Maximum Value-at-Risk

Peter Mitic[1,2]

[1] Department of Computer Science, University College London, UK,
p.mitic@ucl.ac.uk
[2] Santander Bank, London UK

Abstract. Some Value-at-risk estimates can be so large that their validity is questionable, and are subjectively rejected. An objective rejection criterion, based on a comparison of empirical data with a Generalised Pareto model of the data tail and applying the Pickands-Balkema-deHaan Theorem is presented. A consequent definition and measure of 'Maximum Value-at-risk' is developed and validated.

Keywords: Value-at-Risk, Generalised Pareto, Pickands-Balkema-deHaan Theorem, Single Loss Approximation, order statistic

1 Introduction

The context of this paper is *Operational Risk* (hereinafter *OpRisk*), which is, informally, the risk of "things going wrong". More rigorously, the European Banking Authority defines *OpRisk* as "the risk of losses stemming from inadequate or failed internal processes, people and systems, or from external events." [1]. Regulated firms (banks, insurers, financial advisors etc.) must, annually, calculate the amount of retained capital to cover *OpRisk* losses that it might incur in the following year. This is done by fitting an appropriate *fat-tailed* distribution to data, drawing random samples from that distribution, and calculating *value-at-risk* (*VaR*). Sometimes the calculated value is considered, subjectively, "excessive". In that case, an alternative distribution must be selected. Here we propose a simple objective criterion for distribution acceptance or rejection.

1.1 Minimum and Maximum Value-at-Risk

A 'natural' theoretical minimum for *VaR* exists. It is best understood by referring to the *Loss Distribution Algorithm* [2] (usually abbreviated to *LDA*), which is a generally applicable method for *VaR* calculation. The *LDA* comprises a convolution of a frequency and a severity distribution, in which a single draw from the frequency distribution is used in the severity distribution to generate a random sample of losses. If *VaR* is calculated using the empirical data only, the maximum draw from any random sample cannot be greater than the maximum of the empirical data. Sampling in this way is known as the *Empirical Bootstrap*. In contrast, if a distribution is fitted to the data, and a sample is drawn from the fitted distribution, the sample could contain elements that are much greater

(even by orders of magnitude) than the maximum of the empirical data. Therefore, *VaR* derived using the *Empirical Bootstrap* is expected to be less than *VaR* derived from a fitted distribution.

The concept of *maximum VaR* is problematical because a draw from a distributional random sample is theoretically unbounded. Informally, we refer to a *maximum acceptable VaR*, which is the maximum that would be tolerated in normal practice. If an extremely high *VaR* value emerges, it might be immediately rejected because it exceeds some established statistic (e.g. a country's GDP or capitalisation of a large corporation). Such cases are obvious, but others are not. An acceptability decision for cases that are higher than usual are harder to judge, especially if there has been no significant data change compared to prior calculations. A rigorous definition will be formulated in Section 3. *Maximum VaR* (hereinafter *MaxVaR*) is calculated by optimising a function of the *Generalised Pareto* maximum order statistic.

2 Literature Review

In this review we concentrate on the relevance of the *Generalised Pareto Distribution* (*GPD*) to tail *VaR* calculations, and on associated approximations. See [3] and [4] for details of the latter.
The relevance of the *GPD* arises from the theory of *Threshold Exceedances*, which links the *GPD* to the tail data (a subset comprising the largest data elements). The basic formulation was done in the mid-1970s by Pickands [5], and Balkema and de Haan [6]. They defined an *Excess* distribution, which models the distribution of data that exceeds some threshold u. Equation 1 gives the definition of an *Excess* distribution $H_u(x)$ when the distribution function of X is $F(x)$. $H_u(x)$, is equivalent to the distribution of a random variable X-u, given $X > u$.

$$H_u(x) = P(X - u < x | X > u) = \frac{F(u + x) - F(u)}{1 - F(u)} \tag{1}$$

The ensuing *Pickand-Balkema-de Haan* Theorem (hereinafter abbreviated to *PBH*), states that the *Excess* distribution tends to a function G as the tail size decreases (Equation 2). G is the *GPD* , with density and distribution functions given in Equation 3. Its three parameters are location (μ), scale (σ) and shape (ξ). G is particularly sensitive to ξ. See [7] or [8] for the case $\xi \leq 0$.

