



A Credibility Framework for Extreme Value-at-Risk

Peter Mitic

Received: 29 June 2023 / Revised: 15 February 2024 / Accepted: 15 March 2024 / Published online: 21 May 2024
© The Author(s) 2024

Abstract Value-at-risk estimates derived from extreme value data by fitting fat-tailed distributions can be so large that their validity is open to question. In this paper, an objective criterion, and a framework from which it was developed, are presented in order to decide whether or not a fitted distribution is inappropriate for the purpose of value-at-risk calculation. That criterion is based on established extreme value theory (principally the Pickands-Balkema-deHaan Theorem), which is used to calculate a sequence of reference value-at-risk estimates using Generalised Pareto distributions. Those estimates are used to develop a closed-form formula for calculating a theoretical 'maximum' value-at-risk. The method is validated by generating 100 random data sets and testing them against the framework for varying input parameter values. Approximately 75% of those cases passed the validation test.

Keywords Value at risk · Generalised Pareto · Gauss surface curvature · Mean surface curvature · Pareto–Pickands · Extreme value distributions · GPD-surface

1 Introduction

Every year, regulated financial institutions (banks, insurance companies, financial advisors etc.) must assess the amount of retained capital to cover operational risk losses that might be suffered in the following year. The European Banking Authority [7] defines operational risk as "the risk of losses stemming from inadequate or failed internal processes, people and systems or from external events". Informally, operational risk is the risk of "things going wrong". Such capital must be retained by the bank and cannot be used for lending. Consequently, operational risk capital must be sufficient to cover expected and anticipated losses, but should not be over-estimated substantially. Calculating a satisfactory amount is an important part of a bank's risk control. This paper concerns a particular problem that often arises when calculating operational risk capital. There are cases when the result of a capital calculation is judged subjectively to be 'too large'. These cases arise when the Operational Risk loss data contains *very* extreme values, and alternative calculations have to be done. There is currently no *objective* way to decide whether or not the result of a capital calculation is 'too large', and the purpose of this paper is to present an objective decision process.

P. Mitic (✉)
Computer Science, University College London, Gower Street, London WC1E 6B, UK
e-mail: p.mitic@ucl.ac.uk

1.1 Structure of this Paper

The introduction is followed by a summary of the characteristics of operational risk and problems associated with measuring it. The literature review (Sect. 3) covers prior research on modelling probability distributions that are appropriate to operational risk data.

In the main body of this paper (Sect. 4), a framework is presented to assess whether or not a *Value-at-Risk* calculation for any appropriate data set is "excessive". An objective method to make that assessment is proposed. Validation results are presented in Sect. 5, and the framework is evaluated in Sect. 6.

1.2 Nomenclature

The following terms are used extensively throughout this paper.

1. *OpRisk* is an abbreviation for Operational Risk.
2. 'Operational Risk Data' (usually abbreviated to 'data' in this paper) comprising sets of 'fat-tailed' data, each with a date stamp.
3. *VaR* is the usual abbreviation for *Value-at-Risk*. In this paper, *VaR* refers to the particular instance of *Value-at-Risk* specified by international financial regulators, which is *Value-at-Risk at 99.9%, with a 1-year time horizon*.
4. 'Loss' ('Losses') refers to payments made to customers/clients in respect of Operational Risk events.
5. 'Tail' refers to the subset of all losses, comprising the largest losses, in our case determined by a percentage.
6. 'Body' is the set complement of the Tail losses.
7. *GPD* is the Generalised Pareto Distribution with location, scale and shape parameters μ, σ, ξ respectively
8. 'Annual frequency' is the number of elements in a data set divided by the number of years spanned by the data.
9. 'Distribution *VaR*' refers to *VaR* calculated for all data (i.e. not restricted to the body or tail)
10. 'GoF' means *Goodness-of-Fit* in the context of significance testing

Closely connected are combinations of the above terms which refer to 'loss' or *VaR* applied to a distribution tail, body.

2 Operational Risk and Value-at-Risk

The Basel Committee on Banking Supervision (the "Basel Committee") is the source of regulations that govern the management of operational risk, and has approved a calculation method known as the *Advanced Measurement Approach (AMA)* [2]. Under *AMA* regulations, banks can implement their own risk models, subject to broad principles specified by the national financial regulators. The capital calculation is usually done by fitting an appropriate probability distribution to data, and calculating a standard risk metric, *VaR*. *VaR* can be thought of, informally, as the 'largest loss that could be tolerated'. Formally, it measures a maximum amount of money that could be lost over a given time horizon, given a probability of loss [15]. The probability set by international financial regulations [2] is 0.001 (usually expressed as '99.9% confidence'). Operational Risk data sets are usually modelled by 'fat-tailed' distributions, which are polynomial-like rather than exponential-like for large losses. A common model for calculating *VaR* is known as the *Loss Distribution Approach* [9], abbreviated here to *LDA*. The *LDA* incorporates a convolution of two statistical models, one for loss frequency (mostly Poisson or Negative Binomial are used), and the other for loss severity.

Table 1 illustrates the problem we are trying to solve. It shows best-fit results with corresponding *VaR* values derived by fitting the distributions shown, all of which are viable candidates. The *VaR* values vary enormously, from very small to huge. The *LogNormal Mixture* distribution gives a *VaR* which exceeds the dollar figure for the gross domestic product of the world (84.75 trillion USD in 2020 - <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>)! These bizarre results are surprisingly common. The result of analysis in Mitic [20] is that a *VaR* value should

Table 1 Operational Risk p-values and *Value-at-Risk* for candidate fat-tailed distributions. Only one is a Goodness-of-fit fail (p-value > 0.05), and is indicated in bold

| Distribution | p-value | Value-at-Risk (mEUR) |
|-------------------------|--------------|----------------------|
| LogNormal Mixture | 0.018 | 18620000 |
| LogNormal | 0.047 | 10.2 |
| Weibull | 0.026 | 2.8 |
| LogGamma | 0.012 | 283.0 |
| Burr | 0.004 | 3893 |
| Generalised Pareto | 0.006 | 1862 |
| G-and-H | 0.002 | 30.5 |
| Gamma | 0.144 | 97.8 |
| LogNormal Gamma Mixture | 0.047 | 10.2 |
| LogLogistic | 0.005 | 972.8 |
| Frechet | 0.007 | 11505 |
| Gumbel | 0.013 | 115.7 |

reflect the overall 'size' of the data. Specifically, it should be not more than $7\frac{1}{3}$ times the annualised data sum. For the data set used in Table 1, the data sum was 2144 mEUR, and data spanned 10.5 years. The upper limit for its *VaR* should then be approximately 1500 mEUR, which eliminates the highest estimates in Table 1.

There are two principal reasons for the discrepancies noted in Table 1: distribution parameter estimation, and the gradient of the derived distribution CDF for large losses. In the former case, parameter estimation can be difficult in cases where there is little variation in loss severity within a large subset of losses. In those circumstances, there can be little change in a maximum likelihood value even if a target parameter changes by a relatively large value. The maximum likelihood estimation then terminates because it has reached a specified maximum number of iterations, and not because it has converged. In particular, estimation of the ξ parameter for Generalised Pareto, Gumbel, Frechet distributions is subject to this problem. Usually the Hill estimator is a reliable way to estimate the ξ -value, provided that there is sufficient data. In the second case, the CDF gradient for some distributions is very flat for large data values. The CDF then attains the required 0.999 value at extremely large data values. A particular case is the Generalised Pareto distribution when ξ is greater than 1. A potential solution to this problem is to reject values in a sample that exceed the maximum observed datum by some predetermined amount. For example, a factor of 10 times the maximum observed datum is one possibility. Without investigation, it is hard to tell which effect operates in any particular case. Parameter estimation for some distributions is known to be reliably convergent (LogNormal, Weibull), whereas for others (Generalised Pareto) it is not. Similarly, the CDF gradient for large loss is known to be near to zero for distributions such as Burr and Generalised Pareto. Further investigation on this topic is merited.

