047
Empirical			0.051	0.023	0.005

The resulting residual estimation risk figures are shown in Table 10. For η_{emp} results are only reported for $n \geq 200$, as estimation of the given RVaR measure is not meaningful on smaller samples. The results demonstrate how the residual estimation risk of the unadjusted estimator η is nearly eliminated by the bootstrap corrected estimator η_{bs1}. The estimator η_{norm}, accounting for the case of model error, presents a substantially higher residual risk, due to the underestimation of required capital arising by the lighter normal tail that it assumes. Furthermore, residual risk does not tend to zero for increasing sample sizes. On the other hand, for $n \geq 200$, the model-free empirical estimator η_{emp} beats η_{norm}, while at the same time performing worse than the estimators η_t, η_{bst} that make use of the parametric family that the data come from. □

5 Conclusions

We introduce the notion of residual estimation risk for measuring the impact that the volatility of risk estimators has on capital adequacy. Residual risk quantifies the capital that needs to be added to a portfolio, consisting of a random loss and a random capital estimator, in order to make the total position acceptable with reference to a risk measure. In a parametric setting, this interpretation motivates the design of modified capital setting procedures, based on bootstrapping and Bayesian predictive distributions. The good performance of these approaches is demonstrated by numerical examples, both for general distributions and for location-scale families, where exact elimination of residual risk is always possible.

While our focus here is on parameter uncertainty, the idea of residual risk retains its meaning in the broader context of model uncertainty. Under model uncertainty, it is customary to consider a number of candidate models (families of distributions) for the loss (Cairns, 2000; Kerkhof et al., 2010; Barrieu and Scandolo, 2013; Breuer and Csiszár, 2014; Alexander and Sarabia, 2012; Boucher et al., 2014). Based on a random sample, a suitable model may be chosen using either statistical criteria (e.g. goodness of fit) or a worst-case scenario approach. Any such estimation procedure can be expressed via a capital estimator as in this paper, such that the corresponding residual risk can be quantified. An investigation of residual estimation risk in the context of model uncertainty is performed by Bignozzi and Tsanakas (2013), where a disentangling of the distinct impacts of parameter and model uncertainty is attempted.

Finally, we note that the proposed capital estimation procedures typically lead to an increase in the calculated capital requirements, compared to e.g. MLE. However, this does not mean that, if one of those procedures is followed, sufficient capital will certainly be present for each individual portfolio. The proposed capital increases are designed to be effective at an aggregate (e.g. market) level, with the outer risk measure in the definition of residual risk calculated under the true loss model. A regulatory perspective

is thus implicit in our use of a frequentist statistical framework, where the volatility of random samples may be best understood as variability in the experience of a group of economic/statistical agents.

6 Appendix

6.1 Numerical evaluation of residual risk

Here we explain how the residual risk for different estimators is obtained via Monte-Carlo simulation, using a simple importance sampling scheme. In all examples, we need to calculate the quantity $\rho(Y - \eta(\mathbf{X}))$, where ρ may be VaR_p, TVaR_p or $\mathrm{RVaR}_{p1,p2}$.

First $m = 10^7$ samples are simulated from the random variable $\eta(\mathbf{X})$; denote these as $\eta_1, \ldots, \eta_m$. (Details about simulation of $\eta(\mathbf{X})$ are given in Section 6.2.) Subsequently, λm simulated values, $y_1, \ldots, y_{\lambda m}$, are obtained from the conditional distribution $Y|Y > \mathrm{VaR}_u(Y)$ and $(1-\lambda)m$ simulated values, $y_{\lambda m+1}, \ldots, y_m$, from $Y|Y \leq \mathrm{VaR}_u(Y)$. Throughout, we use the values $\lambda = 0.9$ and $u = 0.9$. This ensures that a high fraction of simulations is obtained for high values of Y leading to more frequent exceedances of $\eta(\mathbf{X})$ by Y. For each $i = 1, \ldots, m$ set $z_i = y_i - \eta_i$. Note that it is

$$\begin{aligned}\mathbb{P}(Y - \eta(\mathbf{X}) \leq z) = (1-u)\mathbb{P}(Y - \eta(\mathbf{X}) \leq z | Y > \mathrm{VaR}_u(Y)) \\ + u\mathbb{P}(Y - \eta(\mathbf{X}) \leq z | Y \leq \mathrm{VaR}_u(Y)).\end{aligned} \tag{24}$$