$$H_u(x) \to G(x, \mu, \sigma, \xi); \quad u \to \infty \tag{2}$$

$$\begin{aligned} f(x : \mu, \sigma, \xi) &= \Big(\frac{1}{\sigma}\Big)\Big(1 + \frac{\xi(x - \mu)}{\sigma}\Big)^{-1-\frac{1}{\xi}} \quad x \geq \mu, \sigma > 0, \xi > 0 \\ F(x : \mu, \sigma, \xi) &= 1 - \Big(1 + \frac{\xi(x - \mu)}{\sigma}\Big)^{-\frac{1}{\xi}} \quad x \geq \mu, \sigma > 0, \xi > 0 \end{aligned} \tag{3}$$

A formal proof was given by Leadbetter [9], and an outline of the proof may be found in the books by Coles [7] (section 4.2.2), Resnick [10], or Embrechts et al [8]. The *PBH* theorem has remained largely static since it was first proved, and enhancements are few and more recent. An example is a 2-dimensional extension using copulas to model marginal distributions [11].

The proposed solution in this paper assumes that a set of *OpRisk* losses may be replaced by the largest loss in the set. This replacement is the basis of the closed form *VaR* approximation formulated by Boecker et al. [3], known as the *Single Loss Approximation* (*SLA*). For confidence level $0 < \kappa < 1$, with $\kappa \sim 1$, the *SLA* takes the form in Equation 4 for a fitted distribution F:

$$VaR(\kappa) \sim F^{-1}\Big(1 - \frac{1-\kappa}{\nu}\Big) \;\; ; \quad \kappa \to 1^{-} \tag{4}$$

The *SLA* assessment by Hess [12] concluded that accuracy is best for high frequency data. Boecker later added a refinement by including a "data mean correction" [4]. Degan [13] further modified the *SLA* by analysing the asymptotic behaviour of the underlying distribution. Further *SLA* enhancements were made by Opdyke [14] and Hernandez [15].

3 Proposed Solution

In order to deal with problem cases where a calculated *VaR* value appears excessive, we propose the following definition. It is suggested by the observation that *VaR* is largely determined by tail losses, for which a *GPD* is, by the *PBH* Theorem, a good model.

Definition 1.
Maximum VaR, abbreviated to MaxVaR with notation $\mathbb{M}$*, is a scaling of the median value* $\mathbb{M}'$*, by a scale factor* λ*, of the distribution of random sample maxima of given size drawn from a GPD distribution.*

This definition enables us to develop an objective acceptability criterion (Section 3.3). $\mathbb{M}(\mu, \sigma, \xi, \nu, \lambda)$ is a function of the three *GPD* parameters, the sample size ν, and the scale factor λ. For convenience, we omit them and write $\mathbb{M}$. The scale factor, λ, has to be determined by a regression of the derived median $\mathbb{M}'$ on the fitted distribution *VaR*. We will refer to that regression line as the *Median Line*.

3.1 Overall Procedure

In order to calculate a value for λ, the following procedure is used. It makes use of the *PBH* Theorem by modelling the distribution tail by a *GPD*, and uses the order statistic of the maximum loss via the *SLA*.

1. Generate multiple data sets, partition each into body/tail at predetermined percentages. For each data set:

   (a) Fit an optimal distribution to the body, and a *GPD* to the tail
   (b) Calculate the median of the maximum order statistic of the data
2. Formulate a regression model *scaled median* against *fitted VaR*
3. Calculate the best linear fit line with error bounds
4. Optimise the scale factor such that lower error bound gradient is 1. This forces *MaxVaR* to be numerically equal to the scaled median.

Assumptions

This above procedure requires two assumptions. Both concern the convolution of the frequency and severity distributions.

Assumption 1

ν can be treated as a constant, not a random variable. Therefore there are ν terms in the sum of severities for all Monte Carlo iterations in the *LDA*. This assumption is justified empirically for medium/large frequencies. For a theoretical proof, see [16].

