Research Paper

A central limit theorem formulation for empirical bootstrap value-at-risk

Peter Mitic[1,2,3] and Nicholas Bloxham[1]

[1]Santander UK, 2 Triton Square, Regent's Place, London NW1 3AN, UK;
emails: peter.mitic@santandergcb.com, nicholas.bloxham@santander.co.uk

[2]Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK

[3]Laboratoire d'Excellence Régulation Financière (LabEx ReFi), 12 Place du Panthéon, 75231 Paris Cedex 05, France

ABSTRACT

In this paper, the importance of the empirical bootstrap (EB) in assessing minimal operational risk capital is discussed, and an alternative way of estimating minimal operational risk capital using a central limit theorem (CLT) formulation is presented. The results compare favorably with risk capital obtained by fitting appropriate distributions to the same data. The CLT formulation is significant in validation because it provides an alternative approach to the calculation that is independent of both the empirical severity distribution and any dependent fitted distribution.

Keywords: central limit theorem (CLT); operational risk; value-at-risk (VaR); minimum risk capital; loss distribution.

1 INTRODUCTION

In this paper, we describe an alternative method to calculating minimal operational risk capital: one that can replace the empirical bootstrap (EB). The way in which

---

operational risk capital should be calculated is governed by the regulations of the Basel Committee on Banking Supervision (2011, Guideline 196), which gives us a great deal of flexibility when building a capital model. The overall aim of this process is to estimate how much capital should be retained in order to cover future expected and unexpected operational risk losses. There is a de facto standard methodology for doing this, consistent with the aforementioned Basel regulations. An appropriate severity distribution is fitted to historical losses, samples are taken from that distribution, and a Monte Carlo process is used to simulate annual losses. The Monte Carlo method was established by Frachot *et al* (2001), and it is now widely known as the loss distribution approach (LDA). The 99.9% value-at-risk (VaR) is then extracted. Internal data may be augmented using external data in order to model the losses that could possibly have been incurred but have not been. Scenarios provide a forward-looking view of losses that are anticipated in the near future. The thrust of operational risk modeling is to combine the three elements – internal data, external data and scenarios – in order to produce the VaR value. A well-known validation check on the result is to apply the LDA process to the empirical data rather than samples taken from a fitted curve. This step is known as the EB. The purpose of this paper is to discuss an alternative to the EB that has certain advantages over it.

Calculating VaR using empirical data (as opposed to a fitted distribution) is a useful tool in the process of calculating operational risk capital for one principal reason: to provide an estimate of the minimum capital. Samples are taken only from losses that have been incurred, so it is not possible to draw a sample that contains a value greater than the maximum empirical loss. If an appropriate distribution is fitted to operational risk loss data and samples are drawn from that distribution, then it is possible – although unlikely – to draw values that are much larger than the maximum empirical loss. The expectation is therefore for a larger VaR result than would be obtained using empirical losses only, although it is impossible to predict from the outset how much larger the VaR derived from a fitted distribution might be.

In this paper, we employ the central limit theorem (CLT) to develop an "alternative bootstrap" that does not involve sampling. The results are then compared with those obtained via a "traditional" EB as well as with those obtained by fitting appropriate severity distributions. The advantages of using this method are stated in the following section.

1.1 Bootstrapping

"Bootstrap" sampling techniques were originally developed by Bradley Efron and are discussed in Efron and Tibshirani (1986). Such methods depend on resampling from a single initial sample. We can imagine a theoretical infinite population, and we want to make inferences about the structure of that population. For example, its

distribution, descriptive statistics and confidence intervals are all of interest. To make inferences about such a population, we would normally draw random samples from it. It is assumed that these samples will reflect the structure of the parent population, and that the way to ensure this is to draw a sufficient number of samples. However, in the context of operational risk losses, there is no such infinite population. The number of losses is finite and may even be small. To draw random samples from such a finite population may not provide reliable inferences about the population. A common adjustment is to apply the following multiplicative correction factor to the estimator for the sample variance,

$$\sqrt{\frac{N - \nu}{N - 1}},$$

where N is the population size and ν is the sample size (see, for example, Isserlis 1918). A "rule of thumb" is to do this if ν is greater than 5% of N. Thompson (2012) presents examples of sampling from small populations and gives an account of associated bias. The present discussion is more applicable to larger population sizes, for which a variance correction may still be appropriate.

1.2 Advantages for model validation

The use of the EB in operational risk modeling is recommended for validation purposes by the US Board of Governors of the Federal Reserve System (2014, p. 17). The alternative that we propose goes further: with a suitable configuration, it could be used as an alternative way to calculate operational risk capital rather than minimum operational risk capital.

We stress the following advantages for model validation.

- Using an independent methodology is a good way to validate the efficacy of a model and the results it produces. In this case, the input data is the only common element shared by the EB method and the proposed CLT method.

- The proposed methodology is not stochastic, so there is no need for extensive simulations to achieve a given degree of accuracy.

- In order to estimate VaR, the proposed methodology integrates the three elements – internal data, external data and scenarios – into a common method. This unified methodology takes advantage of well-established statistical results, and it is easy to formulate and implement.

- We give examples in the ensuing discussion of distributions of VaR that do not appear to fit common probability distributions well. The CLT approach is distribution independent because all calculations are based on "normal" distributions.

In addition, there are some more general reasons for stressing the applicability of the CLT as an alternative validation method to the EB. If the VaR calculated by the EB is greater than the VaR calculated by fitting a distribution, then this is an indication that the fitted distribution underestimates tail VaR. In these circumstances, a quick and convenient check of the EB value is a useful validation technique. Indeed, there have been previous warnings about the use of the EB with heavy-tailed distributions. Balta *et al* (2009) say that "traditional resampling methods typically fail for heavy-tailed loss models and rare event estimation" and call for further investigation. To say that these methods "typically fail" is to be overly pessimistic; however, it is not hard to observe failure in the sense that a fitted distribution produces a "VaR that is too low".

Balta *et al* (2009) also suggest that alternative sampling methods (such as importance sampling) might be used in an EB. We suggest that amendments to these sampling methods cause further validation problems. Instead, more parameters must be introduced. The parameters chosen and their values should be validated and sensitivities to them should be calculated. The CLT approach is more "validation friendly", as sampling is not applicable.

1.3 The empirical bootstrap procedure

In the EB procedure, resamples with replacement are drawn from a base sample that constitutes the set of empirical losses in the case of operational risk. With replacement, individual losses can be selected multiple times; this is effectively a proxy for not being able to include any loss greater than the largest empirical loss in a given sample. There is an assumption that resamples – if chosen randomly – will resemble the population that they came from. Resampling gives an estimate of the sampling distribution of the sample statistic in question (usually the sample mean and variance), rather than providing an estimate of the statistics and distribution of a population. In passing, resampling is of primary importance in the context of sampling from organic material (museum items, for instance), which often involves destructive sampling. Items that are sampled cannot be replaced, so one small sample is often all that exists.

Dutta and Perry (2007) amplify this point (that it is not possible for a future loss to be greater than the maximum empirical loss); however, they also give some caveats. First, sampling can be unduly affected by the existence of a single large loss, leading to an overestimate of capital. Second, smaller financial institutions are less likely to have very large losses, leading to an underestimate of capital. They do not discuss the dependence of any capital estimate on the modeling threshold. Experience shows that capital value is heavily dependent on the modeling threshold. In order to not bias its results, an empirical method should use a modeling threshold of zero: the value used in all the calculations presented here.

ALGORITHM 1

---

(1) Calculate an annual loss frequency, $f = N/y$.

(2) Repeat n times.*

    (a) Obtain a sample size z by drawing a random sample of size 1 from a Poisson(f) distribution.**

    (b) Generate a uniform random sample of size z, $S_z = \{N_1, N_2, \ldots, N_z\}$ from the set of index numbers $\{1,2,\ldots,N\}$, with replacement.