2.1 Theoretical Background: Tail *VaR*

A distribution tail is generally a principal determinant of *VaR*. Furthermore, the *Pickands-Balkema-deHaan* Theorem [25] shows that the tail distribution approaches a *GPD* as the tail size decreases. Fitting a *GPD* to tail data is a fundamental part of our framework for assessing whether or not a *VaR* calculation is "excessive". The basis of the *Pickands-Balkema-deHaan* Theorem is the *Excess* distribution. Given a high-value threshold u , $F_u(x)$ is the conditional distribution where a random variable X exceeds $x + u$, given that X exceeds u . That is:

$$F_u(x) = P(X > u + x | X > u) \quad (1)$$

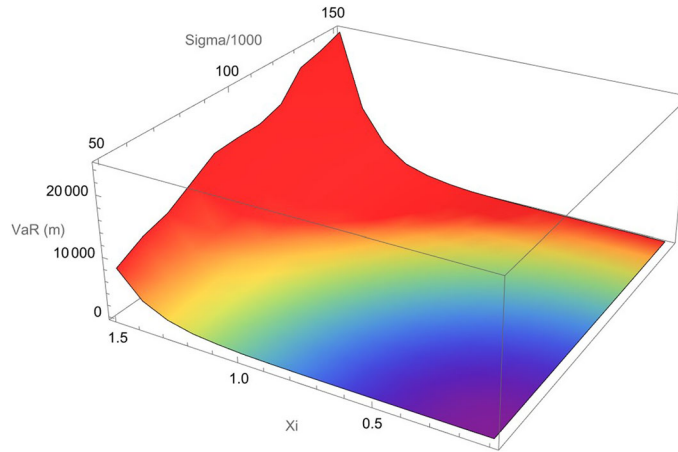


Fig. 1 Typical *GPD Surface*: $\sigma \in [50000, 150000]$, $\xi \in [0, 1.5]$, $\nu = 4$

The *Pickands-Balkema-deHaan* Theorem states that the distribution $F_u(x)$ is approximated by the *GPD*, with the density and distribution functions f and F respectively in Eq. 2. A *GPD* is characterised by 3 parameters: location (μ), scale (σ) and shape (ξ). Of these, ξ is the principal determiner of *VaR*.

$$\begin{aligned}
 f(x : \mu, \sigma, \xi) &= \left(\frac{1}{\sigma}\right) \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1-\frac{1}{\xi}} \quad x \geq \mu, \sigma > 0, \xi > 0 \\
 F(x : \mu, \sigma, \xi) &= 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}} \quad x \geq \mu, \sigma > 0, \xi > 0
 \end{aligned} \tag{2}$$

2.2 Proposed Solution

We assert that maximum appropriate *VAR* may be determined using the tail of a data distribution, modelled by a *GPD*. The proposed solution is to implement a framework that introduces two concepts: a *GPD Surface* and a *Credible Region*. A *GPD Surface* is a 3-D plot of *VaR*, for a given frequency parameter ν , derived from a *GPD* on the vertical axis, with independent *GPD* variables ξ and σ on the horizontal axes. Each *GPD Surface* has a 'near-horizontal' portion when ξ is small, with a marked 'upturn' for larger ξ . Examples are shown in Sect. 5.2. The *Credible Region* is a subset of the 'near-horizontal' portion, and its boundary partitions *VaR* values that are subjectively "too big" from those that are not. The framework comprises the following steps for each value of the frequency parameter ν .

- Assume that maximal distribution *VaR* can be estimated using tail *VaR*.
- Model tail *VaR* using a *GPD* at frequency ν .
- Find a best-fit distribution, F , for the entire data set.
- Compare the calculated *VaR* using F to a reference set of data-independent pre-calculated *GPD Surface* values derived from a *GPD* distribution.
- Formulate a decision rule: "If the distribution *VaR* exceeds the *GPD*-derived *VaR*, reject F and seek an alternative (non-optimal) fit. Otherwise, accept F ."

These steps will be explained in the Methodology section, 4. To give a clearer idea of the two essential concepts in the framework, Fig. 1 shows a typical *VaR Surface*. The *Credible Region* corresponds approximately to the purple-blue "flat" part of the surface. The sharp surface curvature for large ξ is also apparent. The orange-red part of the surface represents *VaR* values that are not "credible".

3 Literature Review

The literature review concentrates on the parts of *Extreme Value Theory (EVT)* that are relevant to *GPD* tail loss modelling. There were two principal approaches to modelling the extremes of a set of losses in the mid-20th century. The development of the second depends on the first, and both are fundamental to the study of distribution tails.

The *Block Maxima* approach [13] leads to the *Generalised Exponential Distribution (GEV)*, and its three subtypes, *Frechet*, *Gumbel* and *Weibull*. It is so-called because it relates to distributional properties of the maximum loss in a set of partitions of the data. The groundwork was established earlier. In 1923, [6] calculated an expression for the median of the tail distribution. Also in 1923, [17] calculated an expression for the expected value of the tail distribution. In 1927, [10] advanced the theory with a study of the asymptotic distribution of the tail. A year later, [8] showed that extreme limit distributions of the tail can only be one of the three types. That result was not proved rigorously until 1943 [11]. The final result of mid-20th century work has become known as the *Extremal Types Theorem*.

A formal distribution function for the *Extreme Value Distribution (EVD)* was given by [14], following an earlier basis [18]. [5] (Sect. 3.1.4) provides an outline proof of the form of an *EVD*.

The second approach, *Threshold Exceedances*, extends the research described above, and leads to the link between the *GPD* and tail data. The essential formulation was done in the mid-1970s [25], [1]. The basis of the advance was to derive the distributional form of Eq. 2 from the *Excess* formulation in Eq. 1.

An alternative approach was to characterise convergence in terms of moments. [4] suggested using the distribution mean and standard deviation as loss scaling constants. Validity conditions were given in [24]. [5] (Sect. 4.2.2) provides an outline proof of convergence, and a formal proof is given by [16]. Coles also has many further references to the development of *EVT*, and also to its early applications. [27] gives a similar overview, with details on parameter estimation.

3.1 Recent Advances

Before 2023, no direct attempts to characterise *VaR* in terms of the data from which it was calculated had been made. In 2023, two such studies were published. In Mitic [22], an empirical relationship between *VaR* and the annual loss sum, S , was established. The relationship was expressed in a very simple linear formula: $VaR \sim \frac{S}{2}$. No indication of a *VaR* ceiling is suggested.

An alternative way of estimating a *VaR* ceiling was used in Mitic [21]. The median of the distribution of the *Maximum Order Statistic* of a *GPD* fit for tail losses is associated with a *VaR* ceiling via a linear scale factor. Overall, it is easier to apply than the method proposed in this paper, at the expense of a slightly impaired success rate. A third result, Mitic [20] presents a further simple formula to determine whether or not a *VaR* value is "excessive": $VaR \sim 7\frac{1}{3}S$. This result has the considerable advantage of simplicity, but also at the expense of an impaired success rate.

4 Methodology

Our 'solution' to the problem of answering the question "is the calculated *VaR* too big?" is to compare the calculated *VaR* to a reference set of data-independent pre-calculated *VaR* values derived from a *GPD* distribution. The latter set is termed a *GPD Surface*. If the calculated *VaR* exceeds the *GPD*-derived *VaR*, the distribution used to calculate *VaR* is rejected. Otherwise it is accepted. In order to avoid excessive time-consuming calculations, we propose a framework of pre-prepared *GPD surfaces*, each linked to a given annual loss frequency. An empirically-calculated *VaR* value can be compared with a *GPD surface* (which may have to be derived by interpolation) with the same annual frequency. In doing so, the following assumptions are made.

Assumptions

1. Tail *VaR* can be calculated using a *GPD*.
2. Tail *VaR* is representative of *VaR* for an entire data set (i.e. body *VaR* is small compared to tail *VaR*). An outline proof of this assumption for large data sets is given in Appendix A. It depends on assumptions that the mean and standard deviation of the body are smaller than the mean and standard deviation of the tail.
3. There is a boundary, such that if *VaR* exceeds the boundary value, the distribution that was used to calculate the *VaR* value should be rejected in favour of an alternative distribution that gives a *VaR* value that does not exceed the boundary.
4. The boundary may be approximated as a quadratic
5. An appropriate model that expresses *VaR* in terms of the *GPD* parameters $\{\mu, \sigma, \xi\}$ can be fitted to empirical data.

4.1 The *GPD Surface*

We first define the essential terms that underpin the framework that follows.

Definition: *GPD Surface*

A *GPD Surface* is a mapping from a set of n pairs in a region defined in terms of a quadratic function $\sigma = Q(\xi)$ by $P_{\xi}^{\sigma} = \{\sigma_i, \xi_i \mid \xi_i > 0, \sigma_i > 0, \sigma_i \leq Q(\xi_i)\}; i = 1..n$ and constants $\{\mu, \nu\}$ to a set of *VaR* values V_i calculated using *GPD* parameters $\{\mu, \sigma, \xi\}$, with an annual frequency ν . We denote it by Γ .

$$\Gamma(\mu, \sigma, \xi, \nu) = \left\{ \{\sigma_i, \xi_i, V_i\} \mid \mu, \nu \right\} \quad \sigma_i, \xi_i \in P_{\xi}^{\sigma}; \quad i = 1..n \quad (3)$$

This definition therefore extends a *GPD* to a "pseudo distribution" which has four parameters: the three *GPD* parameters, plus a frequency parameter. The frequency dependence arises because *VaR* calculations depend on the data time span. For convenience, we sometimes shorten the notation to $\Gamma(\nu)$ if an interpolation on ν is involved. See Algorithm: *GPD-CRED*.

Definition: *Credible Region*

A *Credible Region* is a subset of a *GPD Surface*, such that the *VaR* values that fall within that region satisfy an acceptance criterion that limits the values of σ and ξ on the *GPD Surface* to σ_C and ξ_C respectively. We denote it by $\Gamma_C(\mu, \nu)$. Note that the term *Credible Region* does not refer to a term with the same name in Bayesian analysis.

$$\Gamma_C(\mu, \sigma, \xi, \nu) = \left\{ \Gamma(\mu, \sigma, \xi, \nu) \mid \mu, \nu, \sigma \in \sigma_C, \xi \in \xi_C \right\} \quad (4)$$

The precise way to define the *Credible Region* will be discussed in Sect. 4.6. In parallel with the abbreviated notation $\Gamma(\nu)$ for a *GPD Surface*, we also use the abbreviated notation $\Gamma_C(\nu)$ for a *Credible Region*.