Assumption 2

The *SLA* approximation can be applied to the sum of severities. The sum of severities can then be replaced by the largest element in the sum. This assumption has been justified for medium/large frequencies in [13] and [12].

3.2 Theoretical basis

We would like to quantify an upper bound for "acceptable" *VaR* by considering the distribution tail. Theorem *PBH* shows that a good model for a distribution tail is a *GPD*. In particular, the tail maximum is significant in determining the *VaR* for all losses. Ideally, we would prefer to designate the expected value, $\mathbb{E}(\mathbb{M})$, of the distribution of the maximum as representative of the required upper bound. However, the generalised form of the integral involved in calculating that expected value is intractable.

Rather than attempt an approximation for the expected value, we use the median in place of the mean. This gives an expression which is much easier to simplify, and uses a principle that often applies in *Oprisk*: the median is often a more representative distribution property than is the mean. The median is not subject to influence by extreme outliers. Using the median prompts the following proposition.

Proposition 1.
Let $\mathbb{M}$ be measured by the scaled median value of the distribution function of the maximum datum in a random sample of size ν, $S_\nu = \{X_1, X_2, ..., X_\nu\}$, drawn from a $GPD(\mu, \sigma, \xi)$ distribution. Then, if the scale factor is λ and $\mathbb{M}'$ is the median, $\mathbb{M}$ is given by Equation 5.

$$\mathbb{M}' = \Big[\Big(\frac{2^{1/\nu}}{2^{1/\nu} - 1}\Big)^{\xi} - 1\Big]\frac{\sigma}{\xi} + \mu \;\; ; \qquad \mathbb{M} = \lambda\mathbb{M}' \tag{5}$$

Proof.
Consider the *LDA* component in the r^{th} Monte Carlo iteration. With *Assumption 1*, there are a constant number ν of severity draws (from a *GPD* random variable X) The sum of severities is:

$$S_{(r)} = X_{r1} + X_{r2} + ..., + X_{r\nu} \tag{6}$$

With *Assumption 2*, replace $S_{(r)}$ by $max\{X_{r1}, X_{r2}, ..., X_{r\nu}\}$. Then the distribution function of $S_{(r)}$ can be expressed in terms of the distribution function of a *GPD*.

For any loss x, with a standardised random variable $z = \frac{x-\mu}{\sigma}$, denote the distribution function of a $GPD(\mu, \sigma, \xi)$ by $F(x)$. Then the distribution function of the maximum of the random samples of size ν, $S_{(r)}$, is $F^{\nu}(x)$. Then using the standard result for the maximum order statistic of a *GPD*, the median value, $\mathbb{M}'$ is given by Equation 7.

$$\frac{1}{2} = F^{\nu}(\mathbb{M}') = \left[1 - (1 + z\xi)^{-1/\xi}\right]^{\nu} \tag{7}$$

Solving for $\mathbb{M}'$ gives

$$\mathbb{M}' = \left[\left(\frac{2^{1/\nu}}{2^{1/\nu} - 1}\right)^{\xi} - 1\right]\frac{\sigma}{\xi} + \mu \tag{8}$$

Now set $\mathbb{M} = \lambda\mathbb{M}'$ for some scale factor λ, which is to be optimised empirically. The result, Equation 5, follows immediately. □

Observations from empirical data indicate that there is a lower limit for $\mathbb{M}$. That limit may not be useful in practice (it may be too low), but its existence is quite marked. We therefore propose a corollary to the above proposition that quantifies the lower limit. Figure 1 shows the relationship between the *Median Line*, and its lower confidence bound, which forms a *MaxVaR* line. The horizontal axis records the *VaR* calculated by fitting a distribution to all the data, *V*, and the vertical axis records the calculated *MaxVaR*, $\mathbb{M}$. The figure is also used in the corollary.