Hence we can approximate the distribution of $Y - \eta(\mathbf{X})$ using the conditional empirical distributions

$$\mathbb{P}(Y - \eta(\mathbf{X}) \leq z) \approx \frac{1-u}{\lambda m} \sum_{j=1}^{\lambda m} \mathbb{I}_{\{z_j \leq z\}} + \frac{u}{(1-\lambda)m} \sum_{j=\lambda m+1}^{m} \mathbb{I}_{\{z_j \leq z\}} := \xi(z). \tag{25}$$

An estimate of $\mathrm{VaR}_s(Y - \eta(\mathbf{X}))$ at some confidence level $s \in (0,1)$ is the value z_s that makes the right-hand-side of (25) equal to s, i.e. $\xi(z_s) = s$. A numerical search is carried out (e.g. via MATLAB's `fzero` function) to obtain such z_s.

After estimating VaR at levels p, p_1, p_2, by z_p, z_{p_1}, z_{p_2} as above, the TVaR_p and RVaR_{p_1,p_2}

measures are estimated by

$$\mathrm{TVaR}_p(Y-\eta(\mathbf{X})) \approx \frac{1}{1-p}\left(\frac{1-u}{\lambda m}\sum_{j=1}^{\lambda m} z_j \mathbb{I}_{\{z_p<z_j\}} + \frac{u}{(1-\lambda)m}\sum_{j=\lambda m+1}^{m} z_j \mathbb{I}_{\{z_p<z_j\}}\right), \tag{26}$$

$$\mathrm{RVaR}_{p_1,p_2}(Y-\eta(\mathbf{X})) \approx \frac{1}{p_2-p_1}\left(\frac{1-u}{\lambda m}\sum_{j=1}^{\lambda m} z_j \mathbb{I}_{\{z_{p_1}<z_j\le z_{p_2}\}} + \frac{u}{(1-\lambda)m}\sum_{j=\lambda m+1}^{m} z_j \mathbb{I}_{\{z_{p_1}<z_j\le z_{p_2}\}}\right). \tag{27}$$

6.2 Simulation of $\eta(\mathbf{X})$

We now describe how to compute $\eta(\mathbf{X})$ for different capital estimation procedures.

General method: Since the function η is given (either explicitly or in a form that can be numerically evaluated), it is always possible to simulate m realizations of the random vector $\mathbf{X}$, $(\mathbf{x}_1,\ldots,\mathbf{x}_m)$ leading to m realizations of $\eta(\mathbf{X})$, $\eta_1=\eta(\mathbf{x}_1),\ldots,\eta_m=\eta(\mathbf{x}_m)$.

MLE: In Examples 1, 2, 3, some saving in computational time is made by exploiting the known distributions of estimators. In those examples we have $\eta(\mathbf{X}) = \rho[F(\cdot,\hat{\theta})]$, with $\hat{\theta}\sim G(\cdot;\theta)$. To obtain a sample of size m from $\eta(\mathbf{X})$ it is sufficient to simulate m random numbers, $\hat{\theta}_1,\ldots,\hat{\theta}_m$ from the distribution $G(\cdot;\theta)$ and then set $\eta_i=\rho[F(\cdot,\hat{\theta}_i)]$ for $i=1,\ldots,m$.

Bayes: For the Bayes approach implemented in Examples 6, 7,10, the predictive distribution takes a parametric form that once more depends on the MLE $\hat{\theta}\sim G(\cdot;\theta)$. Consequently m values are again sampled from $\hat{\theta}$ and m values of $\eta_{bay}(\mathbf{X})$ are subsequently obtained. In the case of the log-normal and Pareto distributions, the integrals in (12) and (14), which are functions of $\hat{\theta}$, are evaluated numerically (using MATLAB's `quadv` function).