    (c) If $z = 0$, set a sum of losses variable, s_z, to zero and do not apply steps (d) and (e).

    (d) If $z > 0$, extract the losses corresponding to the index set S_z.

    (e) If $z > 0$, set the sum of losses variable, s_z, to the sum of the losses in S_z (that is, an estimate of annual loss) and retain the result.

(3) End repeat.

(4) Calculate the 99.9th percentile of the retained sums.

---

*Generally, we use $n \sim 100\,000$, which provides sufficient confidence in the result without taking too long to calculate. (If $n < 1000$, the 99.9th percentile is always the largest value of the retained sums.)
**A Poisson distribution is not the only way to do this. A negative binomial distribution is often used. For large f, a normal approximation to the binomial is appropriate.

1.3.1 Sampling methodology

In the context of a Monte Carlo VaR process, Efron's resampling method is applied in the same way that it is applied in the LDA calculation effected by Frachot *et al* (2001), except that only empirical losses are used, rather than losses generated by fitting a distribution. The EB algorithm is as follows: let $L = \{L_1, L_2, \ldots, L_N\}$ be a set of N (> 0) losses covering a period y years, and let n be the number of Monte Carlo trials used.

Algorithm 1 is discussed from a theoretical point of view in the context of operational risk by Cruz *et al* (2015).

1.4 Summary of previous work on the empirical bootstrap

Shevchenko (2011) gives an outline of how a "vanilla" bootstrap method may be applied in operational risk modeling. This is a very general account of operational risk modeling focusing, in particular, on two "flavors" of bootstrap: parametric and nonparametric. In this paper, the nonparametric bootstrap, which does not assume any severity distribution, is used. Homolka (2012) gives a similar account but rather misses the point that the EB should be used to estimate minimum capital. This point was stated in the introduction to this paper and we stress it again here. Homolka's

account compares the EB with sampling from an extreme value distribution; it is not surprising that the capital and error bounds from the latter exceed those from the EB.

Previous research on the EB has concentrated on amending the sampling technique described above. One example comes from Pfeffermann and Correa (2012), who developed a method for bias correction by considering the error in estimating model parameters. They considered a number of plausible parameter values and sought an appropriate functional relationship between the original sample for each parameter and the corresponding bootstrap samples. A comparison with the bootstrap method of Section 1.3.1 shows a good agreement.

Hall and Presnell (1999) tackle another well-known problem in operational risk capital calculations. Samples drawn from heavy-tailed distributions can contain huge values that affect statistical estimators adversely. Hence, Hall and Presnell (1999) use a weighted bootstrap to moderate the "contaminated" distribution, in which each member of a bootstrap sample is assigned a weight according to an assessment of its influence on dispersion. Their paper is interesting in that, in principle, its technique might be applied in reverse in order to estimate a maximum capital value.

The accuracy and stability of variance estimates in bootstrap processes was investigated by Manzi and Mecatti (2008). They accounted for auxiliary variables and the sampling design in resampling, tying results to the strategies, decisions and policies that depend on them.

In the following section of this paper, we will highlight the different characteristics of the distribution of operational risk losses. We will then formally model this distribution as a bipartite entity in Section 2.2. Separate treatments of ordinary losses (due to high-frequency, low-severity events) and large losses (due to low-frequency, high-severity events) are undertaken by Wang *et al* (2012) in a sophisticated manner. They combine a bootstrap with a piecewise-defined severity distribution in order to measure small operational risk samples. This approach helps to correct bias in the estimators (caused by the small sample) and – due to the bipartite split – permits a better fit to the distribution as a whole.

The bipartite nature of an operational risk loss distribution is treated in a different way by Böcker and Klüppelberg (2005), who showed that an approximate loss distribution may be obtained under the assumption that the loss severities are subexponentially distributed. Specifically, the condition they use is that, for n losses $\{L_1, L_2, \ldots, L_n\}$ drawn from a distribution arising from a random variable $\boldsymbol{L}$,

$$P(L_1 + L_2 + \cdots + L_n > x) = P(\max(L_1, L_2, \ldots, L_n) > x) \quad \text{as } x \to \infty.$$

This condition is known as the "principle of the single big jump" or the "catastrophe principle". The impact of this approximation is that it formally characterizes the VaR calculation in terms of the distribution tail, but only under the subexponential

assumption. The normal approximation to the EB proposed in this paper confirms the heavy dependence of VaR on the distribution tail. Böcker and Sprittulla (2006) noted that the Böcker–Klüppelberg approximation is prone to significant errors when the confidence level for the sum $L_1 + L_2 + \cdots + L_n$ is not very close to 1. In such cases, the expected loss is not negligible compared with the operational VaR. Böcker and Sprittulla (2006) show that, as the confidence level decreases from 99.9% to 99%, the approximation error reduces rapidly to about 50%. Their method links the probability that the total loss exceeds some x with the probability that the largest loss exceeds a smaller amount, which they estimate. Advancing the method further, Hannah and Puza (2015) extend the Böcker–Klüppelberg–Sprittulla formulas to multiple loss types, where the aggregate loss is a sum of random sums, each defined by a different subexponential distribution. VaR rates of less than 99.9% might typically be used for nonregulatory purposes, such as risk appetite calculations (at 95%, for instance). The decreasing accuracy of the EB at VaR $< 99.9\%$ makes the proposed alternative CLT method all the more applicable.

In a more general context, Abad *et al* (2015) discuss the relationship between VaR and financial time series returns, linking the two via a loss function. They propose a firm's loss function that measures the opportunity capital cost of the firm when its losses are covered and also find that the VaR model (designed to minimize total losses) is robust to the regulator's loss function but not to the firm's loss function. Abad *et al* (2015) emphasize the importance of the EB, even though the word "bootstrap" does not appear in their paper:

> Supervisors are concerned about how many times losses exceed the VaR and the size of the noncovered losses. However, the risk managers have a conflict between the goal of safety and the goal of profit maximization.

This type of conflict is common, and a minimum VaR estimate serves to tie down what the VaR value "should" be. Of course, an estimate of the maximum VaR would also be useful, but that is a matter for further research.

1.5 Regions in the frequency distribution

Figure 1 shows a typical frequency distribution for operational risk losses. The loss frequency for high-value losses is very small compared with the loss frequency for low-value losses. In addition, a small number of extremely high-value losses are only hinted at in Figure 1. These three categories of loss – low, high and extremely high – are treated separately in the ensuing discussion, as each has a distinct effect on the VaR calculation of Algorithm 1.

The illustration in Figure 1, although originating from genuine loss data, does present a somewhat idealized view of a frequency distribution. In other cases, low-value losses occupy a single, left-most bar in the histogram; an extended loss

FIGURE 1 Typical loss frequency distribution.
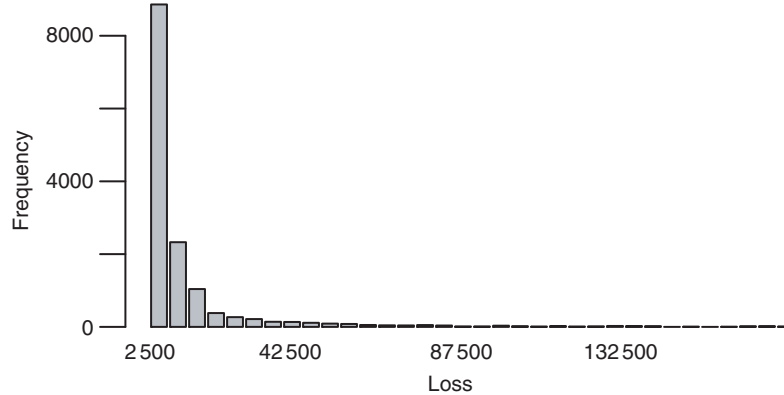
(horizontal) axis is also shown. This indicates that a small number of extremely large losses (perhaps only one) are present, but their frequency is so small as to be invisible in the figure.