4.2 Overall Strategy

The overall strategy is to associate fitted *GPD* tail parameters with a point on a *GPD Surface* that is appropriate for the annual frequency of the tail. The *VaR* derived from a best fit distribution to all the data (not just the tail) can then be compared with a theoretical *GPD*-derived *VaR*. The detailed steps are listed below. The actual comparison is done by a calculation of surface curvature. A marked discontinuity in curvature demarcates a region in which a *VaR* value is of acceptable size from one in which it is not. A marked discontinuity is not so apparent on the *GPD-Surfaces*, except at very high ξ values.

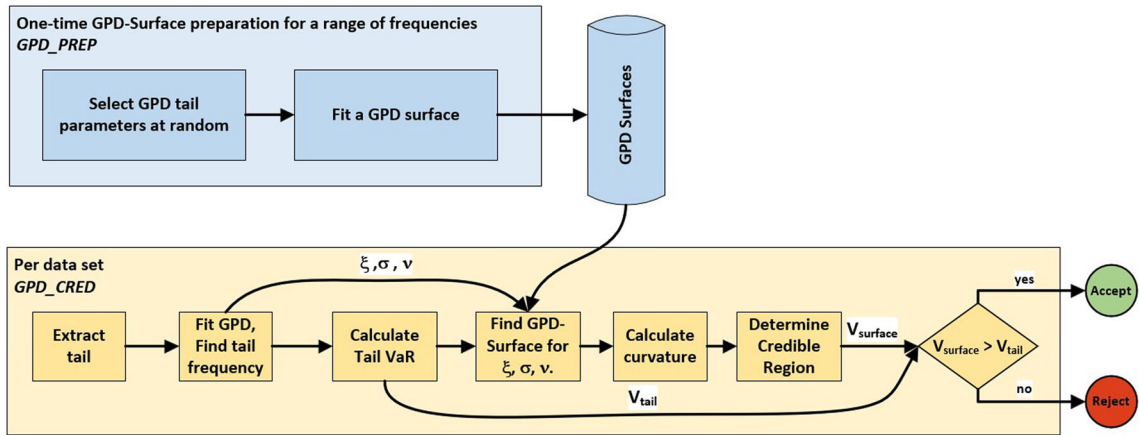


Fig. 2 Process flow: One-time surface preparation stage *GPD-PREP* (in blue), and decision stage *GPD-CRED* (in yellow)

There are two parts to the overall calculation. The first (Algorithm *GPD-PREP*) needs to be done once only, is not data-dependent, and is reused for every data set. In the second (Algorithm *GPD-CRED*), tail *VaR* calculated from data is compared with the *VaR* derived from a *GPD Surface* and its *Credible Region*, and a credibility calculation is applied to decide whether or not the *VaR* calculated from data is "too big". An overview of the process is shown in Fig. 2. Each stage is explained in further parts of this section.

Algorithm: GPD-PREP

The steps below are data-independent, and apply for a single annual frequency ν .

1. Set a value for μ which must be less than or equal to any minimum loss in a data set. $\mu = 1000$ is suggested for tail data.
2. Define a range of σ and ξ values that are appropriate for a distribution's *GPD* tail. We suggest $\xi \in (0, 2.5)$ and $\sigma \in (5 \times 10^6, 50 \times 10^6)$
3. With the frequency ν , for each *GPD* parameter combination σ_i, ξ_i , calculate *VaR* V_i using the fixed value of μ and a Poisson frequency model with 1 million simulations.
4. Fit a 2-D surface to the triples σ_i, ξ_i, V_i , with the pair σ_i, ξ_i as independent variables, and the set V_i as the dependent variable.

The result of applying the steps above is a *GPD Surface* conditioned on an annual frequency ν . The algorithm is applied to multiple annual frequencies $\nu = \nu_1, \nu_2, \nu_3, \dots$, appropriate for the distribution tails commonly encountered. Annual frequencies in the range 1 to 25 are suggested. In this way, a set of *GPD Surfaces*, each conditioned on an annual frequency, is defined. They serve as reusable reference sets, against which *VaR* calculated from data can be tested. *GPD Surfaces* at other frequencies are derived by interpolation or extrapolation.

Algorithm: GPD-CRED

In Algorithm *GPD-CRED*, the *GPD Surfaces* are used in conjunction with fat-tailed data sets to examine the suitability of data fits for the purpose of *VaR* calculations. The steps below form the required framework.

1. Extract the data tail and fit a *GPD* to it (giving parameters μ, σ and ξ), using a Poisson frequency model and 1 million simulations
2. Determine the annual *data* frequency ν
3. Locate a *GPD Surface*, $\Gamma(\nu)$ either by selecting an existing one, or by interpolation (with respect to frequency) on the set of *GPD Surfaces* from Algorithm *GPD-PREP*
4. Calculate surface curvature *Gamma* using σ, ξ and ν
5. Determine a *Credible Region* $\Gamma_C(\nu)$ from Γ
6. Determine boundary values σ_B and ξ_B for σ and ξ respectively, using $\Gamma_C(\nu)$ and parameters σ and ξ

7. Compare σ_B with σ , and ξ_B with ξ in a binary decision process, the outcome of which is to either accept the VaR calculated from data, or reject it.

In Sect. 4.8, a method to validate the decision using distributional statistics of the data is presented. This step is not strictly needed in Algorithm *GPD-CRED*, but provides a measure of the probability that the decision is correct.

4.3 The *GPD Surface* Fit

Empirical evidence (see Sect. 5.3) shows that VaR increases exponentially with the *GPD* ξ parameter. The VaR dependence on the *GPD* σ parameter is essentially linear for a distribution body, but is better modelled by an exponential for the tail. Therefore, we propose the functional form, denoted by $\hat{\Gamma}$, for a tail *GPD Surface* Γ . It comprises an exponential expression representing a plane, modified by a linear term in σ . In Eq. 5, μ and ν are held constant, with σ and ν as variables. Parameters a , b and c are to be determined by a non-linear fit.

$$\hat{\Gamma}(\mu, \sigma, \xi, \nu, a, b, c) = c\xi e^{(a\xi + b\sigma)} \quad | \mu, \nu \quad \{a, b, c \in \mathbb{R}\} \quad (5)$$

If *GPD* parameter values that are appropriate to all of the data (i.e. not just the tail) are selected to define a *GPD Surface*, a more complex form for the *GPD Surface* is needed (Eq. 6). The 'all data' fit, $\tilde{\Gamma}$, incorporates the following components (determined empirically):

- VaR varies approximately exponentially with ξ
- VaR varies approximately linearly with σ
- A 'square root of sigma' modifier to the exponential term to improve the fit
- A 'power of ξ ' additive term to improve the fit further

$$\tilde{\Gamma}(\mu, \sigma, \xi, \nu, a, b, c, d, n) = a\sigma + d\xi^n + c\sqrt{\sigma}\xi e^{b\xi} \quad | \mu, \nu \quad \{a, b, c, d, n \in \mathbb{R}\} \quad (6)$$

In practice, surface fits tend to underestimate VaR . Despite this shortcoming, they do pass goodness-of-fit tests, and serve the purpose of validating an empirical VaR value. We have noted that the VaR surfaces are very smooth, despite the stochastic calculation involved. That is, the numeric values of the partial derivatives of VaR with respect to parameters σ and ξ are small. Therefore we do not expect instability in the surface fit process. Nor do we expect convergence to a non-optimal solution.

4.3.1 Adjustment for small ξ and σ

The fit from Eq. 6 is relatively poor if σ and ξ are 'small' ($0 < \sigma \leq 25, 0 < \xi < 0.5$). An improvement is to approximate the exponential part in Eq. 5 to obtain a simpler form $c\xi(a\xi + b\sigma)$. The fit is greatly improved if an addition additive linear term in σ is included. The following is applicable for a fit to such a restricted region.

$$\hat{\Gamma}(\mu, \sigma, \xi, \nu, a, b, c, d) = c\xi(a\xi + b\sigma) + d\sigma \quad | \mu, \nu, \quad \{a, b, c, d \in \mathbb{R}\} \quad (7)$$

See 5.2 for illustrations of fits in these cases.

4.4 Goodness-of-Fit Tests

Two *GoF* tests are used. To test distributional fits to data, the *TNA*-test ([19]) is effective for all distributions concerned. This test is a formalisation of a *Q-Q* plot: a comparison of the empirical quantiles of a distribution with

the quantiles of a fitted instance of the distribution. It is robust with respect to both small and large populations. The latter requirement is particularly important, since most data sets used in this analysis are too large for a *GoF* test such as *Kolmogorov-Smirnov* or *Anderson-Darling* to be reliable if all data are used. Appendix B shows an outline of how it operates.

To test equality of surface fits, a χ^2 test is used, using an empirical array of k 'observed' coordinate pairs $\{p_1, p_2, \dots, p_k\}$, and corresponding fitted values $\{\bar{p}_1, \bar{p}_2, \dots, \bar{p}_k\}$ (the 'expected' values).