Bootstrap: In order to compute the first order bootstrap capital estimator in Examples 4 and 5, we need to simulate from $\eta_{bs1}(\mathbf{X}) = \eta(\mathbf{X}) + r_1(\hat{\theta})$, where $\eta(\mathbf{X})$ is the MLE as above. The adjustment $r_1(\hat{\theta})$ is simulated as follows. First a sample of size m is simulated from $\hat{\theta}\sim G(\cdot;\theta)$, call this again $\hat{\theta}_1,\ldots,\hat{\theta}_m$. For for each $i=1,\ldots,m$, $r_1(\hat{\theta}_i)$ is numerically

evaluated as follows.

i) Simulate $m' = 10^4$ samples from the distributions $F(\cdot;\hat{\theta}_i)$ and $G(\cdot;\hat{\theta}_i)$; denote these respectively as y^*_{ij} and $\hat{\theta}^*_{ij}$ for $j = 1,\ldots,m'$.

ii) Evaluate $z^*_{ij} = y^*_{ij} - \rho[F(\cdot;\hat{\theta}^*_{ij})]$ for $j = 1,\ldots,m'$.

iii) Estimate $r_1(\hat{\theta}_i) \approx \rho[\hat{F}_{\mathbf{z}^*_i}]$, where $\hat{F}_{\mathbf{z}^*_i}$ is the empirical distribution of the sample $z^*_{i1},\ldots,z^*_{im'}$.

In the case of location-scale families, (equations (18), (19) and Example 9), the capital estimator takes the form $\hat{\mu}+\hat{\sigma}c$, where c is a constant that can be calculated by simulation without reference to the true parameters; hence the above bootstrap scheme is not used. Estimators in Example 11: The unadjusted capital estimator is $\eta(\mathbf{X}) = \hat{\mu}+\hat{\sigma}\mathrm{RVaR}_{p_1,p_2}[t_{\hat{\nu}}]$. The samples $(\mathbf{x}_1,\ldots,\mathbf{x}_m)$ are simulated as in the general method above. From that consequently m sets of parameter estimates $(\hat{\mu}_1,\hat{\sigma}_1,\hat{\nu}_1),\ldots,(\hat{\mu}_m,\hat{\sigma}_m,\hat{\nu}_m)$ are obtained by the estimation method of Section 6.3. The simulated values of the capital estimator then are $\hat{\mu}_i + \hat{\sigma}_i\mathrm{RVaR}_{p_1,p_2}[t_{\hat{\nu}_i}]$ for $i = 1,\ldots,m$.

To simulate the bootstrap corrected estimator η_{bs1}, the function $\zeta(\hat{\nu})$ needs to be evaluated at the m simulated values of $\hat{\nu}$ obtained as above. For each different sample size n, the function $\zeta(\nu)$ is numerically evaluated at points $\nu = 0.7, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. First, for each ν, $m = 10^6$ values of an n-dimensional vector $\mathbf{X}^*$ and a variable Y^* are simulated from a standard Student t_ν distribution, using the importance sampling algorithm of Section 6.1. For each simulated sample of $\mathbf{X}^*$, parameter estimates $(\hat{\mu}^*,\hat{\sigma}^*,\hat{\nu}^*)$ are derived and the corresponding capital is evaluated as $\eta(\mathbf{X}^*) = \hat{\mu}^* + \hat{\sigma}^*\mathrm{RVaR}[t_{\hat{\nu}^*}]$. Subsequently $\zeta(\nu)$ is estimated as the RVaR_{p_1,p_2} measure applied to the empirical distribution of $Y^* - \eta(\mathbf{X}^*)$. For intermediate values of ν, $\zeta(\nu)$ is calculated by linear interpolation. For $\nu > 10$ we let $\zeta(\nu) = \zeta(10)$ and similarly for $\nu < 0.7$ we let $\zeta(\nu) = \zeta(0.7)$. These approximations affect the quality of the bootstrap correction but not the accuracy of the results reported in the example.

For the empirical estimator η_{emp}, m samples are simulated from $\mathbf{X}$. Denote the i^{th} simulated sample $\mathbf{x}_i$. $\mathrm{VaR}_{p_1}[\tilde{F}_{\mathbf{x}_i}]$, $\mathrm{VaR}_{p_2}[\tilde{F}_{\mathbf{x}_i}]$ are calculated as in Section 6.4. Then capital in the i^{th} scenario is evaluated as the average of all x_i that are between those two percentiles:

$$\eta_{emp}(\mathbf{x}_i) = \frac{1}{\sum_{j=1}^{n} \mathbb{I}_{\{\mathrm{VaR}_{p_1}[\tilde{F}_{\mathbf{x}_i}] < x_{ij} \leq \mathrm{VaR}_{p_2}[\tilde{F}_{\mathbf{x}_i}]\}}} \sum_{j=1}^{n} x_{ij} \mathbb{I}_{\{\mathrm{VaR}_{p_1}[\tilde{F}_{\mathbf{x}_i}] < x_{ij} \leq \mathrm{VaR}_{p_2}[\tilde{F}_{\mathbf{x}_i}]\}}$$