In order to account for the characteristics of the three loss categories mentioned above, we partition the frequency distribution into three regions as follows, with respect to a percentage p.

- Outlier: a set of the largest losses, as determined by an "outlier test" discussed below. This set may be empty if the "outlier test" returns a result of nil.

- Tail: the largest $(100 - p)\%$ of the remaining losses (ie, total losses after removing the outliers).

- Body value: the smallest $p\%$ of remaining losses, comprising low- and mid-value losses. The largest losses in this category can still be considerable.

In practice, a common value for p is 95%. Usually, there are only a few outliers, if any at all. The distribution "body" typically has considerably more losses than the distribution "tail". The somewhat bizarre situation can arise where the sum of all losses in a 95% tail (that is, only 5% of the total number of losses) can also represent 95% of the sum of all losses. That leaves the bulk in the "body" representing 95% of the total number but only 5% of the total value. Outliers and tail losses can affect the final VaR unduly. Therefore, it is expedient to treat them separately, as described below.

1.6 Treatment of outliers and tail losses

An expedient and simple way to detect outliers is to use a "u-sigma" test. Given a set of losses L, a set of outliers L_O is defined in terms of the mean and standard deviation of the losses in L, and a multiplier u:

$$L_O = \{L : |L - \text{mean}(L)| > u \times \text{sd}(L)\}. \tag{1.1}$$

Values of u are given in subsequent sections. There are alternative ways to detect outliers. Grubbs's test (Grubbs 1950) is a de facto standard for the detection of outliers and is readily available in the R programming language. However, it is only applicable if the underlying distribution is normal, and operational loss distributions are not. The distribution of log(losses) has some chance of being normally distributed but, all too often, even log(loss) distributions have fat tails. Dixon's Q test (Dixon 1950) has the same normality requirement. Further, Grubbs's test yields only the largest outlier per application of the test; it is more useful to detect all at once given an objective detection criterion as in (1.1).

From a validation viewpoint, rejecting outliers may be an undesirable practice. Pyle (2003, Chapter 8), for example, makes the point that validation should always account for outliers, since outliers may be informative. Our proposed CLT alternative is not significantly affected by outliers, so it is a viable alternative to the EB.

2 THE NORMAL APPROXIMATION TO THE EMPIRICAL BOOTSTRAP

In this section, we develop the theory that underpins the "normal approximation to EB" and provide a means of calibrating the method.

2.1 Value-at-risk is asymptotically normally distributed

As per Rao (2001), it can be shown that VaR is normally distributed. The proof provided by Rao assumes the existence of a sequence of n independent and identically distributed (iid) random variables $X_1, X_2, \ldots, X_n$. In the context in which we use them, these random variables represent losses selected at random from a distribution appropriate to operational risk losses. They have a distribution function F, and Rao defines a real number p $(0 < p < 1)$ that serves as a binomial parameter. Rao then shows that the speed of convergence depends on n, p and F. Let μ be the mean and σ^2 be the variance of each of the $X_1, X_2, \ldots, X_n$. Guegan *et al* (2015) prove the same result and use it to build a set of confidence intervals around the true but unknown VaR, producing a "spectral stress VaR". The proof provided by Rao (2001) applies to a continuous distribution, and we extend it to apply to a discrete empirical distribution without assuming the existence of any underlying probability distribution. Our proof

is given in Proposition 1 in online Appendix D. The first key result from this (labeled D16 online) is

$$\sqrt{n}(F_n(l) - F(l)) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2), \tag{2.1}$$

in which l is any fixed (positive) loss, $F_n(l)$ denotes the empirical cumulative distribution function derived from an observed sample of n losses, $F(l)$ denotes a limiting cumulative distribution function, and σ^2 is expressed as $F(l)(1 - F(l))$. Although there are theoretical expressions for the mean and variance of the normal distribution shown in (2.1), what really matters for the validation process is what is observed in practice, that is, why an alternative validation method (derived, in this case, from the CLT) is valuable as a complement to the existing EB method. There are a number of reasons for this. First, the results from the two methods should be "about the same". The loose language here is intentional and appropriate, because the meaning of "about the same" is a matter of judgement. For example, a 50% difference in two very small VaR results might be as acceptable as a 5% difference between two very large VaR results.

Second, the bootstrapped empirical VaR X has an asymptotic normal distribution given by

$$X \sim N(\mu', (\sigma')^2), \tag{2.2}$$

where μ' and σ' are given in D22 of online Appendix D. In practice, μ' and σ' would be estimated from data.

2.1.1 Convergence speed of the empirical bootstrap to normal approximation

Compared with the CLT, the Berry–Esseen theorem (Esseen 1956) gives a more stringent formulation for the convergence of the scaled mean of a random sample to a normal distribution with increasing sample size. It indicates the rate of convergence by giving a bound on the maximal error of approximation between the standard normal distribution and the distribution of the scaled sample mean. The rate of convergence is proportional to both $1/\sqrt{n}$ (for sample size n) and the third moment of the scaled empirical distribution.

Reiss (1974) uses the Berry–Esseen theorem to bound the approximation error between the sample quantiles and the standard normal distribution. With the application of Reiss's result, it can be shown that the approximation error of our normal VaR approximation (2.2) is also bounded. Further, by assuming such lenient conditions as a subexponential distribution for the underlying losses (see Goldie and Klüppelberg 1998) and an easily met requirement on the number of iterations of the bootstrap parameter k, it can be shown that the order of convergence for our

normal approximation of the EB VaR is approximately $O(1/\sqrt{n})$: the same rate of convergence obtained under the Berry–Esseen theorem.

2.2 The body/tail partition

The parameters μ and σ are not known but can be estimated from data. It must be remembered that the normal distribution in (2.1) is derived from a particular set of samples – namely, the ones that result in a VaR at a given confidence level – out of all the samples that could be taken. Many of those samples will not include the largest losses. Consequently, μ and σ^2 cannot be taken as the mean and variance, respectively, of the set of all losses L. In order to model the tail behavior, a body/tail split may be used. Given an appropriate real number p in the range $(0, 1)$, the distribution body comprises the smallest $100p\%$ of all losses, and the distribution tail comprises the remaining largest $100(1 - p)\%$ of all losses (see Wang *et al* (2012) for a similar body/tail split).

Estimators of μ and σ^2 may be found by defining similar estimators for the body and tail of the distribution, and combining them under the assumption that the body and tail portions of the distribution are both normally distributed. In practice, if only the tail estimators are used, the results are almost unchanged. Useful values for the body/tail threshold p are 0.95 or 0.99. This gives a good balance between a tail that is too small (and therefore gives unreasonably small capital values) and a tail that is too large (and therefore gives unreasonably large capital values). Algorithm 2 is used to determine appropriate estimators of μ and σ^2. The estimators of μ and σ^2 are denoted by $\hat{\mu}$ and $\hat{\sigma}^2$, respectively, with subscripts as appropriate.

2.3 Configuration for the body/tail weight

The body/tail weight parameter t was determined by configurations based on known distributions. The assumption is then that the result will be applicable to loss distributions derived from actual data, which may not fit any formal distribution well. The aim of the configuration was to determine, for the configuration data sets, a capital value slightly in excess of the capital value derived from the EB, resulting in a capital value slightly more conservative than the EB capital value.