Corollary
In Figure 1, the lower confidence bound represents the condition in the first part of Equation 9, referred to a value $V = U_{95}$. Its gradient is 1, so that numerically, $\mathbb{M} = U$. Any line with gradient $0 < \alpha < 1$ represents a lower value for *MaxVar* (the second part of Equation 9. One such line is shown.

$$P(\mathbb{M} < U) = 0.05 \ ; \qquad \therefore P(\alpha\mathbb{M} < U) < 0.05 \tag{9}$$

We now seek a value for α such that $P(\alpha\mathbb{M} < U)$ is extremely small. Specifically, the value U is a distance corresponding to a confidence bound of 1.96 standard deviations relative to the *Median Line*. We seek a line that represents

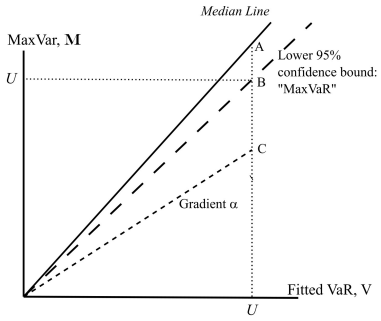

Fig. 1. Median Line with the 95% lower confidence bound (representing "MaxVaR"), optimised such that "MaxVaR" is numerically equal to the Fitted *VaR*. Distance AB corresponds to a confidence bound of 1.96 standard deviations, and distance AC corresponds to 5 standard deviations.

5 standard deviations (i.e. "5-sigma confidence"). Referring to Figure 1, we set a value for α as in Equation 10.

$$\alpha = \frac{AC - AB}{AC} = (5 - 1.96)/5 \sim 0.6$$
$$\therefore P(\alpha\mathbb{M} < U) \sim 2.7 \times 10^{-7} \tag{10}$$

The value $\alpha \sim 0.6$ is apparent in practice. See Section 4.

3.3 Acceptance/Rejection criterion

Given a fitted *VaR*, V_f, an Accept/Reject criterion is formulated as follows.

1. Extract the data tail at $p\%$. Note the tail length ν
2. Fit a *GPD* to the tail, giving parameters μ, σ, ξ
3. Apply Equation 8 to yield $\mathbb{M}'$
4. Multiply by an appropriate factor λ to give $\mathbb{M} = \lambda\mathbb{M}'$
5. Accept the fitted *VaR* if $V_f \leq \mathbb{M}$. Reject otherwise

3.4 Data

Data sets were generated from random samples drawn from distributions commonly encountered in *OpRisk* for a medium-sized European bank:

– LogNormal, Lognormal mixtures and Lognormal Gamma Mixtures;
– The *Max-Stable* distributions: Weibull, Frechet and Gumbel;
– Generalised Pareto, Tukey G-and-H, LogLogistic, Burr (Type VII);
– Others: Gamma, LogGamma, and LogCauchy.

Data sets were then mixed at random, and random data perturbations were made, so that it would not possible to conclude that any particular distribution is optimal. Approximately 2500 mixed data sets were generated, each with between 200 and 2000 elements. Elements that were clearly unrealistic (e.g. greater than 5 billion) were removed before use.

4 Results

We first present overall results in terms of two metrics. The scale factors λ, were calculated from the optimisation in Equation 8. Post-calculation, the number of data sets 'accepted' and 'rejected', according to the criterion in Section 3.3, were enumerated. In Table 1, if n_A is the number of data sets for which the calculated *VaR* was accepted, and n_R is the number rejected, the *Pass Rate* is defined as the ratio $\frac{n_A}{n_A+n_R}$.

Table 1. Summary of scale factors λ and pass rates per tail percentage

Tail %	**Lambda**	**Pass Rate**
1	22.3	70.1
2	25	69
5	27.2	74.2
10	34.6	68.3
15	33.9	68.9
20	28.8	66.8
25	25.6	62.4
33	27.5	66.3
50	17.6	63.8
Mean	26.9	67.8
SD	5.3	3.5

Table 1 shows no clear pattern for the variation of λ with tail size. A linear fit results in $R^2 = 0.157$, and the fit improves as the degree of the fitted polynomial increases. For a quadratic, $R^2 = 0.656$ and for a cubic, $R^2 = 0.803$. However, we caution against increasing the polynomial excessively to avoid over-fitting. Instead, we prefer to use the mean value, and add one standard deviation if higher measured *VaR* values are merited due to significant data changes.