4.5 Curvature Calculations

The *GPD surface* are shown in Fig. 1 indicates that there is a change in curvature throughout, but that it is very marked at certain values of ξ and σ . Therefore, to try to find a boundary, beyond which the *VaR* is deemed to be "inadmissible", we look for changes in curvature on the *GPD* surface. We concentrate on *Gauss Curvature* and *Mean Curvature*. Those curvature metrics apply to a surface in which the dependent variable ϕ (in our case *VaR*) is a function of ξ and σ . In the case of *GPD*, the *GPD*-parameter μ is fixed, as is the reference to a fixed frequency v . The sticking point with this approach is that the functional form of $VaR(\sigma, \xi)$ must be known.

Gauss Curvature (K_G) and *Mean Curvature* (K_M) are defined as in equations 8. Subscripts denote partial derivatives. A full discussion of their derivation may be found in, for example, Sochi [28] or Gray et al [12]. The latter has notes on implementation in *Mathematica*.

$$K_G = \frac{\hat{\Gamma}_{\sigma\sigma}\hat{\Gamma}_{\xi\xi} - \hat{\Gamma}_{\sigma\xi}^2}{(1 + \hat{\Gamma}_{\sigma}^2 + \hat{\Gamma}_{\xi}^2)^2} \quad (8)$$

$$K_M = \frac{\hat{\Gamma}_{\sigma\sigma}(1 + \hat{\Gamma}_{\xi}^2) + \hat{\Gamma}_{\xi\xi}(1 + \hat{\Gamma}_{\sigma}^2) - 2\hat{\Gamma}_{\sigma}\hat{\Gamma}_{\xi}\hat{\Gamma}_{\sigma\xi}}{2(1 + \hat{\Gamma}_{\sigma}^2 + \hat{\Gamma}_{\xi}^2)^{3/2}} \quad (9)$$

We have concentrated on the use of *Gauss Curvature*, as a map of $K_G(\sigma, \xi)$ for the *GPD*-Surfaces in this analysis corresponds more closely to the perceived *Credible Regions*. An advantage of using *Mathematica* is that derivatives of any function ϕ can be formulated dynamically for any functional form for ϕ .

4.6 Credible Region

The credible region is the framework core. It is defined in terms of a contour plot of *Gauss Curvature* of a *GPD-Surface* defined by tail data only, with ξ as independent variable and σ as dependent variable. Using the tail data only makes it easy to select a contour that serves as a boundary for the credible region. There is some doubt as which contour to use if all data are used, but for tail data only, the curvature contours are very tightly spaced, and which to choose is not an issue. In order to select an appropriate contour, a range of contours from 3 to 25 was considered, and contour plots were produced for all surfaces. In all cases it was observed that as the number of contours increased, they became more tightly packed, particularly in the yellow-green/yellow/yellow-orange regions of the contour plot. Those regions corresponded to approximately the middle set of contours out of the total number of contours set. Therefore we anticipated the credible region boundary to correspond approximately to $\frac{1}{2}$ to $\frac{2}{3}$ of the number of contours set. We acknowledge that the choice of the number of contours is subjective. Having selected a contour, the equation of the selected contour can be calculated. The equation of the boundary contour, $\sigma = Q(\xi)$, takes the (quadratic) form shown in equation 10. This formulation was noted in Sect. 4.1.

$$\sigma = Q(\xi) = A\xi^2 + B\xi + C, \quad A, B, C \in \mathbb{R}, \quad \xi, \sigma > 0 \quad (10)$$

To formally define the credible region, we locate the ξ coordinate, ξ_0 , which is the intersection of the selected contour with the ξ axis (i.e. solve $\sigma = 0$ in equation 10). ξ_0 is increased by 10%, giving $\xi_0 \rightarrow \xi_B = 1.1\xi_0$, to allow for a more generous credible region. That change results in a corresponding σ value on what is, in effect, a virtual

contour corresponding to ξ_B : $\sigma_B = Q(\xi_B)$. This admits fewer false negatives. The credible region is then defined in Eq. 11. It corresponds to the yellow region in the example of Fig. 1.

$$\mathbf{R} = \left\{ (\sigma, \xi) : 0 < \sigma \leq Q(\xi) \mid \xi \in (0, \xi_B] \right\} \quad (11)$$

4.7 Acceptance Criterion and the Decision Process

An Acceptance Criterion is used to decide if the measured *GPD* tail parameters σ' and ξ' are 'acceptable'. If not, the distribution from which the tail was extracted is rejected. There are two parts to the Acceptance Criterion. In the first, σ' is compared with an evaluation of the *credible region* equation 10 using ξ' . For $\{\sigma', \xi'\}$ to be within the credible region, $\sigma' < S(\xi')$.

The second part of the Acceptance Criterion is to calculate *VaR* from the fitted surface $V_B = \hat{\Gamma}(\sigma', \xi')$ and compare the value obtained with a "maximum possible" *VaR*, $V_{max}=50$ billion. This value is approx twice as large as the largest seen so far on the ORX¹ database, 27.2 billion. That eliminates *VaR* values that are 'clearly' too high. The Acceptance Criterion is a conjunction of the two parts (equation 12).

$$\begin{aligned} \text{ACCEPT} & \text{ If } \left\{ \sigma' < S(\xi') \mid 0 < \xi' \leq \xi_B \right\} \wedge \left\{ V_B < V_{max} \right\}; \\ \text{REJECT} & \text{ Otherwise} \end{aligned} \quad (12)$$

4.8 Validation

Validation is done using a data-based 'sense test'. Its purpose is to test whether or not the decisions made (Sect. 4.7 are reasonable.) The data components used are (referred to the entire data set X):

- Maximum, X_{max}
- Mean, \bar{X}
- Distribution *VaR*, V_X
- Annual Frequency, ν_X

The validity test is, like the Acceptance Criterion, bipartite. In the first part, the ratio X_{max}/\bar{X} detects huge outliers and rejects them. A coefficient that determines an appropriate limit (with value 30) is conditioned on the first 50% of the data sets, such that 'clear' huge *VaRs* are rejected. In the second, part, we assert that V_X should not be greater than it would be if every draw of a random sample from a fitted distribution resulted in an order of magnitude times the sum of the maximum datum, which is expressed as $10 \times \nu_X \times X_{max}$. The validity condition is the conjunction of these two conditions.

$$\begin{aligned} \text{VALID} & \text{ If } \left\{ \frac{X_{max}}{\bar{X}} < 30 \right\} \wedge \left\{ V_X \leq 10 \nu_X X_{max} \right\}; \\ \text{INVALID} & \text{ Otherwise} \end{aligned} \quad (13)$$

In formulating this validation criterion, we considered that simplicity was paramount. Therefore we have not used more complicated methods based on approximate sums of *GPD* random variables, such as those due to Zaliapin et al [30] or van Zyl [31].

¹ The Operational Risk eXchange: an *OpRisk* data sharing organisation.

Table 2 Best fit distributions, 201 data sets

| Distribution | Count | % of Total |
|---------------------|-------|------------|
| LogNormal Mix | 53 | 26.4 |
| Frechet | 39 | 19.4 |
| LogNormal | 26 | 12.9 |
| LogNormal Gamma Mix | 18 | 9.0 |
| LogLogistic | 17 | 8.5 |
| GPD | 13 | 6.5 |
| Weibull | 10 | 5.0 |
| Gamma | 9 | 4.5 |
| Tukey G-and-H | 8 | 4.0 |
| LogGamma | 4 | 2.0 |
| Gumbel | 3 | 1.5 |
| Burr Type VII | 1 | 0.5 |

5 Results

We first discuss the data used, and then illustrate some of the graphical constructs that are central to our proposed framework. Finally, numerical results are presented,

5.1 Data

Random samples were drawn from distributions that are commonly encountered in operational risk. Having determined suitable parameter ranges, uniform random samples were generated to determine specific parameter values. The sample size was also randomly selected in the range 100 to 1000. Distributions that have no native implementation in Mathematica were implemented using methods available in *R* packages. In particular, package *GK* was used for the *G-and-H* distribution [26], and package *VGAM* for the *Frechet* and *Gumbel* distributions [29]. In *Mathematica*, the *Burr (Type VII)* distribution is known as the *Singh Maddala* distribution, and the *GPD* is known as the *Pareto Pickands* distribution.

Samples were combined in random mixtures in order to make the data more realistic. A best fit distribution for each of the the resulting 201 data sets was found, mostly using the in-built non-linear distribution fitting methods available in Mathematica, using maximum likelihood. In cases where a straightforward method to estimate initial parameters was unclear, random samples were generated within given reasonable ranges, and were tested for optimal *GoF* using the *TNA* test [19]. Random generation of parameters in this way has been shown to be a fast and efficient technique in difficult cases, such as the *Tukey G and H* distribution [3]. Table 2 shows the fitted distributions used, with the number of data sets fitted to each of them.

Each data set is tied to a nominal time window $Y = 5$ years, so that the annual frequency, ν , is the ratio of the number of elements in a data set, $\#(X)$, and the number of years.

$$\nu = \frac{\#(X)}{Y} = \frac{\#(X)}{5} \quad (14)$$

5.2 GPD-Surface Illustrations

Fig. 3 (parts (a)-(d)) shows four views of the same *GPD-Surface*. The flat areas for small ξ and σ are clear, as is the progression to the steepest point (high ξ and high σ). The thinness of the surface indicates an accurate *VaR* estimation.