A range of distributions that are commonly used for operational risk capital calculation (lognormal, Weibull, log gamma, generalized Pareto (GP), lognormal mixture, Burr, loglogistic, Gumbel, Fréchet and *g*-and-*h*) were selected. Tukey's *g*-and-*h* distribution (Jiménez and Arunachalam 2011) is peculiar in that it is very difficult to find an initial parameterization to achieve a data fit; further, it can be useful for fits to distributions with both large outlier losses and small losses. For each distribution, random losses were generated. The resulting distributions covered a range of losses typically found in loss distributions: a large number of small to medium-sized losses,

ALGORITHM 2 Combined body/tail estimators.

---

(1) Define body and tail losses, L_B and L_T (remembering that the losses are ordered):

$$L_B = \{L_i : 1 \leqslant i < pN\}, \qquad L_T = \{L_i : 1 + pN \leqslant i \leqslant N\}.$$

(2) Calculate separate estimators for body and tail mean and variance:

$$\hat{\mu}_B = \text{mean}(L_B), \qquad \hat{\sigma}_B^2 = \text{sd}(L_B),$$
$$\hat{\mu}_T = \text{mean}(L_T), \qquad \hat{\sigma}_T^2 = \text{sd}(L_T).$$

(3) Define a body/tail weight, t (see Sections 2.2 and 2.3).

(4) Calculate the distribution estimators $\hat{\mu}, \hat{\sigma}^2$ (the norm notation $|\cdot|$ denotes the number of elements in a set, and y is the number of years of data used):

$$\left.\begin{aligned}
\hat{\mu} &= \frac{(\hat{\mu}_B|L_B| + t\hat{\mu}_T|L_T|)}{y}, \\
\hat{\sigma}^2 &= \frac{(\hat{\sigma}_B^2|L_B| + t\hat{\sigma}_T^2|L_T|)}{y}.
\end{aligned}\right\} \tag{2.3a}$$

(5) Calculate VaR using

$$\text{VaR} = \hat{\mu} + 3\sqrt{\hat{\sigma}^2}. \tag{2.3b}$$

In stage (5), the multiplier 3 provides the 99.9% VaR ($\Phi^{-1}(3) = 0.9987$, where Φ is the standard normal CDF function).

In practice, the following estimators give numerical results that differ from those obtained with (2.3a) by less than 0.5%:

$$\hat{\mu} = \frac{t\hat{\mu}_T|L_T|}{y} \quad \text{and} \quad \hat{\sigma}^2 = \frac{t\hat{\sigma}_T^2|L_T|}{y}. \tag{2.3c}$$

The above estimators do not require any assumptions of normality for either the body or the tail VaR distribution.

---

with some very large ones. VaR was calculated for each using the EB. The values generated were typical of capital values calculated using actual loss data, ranging from less than one million to billions.

The scale factor calculation was found to be relatively insensitive to the value of p in the $(0.9, 0.99)$. In cases where the number of losses is small (less than twenty), a body/tail threshold of 0.9 was enough to provide sufficient points for all subsequent calculations. In practice, $p = 0.99$ resulted in capital values that were very near to the EB capital values. As the body/tail threshold decreased to 0.9, an increase in capital value relative to the bootstrap capital values was noted; but this increase

was acceptable on the grounds that a larger "minimal" capital value provides a more conservative view.

2.3.1 Scale factor indicators

In order to avoid reliance on only one scale factor calculation methodology, five indicators were used to calculate the scale factor t in (2.3a). They are discussed in this section. The intuition underlying our scale factor calculation is that the "shape" of the underlying distribution CDF is important in determining the influence of tail losses on total capital value, and that this shape may be characterized in several ways. The important characteristics of the shape of the CDF are its curvature and how corresponding statistics (that is, the mean, standard deviation, etc) compare for the body and the tail.

Each indicator, as detailed below, delivers a single numerical result; these results are then combined (see Section 2.3.2) to produce the scale factor value. The first indicator is the Jarque–Bera statistic, which uses a combination of the first four empirical moments of the distribution. The second and third both measure the curvature of the distribution empirical CDF in different ways. The third and fourth compare statistics of the body with those of the tail. In every indicator calculation, the losses for each test distribution are scaled to the range $(0, 1)$ for consistency. Figure 2 shows a scaled idealized empirical CDF, and some key points on the CDF are marked for reference in the detailed discussion that follows.

In subsequent sections, the following notation is used:

- n denotes the total number of losses;

- n_T denotes the number of losses in the tail;

- n_B denotes the number of losses in the body;

- m is the mean of all losses; and

- s is the standard deviation of all losses (with divisor n, not $n - 1$).

2.3.1.1 Jarque–Bera indicator.   The Jarque–Bera (JB) statistic (Jarque and Bera 1980) is a test for the deviation from normality of a distribution, and it uses empirical skewness and kurtosis as measures of this deviation. It is used here because skewness and kurtosis are important characteristics of fat-tailed operational risk loss distributions. JB is formulated as follows.

Let the third and fourth moments of the n losses be

$$m_3 = \sum (L_i - m)^3 / n$$

and

$$m_4 = \sum (L_i - m)^4 / n,$$

respectively. Then, the JB statistic is given by

$$S = m_3 / s^3,$$
$$C = m_4 / s^4,$$
$$\mathrm{JB} = (n - k + 1)(S^2 + (C - 3)^2 / 4) / 6,$$

with $k = 2$, corresponding to two moments. For large samples, the JB statistic has a $\chi^2(2)$ distribution. For smaller samples, it is better to assess critical values using Monte Carlo methods (Jarque and Bera 1981). The indicator used in this analysis is given by

$$J = \sqrt{\frac{\mathrm{JB}}{9.21}}. \tag{2.4}$$

The divisor 9.21 is the 1% $\chi^2(2)$ critical value, and the ratio JB/9.21 measures the effective scaling of the data, so a null hypothesis that the data is normally distributed may be accepted. The square root is compensation for having squared C and S in the JB expression. The critical 5% $\chi^2(2)$ value (5.991) was also considered, but the 1% value proved to be satisfactory in combination with other indicators.

2.3.1.2 Curvature indicator 1.

The first curvature indicator measures the curvature of the empirical CDF in terms of the maximum perpendicular distance, $D_{\max}$, of the scaled CDF (to $(0, 1)$, as above) from the 45° line. This is distance PP′ in Figure 2. An examination of the perpendicular distance of all points on the empirical CDF quickly yields the maximum distance, even when the number of such points is large. There is some possibility of optimizing the search process if assumptions are made about where the maximum might occur. Distance $D_{\max}$ is then scaled by the ratio n_T / n in order to account for the tail volume. Thus, the first curvature indicator is given by

$$\mathrm{C1} = \frac{1}{D_{\max} n_T / n}. \tag{2.5}$$

The inversion brings the value of the indicator more in line with other indicators.

FIGURE 2 Idealized empirical CDF.
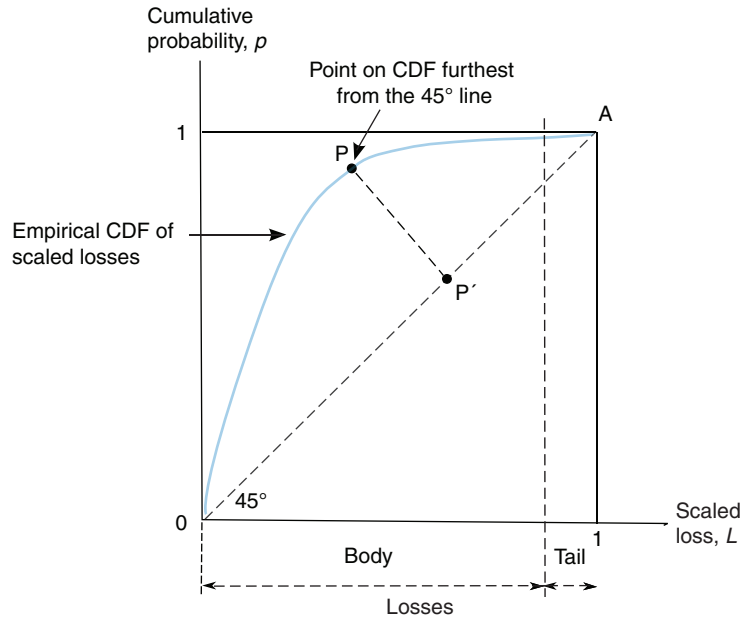
Scaled to (0, 1), this empirical CDF shows division between body and tail as well as a maximum deviation PP′ from the 45 degree line. This figure is referenced throughout Section 2.3.1.