In the case of model error, the capital estimator is $\eta_{norm}(\mathbf{X}) = \overline{X} + S \cdot \mathrm{RVaR}[\Phi]$, where $\overline{X}$ is the sample mean, S is the sample standard deviation from a simulated sample, and $\mathrm{RVaR}_{p_1,p_2}[\Phi] = \frac{\phi(\Phi^{-1}(p_1)) - \phi(\Phi^{-1}(p_2))}{p_2 - p_1}$, which can be obtained by direct integration.

6.3 Robust estimation for the 3-parameter Student t family

Robust estimates of location and scale parameters in heavy tailed models are typically percentile-based, such as the median and interquartile range (see Brys et al. (2006) for a comprehensive discussion). Such arguments are used here to derive a simple robust estimator of the parameters of a Student t distribution. The calculation of estimates is fast, which is necessary for estimation on multiple simulated samples. For a random sample $\mathbf{X}$ of size n, let each observation X_i be equal in distribution to $\mu + \sigma Z_\nu$, where Z_ν has a standard t distribution with ν degrees of freedom.

Empirical percentiles are derived using the method of Section 6.4. The location parameter estimated by the sample median, setting $\hat{\mu} = \mathrm{VaR}_{0.5}[\tilde{F}_{\mathbf{X}}]$. Consider the ratio

$$\hat{h} = \frac{\mathrm{VaR}_{0.95}[\tilde{F}_{\mathbf{X}}] - \mathrm{VaR}_{0.05}[\tilde{F}_{\mathbf{X}}]}{\mathrm{VaR}_{0.75}[\tilde{F}_{\mathbf{X}}] - \mathrm{VaR}_{0.25}[\tilde{F}_{\mathbf{X}}]}.$$

By the location-scale properties of the Student t distribution, $\hat{h}$ only depends on ν, but not on μ, σ. Consider the population version of the same quantity, that is, the function

$$h(\nu) = \frac{t_\nu^{-1}(0.95) - t_\nu^{-1}(0.05)}{t_\nu^{-1}(0.75) - t_\nu^{-1}(0.25)},$$

where t_ν^{-1} is the inverse of the standard t distribution. The degrees of freedom are then estimated by solving numerically the equation $h(\hat{\nu}) = \hat{h}$. Finally, given $\hat{\nu}$, the scale parameter can be simply estimated as

$$\hat{\sigma} = \frac{\mathrm{VaR}_{0.75}[\tilde{F}_{\mathbf{X}}] - \mathrm{VaR}_{0.25}[\tilde{F}_{\mathbf{X}}]}{t_{\hat{\nu}}^{-1}(0.75) - t_{\hat{\nu}}^{-1}(0.25)}.$$

6.4 Calculation of empirical percentiles

Consider a random sample $\mathbf{X}$ of size n. There are different ways of estimating percentiles directly from the sample. Here we describe the method implemented in MATLAB (via the `quantile` function). The empirical distribution of the sample is the step-function $\hat{F}_{\mathbf{X}}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i \leq x\}}$. Let $X_{1:n}, ..., X_{n:n}$ be increasing order statistics. The distribution function $\tilde{F}_{\mathbf{X}}$ is defined by assigning values

$$\tilde{F}_{\mathbf{X}}(X_{k:n}) = \hat{F}_{\mathbf{X}}\left(\frac{X_{k:n} - X_{(k-1):n}}{2}\right),$$

and then performing linear interpolation to obtain $\tilde{F}_{\mathbf{X}}(x)$ for intermediate values x. Then a percentile is estimated by inverting the the piecewise linear function $\tilde{F}_{\mathbf{X}}$, i.e. solving $\tilde{F}_{\mathbf{X}}(x) = p$ for some $p \in (0, 1)$. Therefore, an empirical estimator for VaR_p can be expressed as $\mathrm{VaR}_p[\tilde{F}_{\mathbf{X}}]$.