Figure 2 shows a typical the regression line (in this case for a 10% tail) of *MaxVaR* (calculated using Equation 8) against fitted *VaR*. Each element in the point cloud is one of the generated data sets. In this case, the value $\lambda = 34.6$ was somewhat high. The point cloud shows a typical 'splay' from the origin. Some points with $VaR \sim 100$ have *MaxVar* between 400 and 500. Intuitively, such *VaR* values are likely to be deemed 'excessive'. Note that the region representing mid-to-high fitted *VaR* and low *MaxVar* is empty. A straight line traced

from the origin to the point (600, 1000) (i.e. its gradient is 0.6) approximately demarcates the populated region from the 'empty' region. This observation agrees with the theoretical value obtained in the corollary to Proposition 1, the value α in Equation 10. The empirical results from all other percentage tails also agree with the theoretical result.

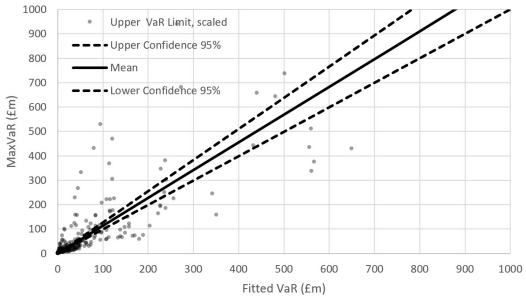

Fig. 2. Scatter of *Fitted VaR* against *scaled VaR Upper Limit*, showing linear fit with lower 95% confidence bound aligned at gradient = 1. The *VaR Upper Limit* is represented by the upper 95% confidence bound.

Figure 3 shows a view of the collective results. The surface shown is of *MaxVaR* (the dependent variable) against the independent variables *Tail %* and *Fitted VaR*. It shows the following features in particular.

1. An association between high *MaxVaR*, and low tail % plus high *Fitted VaR*.
2. At fixed tail %, sharp ridges reflect wide variation of *MaxVaR* with *Fitted VaR*. Variation is most marked for *Fitted VaR* between 1000 and 2000.
3. The region corresponding to high *Fitted VaR* plus a high tail % is empty, with a sharp diagonal linear boundary, as in Figure 3.

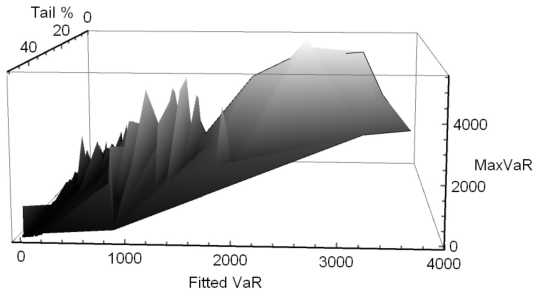

Fig. 3. The surface shows the "choppy" relationship between *MaxVaR* and variables *Fitted VaR* and *Tail %*, including a sharply demarcated 'empty' region, and increasing *MaxVar* with increasing *Fitted VaR* and decreasing *Tail %*

4.1 Informal Validation

For an informal validation, ten additional data sets were generated, and the test of Section 3.3 was applied. Table 2 shows a comparison of the fitted *VaR*, and the *MaxVaR* for the tail percentages indicated. The λ values from Table 1 were used for each corresponding tail percentage. Column *Note* is an "Accept/Reject" decision, qualified in some cases.

Table 2. Validation

	Fitted VaR	**1%**	**2%**	**5%**	**10%**	**15%**	**20%**	**25%**	**Note**
1	104	106	140	135	132	129	108	95	Accept, note 1
2	4101	5667	27201	5191	6338	6534	5687	5444	Accept, note 2
3	3.0	19.4	21.8	42.8	17.1	27.2	24.4	19.3	Accept, note 3
4	828	531	648	730	671	755	843	735	Reject, note 1
5	1017	1471	1649	1874	2251	3049	2922	2231	Accept
6	6125	41292	46292	50365	3794	59787	41291	294855	Accept, note 2
7	1.0	4.1	4.6	5.0	7.0	7.5	6.4	4.9	Accept, note 3
8	201	763	856	898	1131	1073	940	1002	Accept
9	135	3724	4855	1615	1510	1259	1184	722	Accept
10	531	5997	7974	11844	3301	2511	2023	1285	Accept