2.3.1.3 **Curvature indicator 2.** The second curvature indicator estimates the actual curvature at point P in Figure 2. In terms of the coordinates representing loss (L) and cumulative probability (c), let P be the point (L_{max}, c_{max}) on the empirical CDF corresponding with the maximum distance PP′. A neighborhood of $n/100$ points on the empirical CDF surrounding (L_{max}, c_{max}) is selected, and a quadratic curve $Q(L)$ is fitted to the points in the neighborhood. Fitting a curve in this way helps to reduce volatility in the result. The curvature $\kappa(L)$ at a point $(L, c(L))$ on a curve is calculated using the curvature expression

$$\kappa(L) = \frac{c''}{(1 + (c')^2)^{3/2}}$$

(see, for example, Weisstein 1999, Equation 14). This provides the second curvature measure:

$$C2 = Q(L_{max}). \tag{2.6}$$

2.3.1.4 **Ratio indicator.** The ratio indicator R is a comparison of the body and tail of the empirical distribution using the minimum and maximum losses in the tail

and for all losses:

$$R = 1 + \frac{\max(\text{tail}) - \min(\text{tail})}{\max(\text{all}) - \min(\text{all})}. \tag{2.7}$$

2.3.1.5 Quantile indicator. The quantile indicator Q–Q is a comparison of the body and tail of the empirical distribution using the quantile losses. Denoting the 25%, 50% (ie, the median) and 75% quantiles for all losses by q_{25}, q_{50} and q_{75}, respectively,

$$\text{Q–Q} = \frac{q_{25} + q_{75} - 2q_{50}}{\max(\text{all})}. \tag{2.8}$$

This measure is similar to the Bowley–Galton skewness statistic (see Shanmugam and Chattamvelli 2015, Section 4.3.1).

2.3.2 Combined indicators

Once the indicators given by (2.4)–(2.8) have been calculated, they are scaled to a uniform range $(0, 1)$ using a sigmoid function $M(z) = z/(1 + z)$. The combined indicator – the scale factor t used in (2.3 a) – is given by twice the mean of the scaled indicators, as follows:

$$t = 2\frac{M(\text{JB}) + M(\text{C1}) + M(\text{C2}) + M(\text{Q–Q}) + M(\text{R})}{5}. \tag{2.9}$$

The mean value of the indicators, $(M(\text{JB}) + M(\text{C1}) + M(\text{C2}) + M(\text{Q–Q}) + M(\text{R}))/5$, provides a comprehensive indicator value that is not too dependent on any one of the indicators. It was found that, used without factor 2 in (2.9), the resulting VaR was underestimated in some cases. Therefore, factor 2 was introduced to redress this imbalance. The value 2 arises from calibration using the test distributions. It is geared to provide marginally larger VaR results than obtained with the EB, in order to be more conservative. It is expected that, in many cases, the VaR derived from a fitted distribution will be much larger than the EB VaR.

2.3.3 Indicators as a diagnostic tool for validation

Thus far, our discussion has concentrated on the technicalities of the indicators used to calculate the scale factor that determines the weight placed on the tail losses in determining the CLT approximation (parameter t in (2.3 a)). Using the mean of all five indicators (2.9) provides as objective a value for t as possible. Varying t from its indicator-calculated value can then be used as a validation technique for a fitted distribution in the following way. If the VaR value obtained by placing greater emphasis on tail losses increases markedly with increasing t, then this is an indication that the loss distribution has a very fat tail, and that an appropriate fitted distribution might be the

GP, the generalized extreme value (GEV) or similar. If, as t increases, the VaR value increases at a much slower rate, an appropriate fitted distribution might be lognormal (ie, one that rarely produces extreme VaR values). The problem of producing inflated VaR values when fitting very fat-tailed distributions to loss data is less pronounced when using the CLT method, which is less sensitive to outlier loss values. Diagnosing the type of fitted distribution in this manner provides a way to validate the actual fitted distribution. Deviations from the diagnosis would indicate that the VaR calculated from the fitted distribution is either too low or too high.

2.3.4 Indicators for the validation of fitted distributions

Validation of a fitted distribution (rather than the VaR calculated using that distribution) has two useful implications. First, if the best-fit distribution is GP- or GEV-like, and CLT validation indicates that the best-fit VaR is too high, there is a case for rejecting the best-fit distribution and selecting the next-best fit as an alternative. The CLT method is more stable than the EB method when used in this way because it is not prone to stochastic sampling variations. Diagnosing the VaR calculated from a fitted distribution as being "too high" is an unsolved problem in operational risk analysis. VaR that is "too high" often arises from GP- or GEV-like fitted distributions as well as when attempting to fit any distribution to a small number of very large losses. The latter problem is described in Mitic (2017). Whereas the EB and its CLT equivalent are both aimed at estimating minimum VaR, no technique currently exists for estimating maximum VaR.

The second useful validation advantage is for a fitted loglogistic distribution. Estimating VaR multiple times using the LDA method with a loglogistic distribution can result in a bimodal VaR distribution. Looking ahead, Figure 3(b) is a case in point. Such a result is due to a propensity of the loglogistic distribution to occasionally produce excessive very high losses in a random sample. Calculating the mean value of those annual losses produces a VaR value, but it is not necessarily representative of the true VaR, which might be much nearer one peak of the VaR distribution than the other. Deciding which peak to choose is the task of the CLT estimation. A full validation process is then, given a fitted loglogistic distribution, to detect a bimodal VaR distribution in the first place, and to compare VaR values corresponding to the peaks in that distribution with the CLT VaR estimation.

2.4 Configuration results

Algorithm 2 and (2.9) were applied to twenty-one loss distributions, generated from the range mentioned in Section 2.2. The resulting VaR values are given in Table 1. In that table, the abbreviations used for the distribution taxonomy are

- LN = lognormal,

- W = Weibull,

- LG = log gamma,

- GP = generalized Pareto,

- LNMix = lognormal mixture,

- B = Burr,

- LL = loglogistic,

- G = Gumbel,

- F = Fréchet,

- GH = Tukey g-and-h.

The normal approximation results are compared with the results estimated using the EB (see Algorithm 1). Two flavors of the normal approximation are shown. One has the body/tail partition parameter p (see Section 2.2) set at 0.99, and the other at 0.95.

Table 1 shows a good agreement between the empirical VaRs and the two flavors of normal approximation VaR. In most cases, the 99% flavor agrees marginally better. The empirical VaRs are accurate to approximately 1%. In many cases, the normal approximation VaRs fall within the 95% confidence intervals indicated. When they do not, they tend to exceed the upper confidence limit. The normal approximation VaRs are, therefore, more conservative than the empirical VaRs, which is an advantage from a prudential point of view. There are two notable exceptions: LL1 and G1, for which the normal approximation VaRs are underestimates. The loglogistic and Fréchet distributions are not commonly used in capital calculations; thus, while underestimates are not worrying, they should be noted.