6.5 RVaR for the Student t distribution

For standard a Student t random variable Z with distribution t_ν and density g_ν, by direct integration:

$$\begin{aligned}
\mathrm{RVaR}_{p_1,p_2}(Z) &= \frac{1}{p_2 - p_1}\int_{t_\nu^{-1}(p_1)}^{t_\nu^{-1}(p_2)} z g_\nu(z)dz = \frac{1}{p_2 - p_1}\int_{t_\nu^{-1}(p_1)}^{t_\nu^{-1}(p_2)} z\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}} dz \\
&= \frac{1}{p_2 - p_1}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\frac{\nu}{2}\left[\frac{(1 + z^2/\nu)^{-(\nu+1)/2+1}}{-(\nu+1)/2 + 1}\right]_{t_\nu^{-1}(p_1)}^{t_\nu^{-1}(p_2)} \\
&= \frac{1}{p_2 - p_1}\frac{1}{1-\nu}\left(g_\nu(t_\nu^{-1}(p_2))(\nu + (t_\nu^{-1}(p_2))^2) - g_\nu(t_\nu^{-1}(p_1))(\nu + (t_\nu^{-1}(p_1))^2)\right).
\end{aligned}$$

References

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

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

O. E. Barndorff-Nielsen and D. R. Cox. Prediction and asymptotics. *Bernoulli*, 2(4): 319–340, 1996.

P. Barrieu and G. Scandolo. Assessing financial model risk. *Preprint*, 2013.

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

V. Bignozzi and A. Tsanakas. Model uncertainty in risk capital measurement. *Preprint*, 2013.

J. M. Borowicz and J. P. Norman. 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*, 2008.

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

T. Breuer and I. Csiszár. Model uncertainty in risk capital measurement. *Mathematical Finance*, 2014. To appear.

G. Brys, M. Hubert, and A. Struyf. Robust measures of tail weight. *Computational Statistics and Data Analysis*, 50(3):733–759, 2006.

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

A. J. G. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4): 687–718, 2006.

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

G. S. Datta, R. Mukerjee, M. Ghosh, and T. J. Sweeting. Bayesian prediction with approximate frequentist validity. *Annals of Statistics*, 28:1414–1426, 2000.

M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. Actuarial theory for depedent risks. *Wiley*, 2005.

K. Dowd and D. Blake. After VaR: The theory, estimation, and insurance applications

of quantile-based risk measures. *Journal of Risk and Insurance*, 73(2):193–229, 2006.

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

T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer, 1992.

R. V. Hogg, J. W. McKean, and A. T. Craig. *Introduction to Mathematical Statistics*. Pearson, 7th edition, 2012.

P. Jorion. Risk^2: Measuring the risk in value at risk. *Financial Analysts Journal*, 52(6): 47–56, 1996.

J. Kerkhof, B. Melenberg, and H. Schumacher. Model risk and capital reserves. *Journal of Banking and Finance*, 34(1):267–279, 2010.

Z. Landsman and A. Tsanakas. Parameter uncertainty in exponential family tail estimation. *ASTIN Bulletin*, 42(1):123–152, 2012.

A. Mata. Parameter uncertainty for extreme value distributions. *Presented at 2000 General Insurance Convention, 25-28 October*, 2000.

A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, 2005.

J. Nešlehová, P. Embrechts, and V. Chavez-Demoulin. Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, 1:3–25, 2006.

T. A. Severini, R. Mukerjee, and M. Ghosh. On exact probability matching property of right-invariant priors. *Biometrika*, 89(4):952–957, 2002.

R. L. Smith. Bayesian and frequentist approaches to parametric predictive inference. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, volume 6. Oxford University Press, 1999.

E. C. Venezian. Insurer capital needs under parameter uncertainty. *Journal of Risk and Insurance*, 50(1):19–32, 1983.

R. J. Verrall and P. D. England. Predictive distributions of outstanding liabilities in

general insurance. *Annals of Actuarial Science*, 1(2):221–270, 2006.

S. S. Wang, V. R. Young, and H. H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183, 1997.

J. Ziegel. Coherence and elicitability. *Mathematical Finance*, 2014. To appear.