Figure 4 shows a group of layered surfaces, with frequencies 1, 2 and 5. It illustrates the fact that all *GPD-Surfaces* considered in this analysis have the same shape: a largely flat area leading to an abrupt gradient change for large ξ . The layers do not intersect. The graphic is positioned to show the divergence of the surfaces when ξ and σ are both large.

5.3 GPD-Surface Data Fits

Fig. 5 shows a typical surface (in 'rainbow colours') with its best-fit parametrically-defined surface of the form $\hat{\Gamma}_i(\sigma, \xi) = ce^{(a\xi+b\sigma)}\sigma$ (in grey/black). When the grey/black regions are visible, there is an over-fit, which is desirable when assessing maximum 'acceptable' *VaR*. The goodness-of-fit test for surface equality result was $\chi^2 = 5.48$, with a p-value of 0.963. This result shows that the null hypothesis that the fitted and empirical surfaces have the same distribution is *not* rejected at the 5% level. In this case the equation of the fitted surface was $\hat{\Gamma}_i(\sigma, \xi) = 11.566e^{(8.376\xi+0.0344\sigma)}\sigma$ ($\mu = 1000$; $\nu = 10$). Fitted surfaces at other frequencies are similar, and

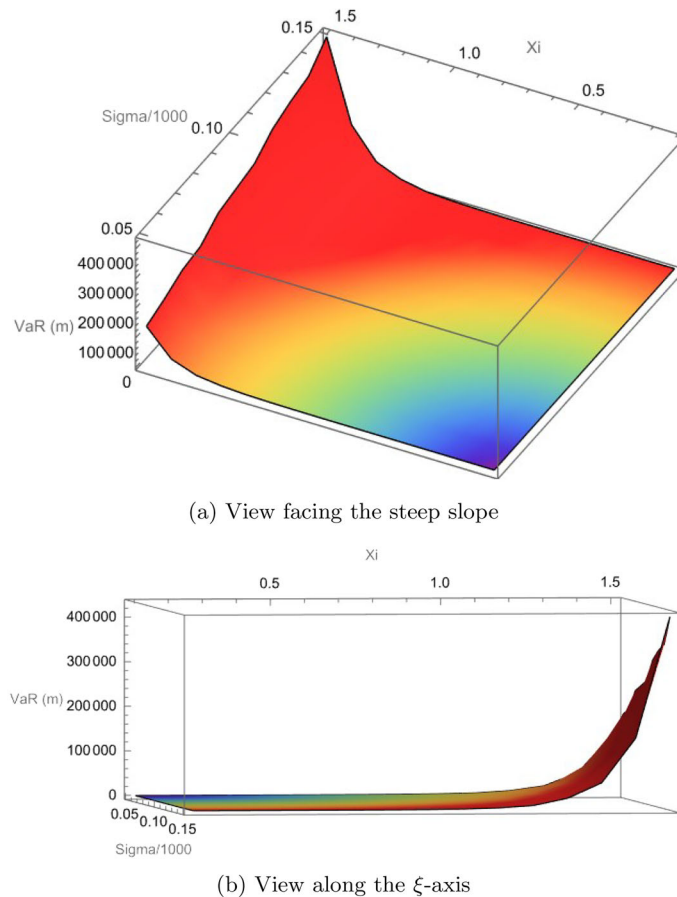
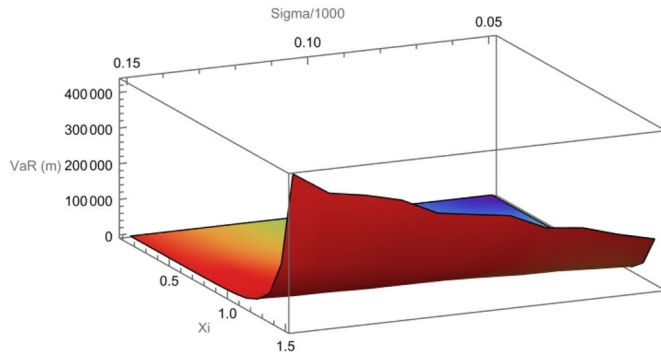
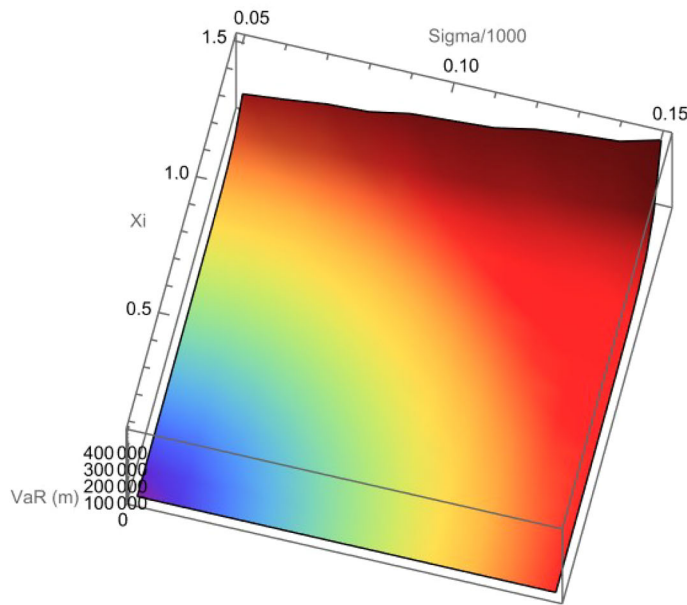


Fig. 3 Views of a *GPD-Surface*, frequency 25 (parts (a) and (b)). Views of a *GPD-Surface*, frequency 25 (parts (c) and (d))



(c) Rear view



(d) View from above

Fig. 3 continued

an ‘under-fit’ is a common feature. The goodness-of-fit conclusion (that the null hypothesis is rejected) is also the same for frequencies ranging from 1 to 40.

Figure 6 shows a close-up view of Fig. 5, restricted to a region of small ξ and small σ . See Sect. 4.3.1 for details of the fit. In this case $0 < \sigma \leq 25 \times 10^6$, $0 < \xi < 0.5$. A notable point is that the fitted surface represents greater VaR values than the corresponding empirical values. Consequently, upper VaR limits are effectively more stringent. That is not a problem, as upper bounds are high anyway.

5.4 Credible Region Results

Figure 7 (left hand) shows a contour map of curvatures for all data (i.e. body plus tail) for a typical *GPD-Surface* (in this case for frequency 10), derived using the method of Sect. 4.5. The illustration shows a gradual progression from the blue credible region (small ξ and σ) into the non-credible region (red). The blue portion of the contour map is the

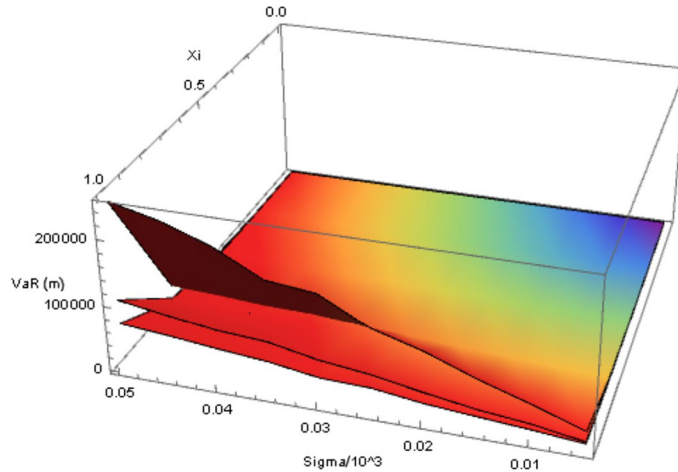


Fig. 4 *GPD Surfaces*: Frequencies 1,2,5, low values for σ and ξ . They are non-intersecting, and all have the same shape profile. The *Credible Regions* are the violet-blue-cyan portions of the surface

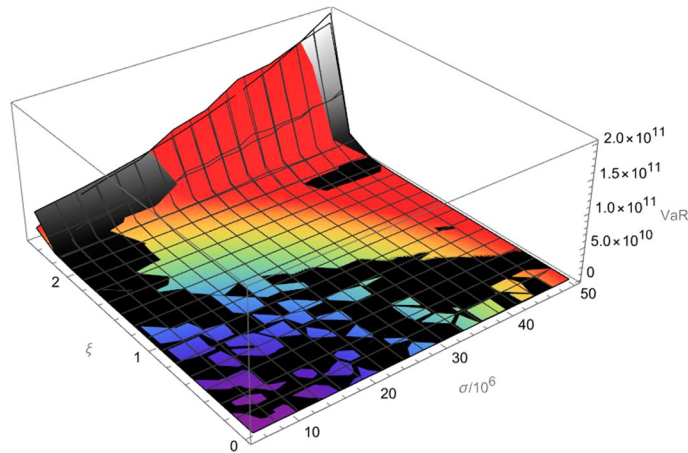


Fig. 5 Surface fit, annual frequency 10. The black/grey regions show where *VaR* for the fitted surface exceeds the actual *VaR* (shown in 'rainbow' colours)

curvature equivalent of a 'flat' region for a *GPD-Surface* such as the one in Fig. 1. Typically, the *Gauss Curvature* and *Mean Curvature* plots are almost identical. In both cases the choice of contour that could formally define the credible region (Sect. 4.6) is has to be made. We have selected the contour that intersects the $\sigma = 0$ axis at the maximum ξ -value where is contour is defined, using 6 contours. The coordinates of that contour can be extracted, and a quadratic $Q(\xi, \sigma)$ can be fitted to them. The credible region is the region bounded by $\{Q(\xi, \sigma), \sigma = 0, \xi = 0\}$. This defines a region that excludes only the red portion in Fig. 1, in order to avoid too many false negative assessments.