We stress that the principal purpose of the CLT alternative to EB is to derive an approximation that is exactly that: an approximation. As such, it is a viable alternative to the EB, and may be used to validate VaR calculated from a fitted distribution. The EB and the CLT VaR values should agree within tolerable limits, despite any concerns due to Berry–Esseen considerations. In this context, such tolerance is a matter of experience and judgement, but the two calculations should certainly not be an order of magnitude apart. No normal approximation VaR in Table 1 differs from its empirical VaR equivalent by more than 20%, which is a sufficient tolerance for practical purposes.

TABLE 1 Capital values for configuration distributions in € (millions).

Distribution	Empirical VaR	95% confidence interval for empirical VaR	Normal approximation VaR (99%)	Normal approximation VaR (95%)
LN1	23.18	(22.33, 23.89)	23.51	26.39
LN2	0.40	(0.39, 0.41)	0.39	0.41
LN3	97.13	(92.34, 101.19)	104.21	111.87
GPD1	0.74	(0.73, 0.75)	0.72	0.74
GPD2	13.68	(13.45, 13.77)	13.13	13.65
GPD3	3367.99	(3192.25, 3551.95)	3584.26	3580.09
GPD4	82.54	(79.72, 85.67)	87.66	93.21
WB1	221.59	(212.7, 232.19)	227.74	254.32
WB2	30.08	(29.18, 30.57)	29.99	33.52
WB3	101.36	(99.7, 103.02)	99.84	105.89
LNMix1	10.98	(10.8, 11.12)	10.95	11.60
LNMix2	282.57	(278.94, 286.66)	276.43	292.82
GH1	2.24	(2.14, 2.3)	2.31	2.59
GH2	8.46	(8.04, 8.85)	8.78	9.97
LG1	0.17	(0.16, 0.18)	0.19	0.19
LG2	34.13	(32.83, 35.88)	30.07	28.78
B1	7084.26	(6632.76, 7518.45)	7689.92	7189.44
B2	336.11	(318.38, 353.92)	355.46	377.36
LL1	8422.43	(8417.13, 8425.34)	7128.96	6545.38
LL2	1.67	(1.58, 1.76)	1.92	1.90
G1	1142.59	(1134.32, 1151.17)	1008.79	1024.19
F1	133.54	(122.1, 142.6)	141.21	132.12

2.5 Cases where VaR is not always normally distributed

Outliers in the data can result in VaR distributions that are clearly nonnormal. Consider the case where one loss, l_{max}, is very much greater than all the others. The VaR value will either include l_{max} or not. If it does not, VaR will be very small compared with l_{max}. If the VaR value contains l_{max}, VaR will be of comparable size to l_{max}. If the VaR value contains two instances of l_{max}, the VaR will be much greater than l_{max}. Further cases can be noted where there are more instances of l_{max} as part of the VaR value. Overall, this gives rise to a multimodal population.

In the numerical examples that follow in subsequent sections, twelve sets of loss data are used. The following example uses the first of them (INT_1) to illustrate the nonnormal distribution phenomenon. INT_1 has one huge loss of €31.15 million that is significantly larger than all other losses. Indeed, this one loss is larger than

FIGURE 3 Three-part figure comparing various modes of VaR distribution.
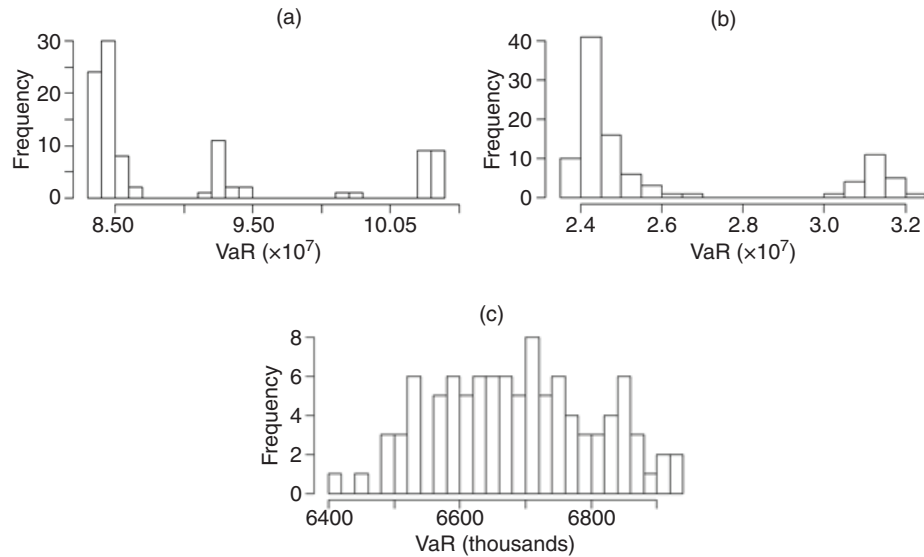
(a) Trimodal VaR distribution: all data. (b) Bimodal VaR distribution: largest loss removed. (c) Unimodal VaR distribution: largest two losses removed.

the sum of all the other losses. The second largest loss is €9.59 million. There are 908 losses in all, and the mean loss is €69 610. INT_1 is unremarkable apart from its largest two losses. If all losses are included and Algorithm 1 is applied, a histogram of 100 identical VaR calculations clearly shows that the distribution is not normal (see Figure 3(a)).

For the full set of data included in INT_1, the VaR obtained was approximately €90 million. This figure is approximately three times the value of the maximum loss, indicating that it is likely that the VaR figures contain two (perhaps even three) instances of the maximum loss.

If the maximum loss of €31.15 million is removed so that there is a new maximum loss of €9.59 million, the equivalent histogram is bimodal (as in Figure 3(b)). The VaR value in this case is approximately €26 million. Again, it is likely that two or perhaps three instances of €9.59 million are included in the VaR figure.

It is not until the largest two losses are removed that the VaR distribution appears normal (as in Figure 3(c)). In this case, VaR is approximately €7 million.

It is clear that the existence of outliers has a dramatic effect on the VaR distribution and, more importantly, on the VaR calculated.

2.6 Validation: sensitivity to model parameters

Investigating the sensitivities of model outputs to input parameters and data is an important aspect of validation. In the context of operational risk capital calculations, this causes particular problems because of (1) the number of parameters involved and (2) the length of time it takes to complete sufficient Monte Carlo iterations (in an LDA process) for one to have confidence in the result.

Each complete Monte Carlo calculation can be very time consuming, especially if a large number of operational risk losses are involved. The calculation may take several hours to achieve a standard 1% confidence for the calculated VaR value. That makes it very difficult to validate the essential parameters, and even more difficult to do extra data validations. Using a mixture of internal and external data and then adding scenarios introduces parameters that determine the relative mix of these three components in the LDA calculation. Further, higher-value losses and scenario values have a significant effect on the calculated VaR, and validation runs are required to assess the effects of each of them. The CLT method makes it very quick and easy to do all of these validations, because the calculations it entails are practically instant. A quick alternative to the EB, the proposed CLT method makes sensitivity analysis easily feasible in all cases. Time dependence is almost not applicable for the proposed CLT method. It involves calculations of means and standard deviations as well as combinations of them, all of which are fast. The slowest step is a data read. The slowest step of all in the LDA procedure, the Monte Carlo simulation, is absent from the CLT method.

2.7 Validation: regulatory aspects

Validation of the operational risk capital calculation has recently come under scrutiny from the Federal Reserve Board (Fed). Their Comprehensive Capital Analysis and Review (CCAR) requires banks to project operational risk losses under stressed scenarios. The Fed has issued recent guidance (Curti *et al* 2016) to the effect that benchmarking should include an EB calculation. As a result, a way to check what may become a new regulatory requirement is an essential tool in the validation process.

3 NUMERICAL RESULTS FOR SEPARATE INTERNAL AND EXTERNAL LOSS DATA

The calibration discussed in the previous section gives some confidence that the normal approximation to the EB provides a reasonable estimate of minimum VaR

TABLE 2 Capital values for internal data in € (millions).