Validation Notes for Table 2

1. Majority decision based on tail %
2. Accepted, but *VaR* is intuitively high given the fitted LogNormal parameters
3. Low *VaR* would not normally be subject to a "maximum" test

5 Discussion

The problem outlined in the Introduction - how to formulate an objective criterion to decide whether or not a fitted *VaR* value is excessive given the data, has been solved by developing an optimisation procedure based on a *GPD* model of tail data. The result is the *MaxVaR* value, $\mathbb{M}$, which should be greater than the fitted *VaR*. The decision criterion is not foolproof, as the comments on the validations in Table 2 show. It should be seen more as guide.

We end with two reasons why a *VaR* calculation might be considered 'excessive'. First, the gradient of the fitted cumulative distribution function might be very small (although positive) for large loss values. The 99.9% centile will therefore be extremely large. Second, the data fitting process may have failed, in the sense that it may have converged to an alternative local optimum. In principle, any marked change in *VaR* on a year-on-year basis should be reflected in a clear change in the data distribution. For example, if one or more severe losses have appeared, an increase in *VaR* would be expected.

References

1. European Banking Authority: Regulation and policy - Operational risk. (2017). https://www.eba.europa.eu/regulation-and-policy/operational-risk
2. Frachot, A., Georges, P., and Roncalli, T.: Loss distribution approach for operational risk. (2001). Working paper, Groupe de Recherche Operationnelle, Credit Lyonnais, France. http://ssrn.com/abstract=1032523
3. Boecker, K. and Kluppelberg, C.: Operational var: a closed-form approximation. (2005) Risk. pp.90-93
4. Boecker, K. and Sprittulla, J.:Operational var: meaningful means. (2006). https://mediatum.ub.tum.de/doc/1072585/file.pdf
5. Pickands, J.: Statistical inference using extreme order statistics (1975), Annals of Statistics, vol. 3, pp. 119-131, 10.1214/aos/1176343003
6. Balkema, A. and Haan, L.: Residual Life Time at Great Age (1974). Annals of Probability vol. 2 10.1214/ao10.2307/1968974p/1176996548
7. Coles, S.: Intro. Statistical Modeling of Extreme Values (2001), Springer
8. Embrechts,P, Klueppelberg,C., and Mikosch,T.: Modelling Extremal Events for Insurance and Finance (1997), Springer
9. Leadbetter, M. R.: Extremes and local dependence in stationary sequences (1983), Zeit. Wahrscheinl.-theorie vol. 65, pp. 291-306
10. Resnick, S.I.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling (2006), Springer
11. Wuethrich, M.V.: Bivariate extension of the Pickands–Balkema–de Haan theorem (2004) Annales de l'Institut Henri Poincare (B) Probability and Statistics 40(1), pp. 33-41, https://doi.org/10.1016/j.anihpb.2003.03.002
12. Hess,C.: Can the single-loss approximation method compete with the standard monte carlo simulation technique? (2011). Jnl. Operational Risk 6(2), pp.31-43, DOI:10.21314/JOP.2011.091
13. Degan,M.: The calculation of minimal Regulatory Operational Risk Capital using single-loss approximations. (2010) Jnl. Operational Risk, 5(4), pp.3-17. DOI: 10.21314/JOP.2010.084
14. Opdyke,J.D.: Fast, Accurate, Straightforward Extreme Quantiles of Compound Loss Distributions (2017). Jnl. Operational Risk 12(4) pp. 1-30 DOI:10.21314/JOP.2017.195
15. Hernandez, L., Tejero, J., Suarez, A. and Carrillo-Menendez, S.: Closed-form approximations for operational value-at-risk (2013). Jnl. Op. Risk 8(4), pp. 39-54
16. Mitic,P., Cooper,J. and Bloxham,N.: Incremental Value-at-Risk (2020) Jnl. Model. Risk Validation, 14(1), pp. 1-37. DOI:10.21314/JRMV.2020.216