Figure 7 (right hand) shows the equivalent *Gauss Curvature* contour map for the tail data only. There is a marked distinction between them. The 'tail' plot has a sharp boundary for the credible region, resulting from closely-packed contour lines. Furthermore, the close-packing of the *Gauss* contours indicates that the choice of contour to define the credible region is not important in these cases. The numerical results in Sects. 5.6 and 5.7 confirm this view. It also seems reasonable to model the boundary of the credible region by a quadratic.

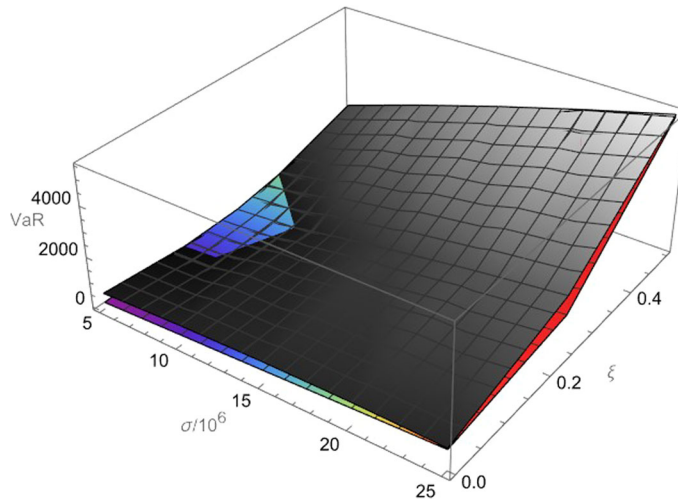


Fig. 6 Surface fit, annual frequency 10. Close-up of the surface in Fig. 5, restricted to the region $0 < \sigma \leq 25 \times 10^6, 0 < \xi < 0.5$

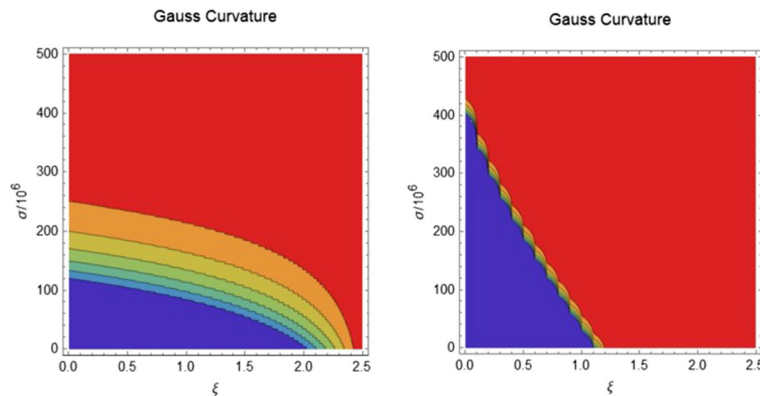


Fig. 7 Left: *Gauss* Curvature Contour plot for all data (body plus tail), for a *GPD-Surface* at frequency 10. Right: *Gauss* Curvature Contour plot restricted to tail data only, also at frequency 10. The sharp transition from the red to the blue region in the right-hand illustration is a strong indicator of the credible region boundary. Non-red parts indicate the credible region

5.5 Validation Method Results

Generally, the success rate drops as the upper limit for acceptable *VaR* increases. Overall, the success rate varies between 75% and 85%, and within that range, the dependence on the tail length is quite variable. There is an optimal range at 5–10%, and another at about 30%. The latter tail length is more suitable for smaller data sets. In all cases, variation of the success rate with the number of contours varies consistently, but in a way that shows no clear trend.

The principal validation method, a comparison of the maximum datum and mean of all empirical data (not just the tail), was designed to eliminate cases where the measured *VaR* (for all data) is 'clearly' excessive. Such cases frequently have *VaR* values in multi-billions, which is not consistent with a maximum value in the order of even a few billion.

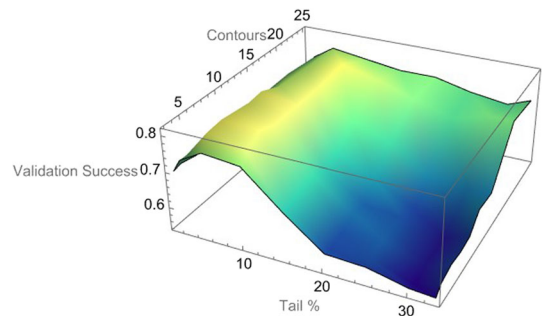
5.6 Validation Details

Overall, 48.3% of data sets were deemed by the decision procedure (Sect. 4.7) to have 'unacceptable' *VaR* values. This number is high, and is difficult to verify with data from practice between 'rejections' are not routinely recorded.

Table 3 Proportion of correct validations if decisions are based on comparison of a *GPD-Surfaces*, using all data

| Tail % | Contours | | | | | | | | |
|--------|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 3 | 4 | 6 | 8 | 10 | 12 | 15 | 20 | 25 |
| 1 | 0.7 | 0.705 | 0.695 | 0.695 | 0.695 | 0.695 | 0.695 | 0.695 | 0.685 |
| 2 | 0.735 | 0.735 | 0.74 | 0.75 | 0.74 | 0.74 | 0.74 | 0.745 | 0.735 |
| 5 | 0.77 | 0.81 | 0.815 | 0.81 | 0.805 | 0.81 | 0.785 | 0.78 | 0.775 |
| 10 | 0.76 | 0.785 | 0.81 | 0.795 | 0.795 | 0.795 | 0.795 | 0.805 | 0.76 |
| 15 | 0.665 | 0.665 | 0.675 | 0.68 | 0.665 | 0.68 | 0.725 | 0.78 | 0.745 |
| 20 | 0.58 | 0.575 | 0.58 | 0.6 | 0.59 | 0.63 | 0.69 | 0.745 | 0.745 |
| 25 | 0.575 | 0.58 | 0.59 | 0.585 | 0.58 | 0.575 | 0.615 | 0.71 | 0.72 |
| 30 | 0.55 | 0.54 | 0.55 | 0.55 | 0.55 | 0.575 | 0.605 | 0.74 | 0.705 |
| 33 | 0.545 | 0.565 | 0.575 | 0.57 | 0.575 | 0.615 | 0.61 | 0.74 | 0.73 |

Fig. 8 Validation Success surface, showing optimal ridge at approximately 10% tail, and worst cases for 5 contours and high tail percentages



However, many high *VaR* results are rejected informally, so the percentage rejected by the objective decision process is not so surprising. Table 3 shows a detailed analysis of the *VaR* variation with both the number of contours and the tail length, expressed as a percentage of the total data count. Since small data tails are more important than larger ones (due to greater applicability of the *GPD* fit in those regions), we settle on 6 contours as the preferred number to produce optimal percentages of correct validations. Eight per cent of the rejections were rejected due to the criterion $V_B < V_{max}$ in Eq. 12.

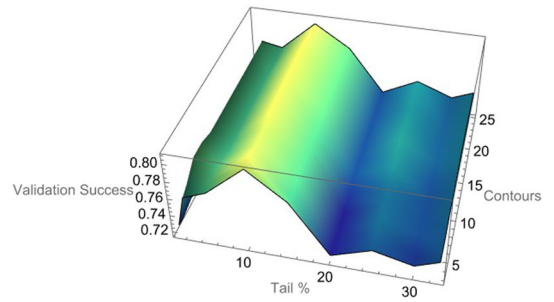
The surface graphic in Fig. 8 illustrates the validation success rates plotted against the number of contours in the *credible region*, and the tail length, expressed as a % of the total data count. There is an optimal ridge at approximately 10% tail, and a descent to a worst case success rate for high tail percentages.

5.7 Validation Results for a Restricted 'Small σ, ξ ' Region

The *feasible region* is of particular interest, because we expect 'acceptable' *VaR* values to originate from it. 'Small' values of the *GPD* parameters σ and ξ generally indicate that a fitted distribution should not be rejected. Within the *feasible region*, the relationship between *VaR* and σ and ξ is essentially linear. We restrict the *credible region*, \mathbf{R} to $0 < \sigma \leq 25 \times 10^6$ and $0 < \xi < 0.5$. The validations are restricted to those cases where the fitted *GPD* σ and ξ parameters fall in the restricted region. Others are ignored since the linearised fit in the restricted region is not applicable elsewhere. The results are noted in Table 4. Overall, they are very similar to Table 3, with marginally improved results. Similarly, the validation surface plot looks very similar to Fig. 8.

Figure 9 shows a surface plot of the data in Table 4. The optimal peak at approximately 10% tail is clear. There is minimal variation with the number of contours, probably due to the near planar nature of the surface.