UoM	Empirical bootstrap VaR	Normal approximation VaR 95%	Normal approximation VaR 99%
1	108.1	93.1	89.5
2	70.4	75.3	70.0
3	15.0	16.8	15.5
4	3.1	3.1	3.0
5	678.7	530.9	528.1
6	1.2	0.8	0.8
7	0.8	0.5	0.4
8	1.7	1.4	0.6
9	64.1	42.5	43.8
10	8.2	6.5	6.5
11	80.6	103.5	93.3
12	0.6	0.5	0.5
Total	1032.6	874.8	852.0

for the test data sets used. In this section, we apply the same method to actual loss data. We could regard an actual loss data set as having some sort of combination of characteristics (as yet undefined) of the test data sets. If so, applying Algorithm 1 with the combined indicator from (2.9) may be expected to yield satisfactory results.

Table 2 shows the EB VaR and the normal approximation to it (of the 99% flavor) for the internal data for twelve units of measure (UoMs).

Informally summarizing the results from Table 1, the normal approximation tends to underestimate VaR with respect to the EB. If the VaR calculated by either method is very small, the differences in percentage terms tend to be greater, but both are immaterial in absolute terms (that is, when compared with the totals). The normal approximation can therefore be regarded as less conservative than an LDA analysis. As such, the normal approximation can be regarded as more applicable for the case of regulatory purposes. This is particularly so in the case of INT_5, which is worth more than 50% of the total VaR. This UoM has a small number of huge losses with thousands of very small ones, and it is thought that the huge losses have an undue influence on all of the VaR calculations in this case. Hence, empirical VaR values for INT_5 are probably overestimates. The normal calculation also has the advantage of being purely deterministic. This makes it fast.

For validation purposes, a general delivery of a lower "minimum VaR" (less than the EB VaR) should be regarded as a more stringent "minimum VaR" criterion when

TABLE 3 Capital values for external data in € (millions).

UoM	Empirical bootstrap VaR	Normal approximation VaR 95%	Normal approximation VaR 99%
1	320.1	357.9	310.3
2	978.5	1102.9	1012.2
3	17.0	16.1	15.8
4	770.1	827.3	777.7
5	2750.1	3556.6	3097.1
6	2714.3	3547.1	3138.5
7	106.9	118.6	104.3
8	105.9	110.7	102.3
9	1189.3	1358.5	1230.4
10	484.1	552.1	492.2
11	414.9	474.5	422.4
12	408.5	462.6	404.0
Total	10 259.8	12 484.9	11 107.1

calculating VaR via a fitted distribution. The standard practice when using EB VaR has been to build in a subjective allowance such that if VaR from a fitted distribution is "about 20%" below the EB VaR, the fitted distribution VaR would be acceptable. This 20% criterion is subjective, and it could be abandoned if the CLT VaR is used.

The corresponding results for external data, derived from the Operational Riskdata eXchange Association (ORX) website (http://bit.ly/2oi9Hgk), are shown in Table 3. External losses are much higher than internal losses and, as expected, both the EB and the normal approximation results are greater than their corresponding internal results. In some cases, overestimation of VaR by the normal approximation is apparent, particularly in the 95% case.

In the context of external data, the question of how to treat outliers becomes particularly significant. All of the VaR values in Table 3 were calculated having removed outliers according to the u-sigma rule (1.1). This process may be challenged when validated, but removal of outliers has been "unofficially" sanctioned recently by Europe's Basel banking authority (Farkas 2015).

With external data, the pattern of overestimating EB VaR has the opposite effect on validation to that which is suggested with the internal data. Were the VaR calculated from it to be considered "too low", one would have to be more cautious in rejecting a fitted distribution. Usually, external data is not processed in isolation and so is not validated separately.

4   EXTENSION TO COMBINED INTERNAL AND EXTERNAL LOSS DATA

The results of Section 2.2 will now be applied to the case where two data sets, assumed to be independent, are available for a VaR calculation. Under Basel II rules, internal and external data may be combined (Basel Committee on Banking Supervision 2011). The same regulations permit the inclusion of scenarios, and these will be added to the model in Section 5. Internal and external data sets have been analyzed separately in the preceding section; our proposal to combine the two in a normal approximation model only amounts to considering a linear combination of the underlying random variables for the corresponding internal and external data sets. Minimal further theory is required. The procedures that follow in this section and in Section 5 (in which scenarios are considered) employ a uniform linear combination approach to combining data. Such uniformity is not always applicable to the LDA calculation, where internal data may be combined with other data in many ways (through maximum likelihood estimates, for instance). Linear combinations of data statistics in the CLT method make intuitive sense.

In the discussion that follows, the subscripts "I" and "E" will refer to "internal" and "external", respectively. As in the previous section, the symbols μ and σ^2, with the appropriate subscripts, will denote unknown means and variances, respectively. The sample size n in the context of empirical VaR refers to the size of the empirical population, from which random samples are drawn with replacement. The notation for VaR, $V_{t(c)}$, becomes rather clumsy once subscripts are added to refer to internal and external data, so the "$t(c)$" subscript will be dropped, and we will simply use V for VaR.

For internal data,

$$V_I \sim N(n_I \mu_I, n_I \sigma_I^2). \tag{4.1}$$

Similarly, for external data,

$$V_E \sim N(n_E \mu_E, n_E \sigma_E^2). \tag{4.2}$$

In practice, the means and variances may be estimated from the unbiased estimators of the means and variances for the empirical internal and external data, denoted in the following equations by the usual "hat" notation. Replacing the expressions for V_I and V_E, we get

$$V_I \sim N(n_I \hat{\mu}_I, n_I \hat{\sigma}_I^2), \tag{4.3}$$

$$V_E \sim N(n_E \hat{\mu}_E, n_E \hat{\sigma}_E^2). \tag{4.4}$$

The mean and variance for the internal data in (4.3) correspond with the combined body/tail mean and variance derived in (2.3a). The mean and variance for external data in (4.4) can be derived in the same way.

Assuming that the internal and external populations are independent, we can form a linear combination of V_I and V_E as follows, using a predetermined weight $w \in (0, 1)$. The numerical value of the weight should be such that the bias is largely in favor of the internal data. Very often, external losses are significantly higher and more frequent than internal losses, and they should not be allowed to influence the overall VaR calculation unduly. A method of determining the weight w, based on credibility theory (Bühlmann and Straub 1970), is given in online Appendix A, although any other objective method might be used:

$$V = wV_I + (1 - w)V_E. \tag{4.5}$$

V also has a normal distribution:

$$V \sim N(wn_I\hat{\mu}_I + (1 - w)n_E\hat{\mu}_E, w^2 n_I \hat{\sigma}_I^2 + (1 - w)^2 n_E \hat{\sigma}_E^2). \tag{4.6}$$

Therefore, VaR for the "internal plus external" combination may be estimated using the 99.9% quantile of V. In practice, however, a more conservative approach is recommended. VaR is not subadditive: given two random variables, A and B, the VaRs generated by each of them, and the VaR generated by their sum, are related by $\text{VaR}(A + B) \leqslant \text{VaR}(A) + \text{VaR}(B)$. Discussions on the implications of the nonsubadditivity of VaR may be found in both Desmedt and Walhin (2009) and Danielsson *et al* (2005). Calculating VaR for the linear combination V is therefore less conservative than evaluating VaR for V_I and V_E separately and weighting the resulting VaRs. The idea of separate evaluations of VaRs, followed by the weighting of results, can be extended further once scenarios are accounted for. Therefore, in the following sections, VaR due to scenarios is described as an "add-on".