Fig. 9 Validation Success surface for a restricted 'small σ, ξ ' range, showing optimal ridge at approximately 10% tail, and almost no variation with the number of contours



5.7.1 Validation 'Sense Checks'

In practice, the simplest path for practitioners is to apply the 'sense check', which implements the validation criteria. The only essential requirements are to select the optimal tail percentage, calculate the necessary descriptive statistics (maximum and mean and frequency), calculate *VaR* (for all the data), and then apply the test in Eq. 13. Subjectively, best fits that are not LogNormal, LogNormal Mixture, or LogNormal Gamma Mixture distributions might be treated with suspicion. In particular, *GPD* and *Extreme Value* distributions often produce 'excessive' *VaR*. Table 5 shows some examples of 'sense checks', with comments to indicate what features in the data a practitioner might look for to make a subjective decision.

Comments

1. *VaR* and Maximum are both huge
 2. *VaR* and maximum are both low
 3. *VaR* is reasonable given maximum value
 4. *VaR* is large but not unreasonably so, but maximum is not excessive: doubtful decision
 5. *VaR* seems reasonable given maximum value but low mean triggered a likely incorrect decision
- Notes 1,2,3 are subjective 'correct decision' verdicts
 - Note 4 is a subjective 'doubtful decision' verdict
 - Note 5 is a subjective 'almost certainly wrong decision' verdict

Table 4 Proportion of correct validations if decisions are based on a comparison of *GPD-Surfaces* with data in a restricted region, $0 < \sigma \leq 25 \times 10^6$ with $0 < \xi < 0.5$

| Tail % | Contours | | | | | | | | |
|--------|----------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 3 | 4 | 6 | 8 | 10 | 12 | 15 | 20 | 25 |
| 1 | 0.725 | 0.725 | 0.73 | 0.72 | 0.725 | 0.725 | 0.715 | 0.72 | 0.725 |
| 2 | 0.76 | 0.765 | 0.78 | 0.78 | 0.775 | 0.775 | 0.775 | 0.775 | 0.775 |
| 5 | 0.77 | 0.775 | 0.775 | 0.77 | 0.77 | 0.77 | 0.77 | 0.77 | 0.77 |
| 10 | 0.805 | 0.81 | 0.805 | 0.805 | 0.81 | 0.81 | 0.81 | 0.81 | 0.805 |
| 15 | 0.775 | 0.77 | 0.775 | 0.775 | 0.775 | 0.775 | 0.78 | 0.78 | 0.78 |
| 20 | 0.72 | 0.72 | 0.72 | 0.72 | 0.725 | 0.73 | 0.73 | 0.73 | 0.73 |
| 25 | 0.735 | 0.735 | 0.74 | 0.74 | 0.74 | 0.75 | 0.755 | 0.75 | 0.75 |
| 30 | 0.725 | 0.725 | 0.73 | 0.73 | 0.73 | 0.73 | 0.735 | 0.735 | 0.735 |
| 33 | 0.735 | 0.735 | 0.735 | 0.74 | 0.74 | 0.74 | 0.745 | 0.745 | 0.745 |

Table 5 Sense check

| Distribution | VaR (m) | Max. (m) | Mean (m) | Frequency | Decision | Note |
|-----------------|---------|----------|----------|-----------|----------|------|
| Burr | 52858 | 53.5 | 1.54 | 69.6 | Reject | 1 |
| LN | 222 | 11.4 | 2.87 | 51.2 | Accept | 3 |
| LN | 612 | 101 | 1.72 | 73.8 | Reject | 4 |
| LN Mix | 179 | 32.7 | 1.14 | 44.2 | Accept | 2 |
| LN Mix | 378 | 8.5 | 0.52 | 129 | Accept | 2 |
| Burr | 100300 | 148.1 | 1.78 | 74.8 | Reject | 1 |
| Frechet | 42682 | 513.3 | 4.31 | 130 | Reject | 1 |
| LN Mix | 4 | 0.1 | 0.03 | 103.8 | Accept | 2 |
| LN | 1 | 0.3 | 0.01 | 53.4 | Accept | 2 |
| LN Γ Mix | 129 | 15.8 | 0.43 | 142 | Reject | 5 |
| LN | 616 | 50 | 4.43 | 51 | Accept | 3 |
| GPD | 7041 | 14.1 | 0.32 | 71 | Reject | 5 |
| LN Γ Mix | 112 | 7 | 0.96 | 68.6 | Accept | 2 |
| LN Mix | 115 | 3.5 | 0.55 | 147.8 | Accept | 2 |
| Weibull | 173 | 9.4 | 1.64 | 67.4 | Accept | 2 |

6 Discussion

An essential component of the *VaR* calculation is to generate data that have values greater than the largest observed datum. If *VaR* is calculated using the empirical data only, the result represents a minimum value, and does not measure a largest "acceptable" value. The numerical values of draws from a random sample from the fitted distribution that exceed the maximum observed value depend on two factors. The first is the fitted distribution from which that sample was drawn. The second is its curvature for large ordinates. It is not possible to calculate a generally applicable expression for curvature, since that depends on the parameters of the fitted distribution.

The computations involved in modelling credibility using the method described are considerable, as is the time taken to do them. The overall task can be partitioned into two distinct stages. Producing the *GPD surfaces* is a one-off process. Once done, the ordinates on each surface can be stored and used as required. If a high degree of accuracy is needed, each ordinate on each surface takes 2–3 min to produce (using an *i7* Windows PC with 48MB RAM), since each complete evaluation requires in excess of 1 million Monte Carlo cycles. With 12 *GPD surfaces*, each with 110 ordinates, the total production time is approximately 44h processor time. The second stage, testing a target *VaR* value against those surfaces is, happily, quick. The *VaR* value to be tested still takes 2–3 min, but the subsequent processes (calculating a corresponding surface-based *VaR*) value and comparison with the target take a few seconds. However, the details of the decision process are complicated, and are unlikely to be easily understood by non-technical risk managers. Therefore we would recommend that the entire calculation process is embedded in a 'black-box' application (written in C++ or similar), accompanied by a detailed explanation of how it works. The *VaR surfaces* can be held in a database.

With the passage of time, operational risk loss severity, and therefore *VaR* values, would be expected to rise in the same way that prices, wages etc. might change. In practice, operational risk losses have been remarkably stable over the past 12 years, although financial crime has been an increasing problem in recent years. We would expect the proposed credibility method to be stable with respect to such changes, provided that the reference data (i.e. the *VaR surfaces*) are still commensurate with the *VaR* levels encountered in practice. *VaR surfaces* can be recalibrated periodically (perhaps every two or three years) by repeating the entire *GPD-PREP* process in Sect. 4.2 with updated reference samples.

7 Conclusion

Practitioners who encounter a calculated VaR value that they consider "unacceptably large" are faced with a dilemma. A huge capital requirement is unsustainable. It inhibits lending, which is how a retail bank makes money. A subjective decision to discard the offending distribution and seek an alternative has to be made. The key points to consider are:

- Is the VaR value consistent with the loss sum? A large VaR should follow from a large loss sum, but not from a small loss sum.
- Is the VaR value consistent with previous similar calculations?
- Has there been a significant change in the data since previous calculations were done?
- Can distribution changes be made, such that the VaR value becomes acceptable?

The analysis presented in this paper is an attempt answer these questions in an objective way. The decision rule in Eq. 12 should not be seen as firm. Rather, it should be regarded as a guideline.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A Dominance of Tail over Body Losses

Proposition A quantile, Q_T , calculated using tail losses only, is significantly larger then the corresponding quantile, Q_B , calculated using body losses only, provided that the sum of body and tail losses, n , is constant. Let the tail losses be a small proportion, p , of n . Specifically, denote the mean and standard deviation of a body distribution by μ_B and σ_B respectively, and the mean and standard deviation of a tail distribution by μ_T and σ_T respectively. We assume that the ratio of means and standard deviations are small, as in Eq. A1, in which ϵ is a small real number.

$$\mu_B = \epsilon\mu_T; \quad \sigma_B = \epsilon\sigma_T \quad (A1)$$

Then, to first order in ϵ , for a quantile defined by a z standard deviations,

$$Q_T - Q_B \approx np\mu_T + z\sigma_T\sqrt{np}. \quad (A2)$$

Proof The proof proceeds by using the *Central Limit Theorem* to define sums of body and tail sampling distributions in the context of the *LDA VaR* calculation. [23] showed that the error induced in *LDA*-sampling by replacing a stochastic value of n by a constant, is immaterial. Using a constant value for n means that the *Central Limit Theorem* applies, as in Eq. A3.

$$\begin{aligned} S_T &\sim N(np\mu_T, \quad np\sigma_T^2) \\ S_B &\sim N(n(1-p)\mu_B, \quad n(1-p)\sigma_B^2) \end{aligned} \quad (A3)$$

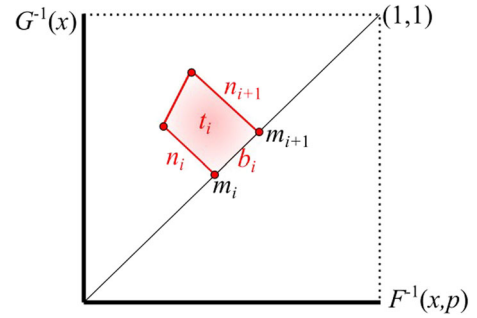
We now approximate a high quantile by an ordinate on the above Normal distributions, defined by selecting a suitable number, z , of standard deviations. For example, $z = 3.09$ is an approximation for the 99.9% quantile. Therefore, using Eq. A1

$$\begin{aligned} Q_T &\approx np\mu_T + z\sigma_T\sqrt{np} \\ Q_B &\approx n(1-p)\epsilon\mu_T + z\epsilon\sigma_T\sqrt{n}\sqrt{1-p} \end{aligned} \quad (A4)$$

The difference is

$$Q_T - Q_B = n\mu_T(p - \epsilon(1-p)) + z\sigma_T\sqrt{n}(\sqrt{p} - \epsilon\sqrt{1-p}) \quad (A5)$$

Fig. 10 Q-Q plot, showing the geometry used in the *TNA* test development in Eq. B2, showing a single trapezium t_i and its vertices and edges



The proportion p is expected to be less than 0.25, but ϵ is expected to be orders of magnitude smaller. Therefore, the terms in ϵ can be safely removed. The result, Eq. A2, then follows. The difference $Q_T - Q_B$ is large because $n\mu_T$ and $\sigma_T\sqrt{(n)}$ are both large. Therefore, Q_B is immaterial compared to Q_T , and body losses can be removed from the *LDA VaR* calculation without incurring serious error.

Appendix B Outline of the *TNA* Significance Test

The basis of the *TNA*-test is an empirical-theoretical data comparison. Specifically, for data x , let $G(x)$ and $F(x, p)$ be the CDFs of the empirical distribution and a theoretical distribution with parameters p . Further, let $G^{-1}(x)$ and $F^{-1}(x, p)$ be their inverses. Then for N data points, Eq. B1 defines a set of empirical-theoretical ordered pairs.

$$S = \left(F^{-1}\left(\frac{i}{N+1}, p\right), G^{-1}\left(\frac{i}{N+1}\right) \right) \quad i = 1..N \quad (\text{B1})$$

For short, write F^{-1} and G^{-1} instead of the two terms in S in the equations that follow. The *TNA*-test first computes the perpendicular distances n_i of points in S to the diagonal line of gradient 1 in the Q-Q plot. Taking their absolute values effectively maps points in S that are 'below' the Q-Q diagonal to above the Q-Q diagonal. In Eq. B2, m_i is the x - (and y -) coordinates of the intersection on the Q-Q diagonal of the perpendicular from (F^{-1}, G^{-1}) to the Q-Q diagonal. The terms t_i define trapezia with sides of length n_i and n_{i+1} , using the 'base' term b_i . Both are illustrated in Fig. 10. Finally, the value of the *TNA* statistic is the sum of those trapezia. Geometrically the *TNA* value is interpretable as the *Enclosed Area*, E , between the envelope of the points in S mapped to 'above' the Q-Q diagonal and the Q-Q diagonal.

$$\begin{aligned} m_i &= \frac{F^{-1} + G^{-1}}{2} \quad i = 1..N \\ n_i &= \sqrt{\left(F^{-1} - m_i\right)^2 + \left(G^{-1} - m_i\right)^2} \\ b_i &= \sqrt{2}(m_{i+1} - m_i) \quad i = 1..(N-1) \\ t_i &= \frac{b_i(n_i + n_{i+1})}{2} \quad i = 1..(N-1) \\ E &= \sum_1^{N-1} t_i \quad (\text{the TNA statistic}) \end{aligned} \quad (\text{B2})$$

The *Enclosed Area* E (more fully $E(x, F, p)$) is zero if all points (F^{-1}, G^{-1}) lie precisely on the Q-Q diagonal. That case constitutes the Null hypotheses. The geometric interpretation of the worst possible instance of the Alternative hypothesis is that the *Enclosed Area* is 0.5 (the area of the upper triangle in the Q-Q plot). The significance level is calculated as a ratio of the measured deviation to the maximum possible area, 0.5.

$$(\text{Null}) \quad H_0 : E(x, F, p) = 0$$

$$(\text{Alternative}) \quad H_1 : E(x, F, p) > 0 \quad (\text{B3})$$

References

1. Balkema, A., Haan, L.: Residual life time at great age. *Annals Probab.* (1974). <https://doi.org/10.1214/ao10.2307/1968974p/1176996548>
2. BCBS.: Supervisory guidelines for the advanced measurement approaches. Basel Committee on Banking Supervision <https://www.bis.org/publ/bcbs196.pdf> (2011)
3. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012)
4. Clough, D.J., Kotz, S.: Extreme value distributions with a special queueing model application. *Ann Math. Statist.* **39**(3), 881–889 (1965). <https://doi.org/10.1214/aoms/1177698320>
5. Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer, London (2001)
6. Dodd, E.L.: The greatest and least variate under general laws of error. *Trans. Amer. Math. Soc.* **2**(5), 525–539 (1923)
7. EBA.: Regulation and policy: Operational risk. European Banking Authority <https://www.eba.europa.eu/regulation-and-policy/operational-risk> (2017)
8. Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Camb. Philos. Soc.* **24**, 180–190 (1928)
9. Frachot, A., Georges, P., Roncalli, T.: Loss distribution approach for operational risk. Working paper, Groupe de Recherche Operationnelle, Credit Lyonnais, France <http://ssrn.com/abstract=1032523> (2001)
10. Frechet, M.: Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math. Cracovie* **6**, 93–116 (1927)
11. Gnedenko, B.V.: Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**(3), 423–453 (1943). <https://doi.org/10.1214/aos/1176343003>
12. Gray, A., Abbena, E., Salaman, S.: *Modern Differential Geometry of Curves and Surfaces with Mathematica*. Chapman and Hall, London (2006)
13. Gumbel, E.: *Statistics of Extremes*. Columbia University Press, Columbia (1958)
14. Jenkinson, A.F.: Frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. R. Meteor. Soc.* **81**, 158–171 (1955)
15. Jorion, P.: *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd edn. McGraw-Hill, Noida (2000)
16. Leadbetter, M.R.: Extremes and local dependence in stationary sequences. *Zeit. Wahrscheinl.-Theor.* **65**, 291–306 (1983)
17. Von Mises, R.: Über Die Variationsbreite Einer Beobachtungsreihe. *Sitzungsber Berlin Math Ges, Lebanon* (1923)
18. Von Mises, R.: La distribution de la plus grande de n valeurs. *Amer Math Soc Reproduced in Selected Papers of Richard von Mises, II* **1954**, 271–294 (1936)
19. Mitic, P.: Improved goodness-of-fit tests for operational risk. *J. Oper. Risk* **15**(1), 77–126 (2015). <https://doi.org/10.21314/JOP.2015.159> <https://doi.org/10.21314/JOP.2015.159> <https://doi.org/10.21314/JOP.2015.159>
20. Mitic, P.: Credible value-at-risk. *J. Oper. Risk* **18**(4), 33–70 (2023a). <https://doi.org/10.21314/JOP.2023.005> <https://doi.org/10.21314/JOP.2023.005> <https://doi.org/10.21314/JOP.2023.005>
21. Mitic, P.: Maximum value-at-risk. In: Vasant, P., Weber, G.-W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Lecture Notes in Networks and Systems*, vol. 569, pp. 981–990. Springer, Cham (2023b). https://doi.org/10.1007/978-3-031-19958-5_58567/eal02030005 <https://doi.org/10.58567/eal02030005> <https://doi.org/10.58567/eal02030005>
22. Mitic, P.: Reasonableness and correctness for operational value-at-risk. *Econ. Anal. Lett.* **2**(3), 35–44 (2023c). <https://doi.org/10.58567/eal02030005> <https://doi.org/10.58567/eal02030005> <https://doi.org/10.58567/eal02030005>
23. Mitic, P., Bloxham, N., Cooper, J.: Incremental value-at-risk. *J. Model Risk Valid.* **14**(1), 1–37 (2020). <https://doi.org/10.21314/JRMV.2020.216> <https://doi.org/10.21314/JRMV.2020.216> <https://doi.org/10.21314/JRMV.2020.216>
24. Pickands, J.: Moment convergence of sample extremes. *Ann. Math. Stat.* **39**, 881–889 (1968)
25. Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**, 119–131 (1975). <https://doi.org/10.1214/aos/1176343003>
26. Rayner, G., MacGillivray, H.: Numerical maximum likelihood estimation for the g -and- k and generalized g -and- h distributions. *Stat. Comput.* **12**, 57–75 (2002)
27. Resnick, S.: *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, Cham (2006)
28. Sochi, T.: *Introduction to Differential Geometry of Space Curves and Surfaces*. Amazon ISBN 978-1546735892 <https://www.amazon.co.uk/Introduction-Differential-Geometry-Curves-Surfaces/dp/1546735895> (2017)
29. Yee, T., Wild, C.: Vector generalized additive models. *J. Roy. Stat. Soc. B* **58**, 481–493 (1996)
30. Zaliapin, I., Kagan, Y.Y., Schoenberg, F.: Approximating the distribution of pareto sums. *Pure App Geophys.* **162**(6), 1187–1228 (2005). <https://doi.org/10.1007/s00024-004-2666-3>
31. van Zyl, J.: Estimation of the shape parameter of a generalized pareto distribution based on a transformation to pareto distributed variables. *J. Stat. Theory Pract.* **9**(1), 171–183 (2015). <https://doi.org/10.1080/15598608.2013.865106>