4.1 Numerical results: internal/external data combination

Using (4.5), with weights for each UoM calculated using the credibility method in online Appendix A, the following results were obtained (see Table 4). The 99% versions for the normal approximations are shown. To calculate the VaRs for external data, the u-sigma rule outlined in (1.1) was applied, with $u = 0.674$, thus excluding the upper quartile of the external data. Higher values for u, in the range (1,3), produce VaR values that are about 10% higher, meaning they are more conservative.

A comparison of the results in the right-most columns of Tables 3 and 4 shows that, in general, the effect of introducing augmented internal data with external data is to increase the VaR. A much higher VaR for UoM 5 might reasonably be expected, but the influence of external data is limited by a credibility weight value that is near to 1.

Using a mix of external and internal data provides a further validation opportunity. With internal/external combinations, external losses should not dominate (provided there is sufficient internal data). If, using the same credibility weight as was calculated

TABLE 4   Capital values for external data in € (millions).

UoM	Credibility weight (w)	$wV_I + (1-w)V_E$
1	0.8789	116.2
2	0.9467	120.3
3	0.9621	15.6
4	0.9814	17.4
5	0.9990	530.7
6	0.9985	5.5
7	0.9976	0.7
8	0.9878	1.8
9	0.8405	233.1
10	0.9582	26.8
11	0.9834	98.7
12	0.9989	0.9
Total		1167.6

for the EB, the CLT approximation with both external and internal data is much larger than the CLT approximation with internal data only, then the calculated weight is suspect and should be investigated. For validation purposes, the percentage increase on calculating VaR when external data is added should be approximately the same for both CLT and EB calculations.

5 EXTENSION TO A SINGLE SCENARIO

The use of historic data (as in Sections 3 and 4) to estimate a future capital requirement must assume that past experience is an adequate predictor of future experience. This assumption is not always justifiable in the context of operational risk because changes to the regulatory environment, operational risk control and antifraud measures are common. A standard way to address this problem is through the use of scenarios. In the context of operational risk, a "scenario" is a qualitative estimate of the largest future loss that could be incurred within a given time horizon. Typical time horizons are five, ten, twenty or forty years. Scenarios effectively answer questions such as: What is the largest operational risk loss expected in the next ten years? Answers typically take the form of a single number, such as "£45.9 million". Scenarios are usually determined by teams of experts – there may be several per UoM – and cover several time horizons. Using scenarios in capital calculations is intended to provide a forward-looking view that accounts for future events, environment changes and procedural advances. As with the internal/external data combination, scenarios are

also combined with data statistics linearly. This practice preserves the consistency of the CLT method. It is not necessarily applicable to the LDA method, for which many methods may be used, such as treating scenarios as weighted losses in maximum likelihood estimation.

Having obtained an EB estimate of VaR using internal and external data, an add-on can be formulated to account for one or more scenarios per UoM. This section contains a discussion of the case of a single scenario and will be followed by an extension to many scenarios.

Appendix B (available online) suggests a way to calculate the scenario weight $w_S(y)$ for a single 1-in-y scenario value $S(y)$ (see equations (B1) and (B2) in Appendix B, available online). The precise method used is not central to this discussion. Having predetermined a scenario weight, the product $w_S(y) \times S(y)$ is simply added to the VaR calculated from the random variable V (as in (4.5)). This method should be seen as a calculation of a minimal scenario add-on, to match the minimality of the empirical VaR. Appendix C (also available online) contains details of a more sophisticated model.

5.1 Extension to many scenarios

Consider the case of a sequence of m scenario values, $S(y(1)), S(y(2)), \ldots, S(y(m))$, with respective time horizons $y(1), y(2), \ldots, y(m)$. Assume further that a corresponding sequence of scenario weights, $w_{S(y(1))}, w_{S(y(2))}, \ldots, w_{S(y(m))}$, has been determined. Then, each triple $\{S_{y(t)}, y(t), w_{S(y(t))} : 1 \leqslant t \leqslant m\}$ can be treated in the way described in the previous section, deriving variances $\tau_t^2 = (0.1 S_{y(t)}/1.96)^2$. Denoting the VaR for each scenario by $V_t(S(y(t)))$, the total scenario VaR add-on, V_S, is then

$$V_S = \sum_{t=1}^{m} w_{S(y(t))} V_t(S(y(t))). \tag{5.1}$$

Each term in this sum is positive and will thus increase the total VaR. However, there are limited merits to including any scenarios with time horizons that are less than the highest time horizon. All others are likely to be effectively represented by the external data. However, Dutta and Babbel (2014) make a valid point that scenarios should not be treated in the same way as data, since scenarios are forward looking and perform a different function. The add-on calculation for lower horizon scenarios should at least be done. It can be ignored, however, if it makes an insignificant difference to the overall VaR.

A practical point arises if a scenario value is generated without recourse to the external data. A rare scenario value (for instance, 1 in 40 or 1 in 50) is intended to represent an event, the severity of which exceeds the generality of external losses.

TABLE 5 Capital values for scenarios in € (millions).

UoM	1-in-40 scenario			1-in-10 scenario			Total scenario VaR V_S
	Weight	Value	Weight × Value	Weight	Value	Weight × Value	
1	0.0028	945.000	2.68	0.0040	77.00	0.31	3.00
2	0.0040	567.000	2.26	0.0020	123.00	0.24	2.50
3	0.0068	348.300	2.38	0.0459	5.40	0.25	2.63
4	0.0131	175.365	2.30	0.0962	3.30	0.32	2.61
5	0.0001	369.900	0.04	0.0007	403.00	0.27	0.31
6	0.0001	114.075	0.01	0.0014	194.00	0.27	0.28
7	0.0031	449.280	1.38	0.1389	0.18	0.03	1.40
8	0.0177	253.125	4.47	0.0066	41.50	0.27	4.74
9	0.0041	589.952	2.39	0.0098	27.60	0.27	2.66
10	0.0023	123.389	0.29	0.0749	4.00	0.30	0.59
11	0.0027	47.925	0.13	0.0052	52.00	0.27	0.40
12	0.0007	328.050	0.22	0.0334	8.50	0.28	0.50
Total			18.54			3.08	21.62

If this is not so, the suggested scenario is effectively modeling with the inclusion of external data. An alternative view is that external data that exceeds the scenario value is unrepresentative of losses that are likely (or even unlikely) to be incurred. We subscribe to the latter view by basing the scenario weight calculation on external data capped at the scenario value. This is particularly important when dealing with scenarios that have shorter time horizons (1-in-5 years, 1-in-10 years, etc).

5.2 Numerical results: scenarios

The results in Table 5 were obtained using the 1-in-40 and 1-in-10 scenarios for the same twelve UoMs of Tables 1–3. The VaRs due to scenarios were calculated using (5.1). The figures in the "1-in-y" columns are given without their weight multipliers, which are recorded in other columns instead. This gives some idea of comparability with the internal and external VaRs in Tables 1 and 2, respectively. The actual VaR to be added on is then the product of the 1-in-y values and their corresponding weights.

6 NUMERICAL RESULTS: COMBINED AND WEIGHTED VAR

Table 6 shows the combined and weighted VaRs, V, using the results from Tables 4 and 5. Given a credibility weight w and a scenario weight w_S, the overall weighting

TABLE 6   Combined capital values in € (millions).

UoM	Credibility weight (w)	wV_I	$(1-w)V_\mathrm{E}$	$wV_\mathrm{I} + (1-w)V_\mathrm{E}$	Scenario VaR (V_S)	Total VaR (V)
1	0.8789	78.7	37.6	116.2	3.0	119.2
2	0.9467	66.3	53.9	120.3	2.5	122.8
3	0.9621	15.0	0.6	15.6	2.6	18.2
4	0.9814	2.9	14.5	17.4	2.6	20.0
5	0.9990	527.6	3.1	530.7	0.3	531.0
6	0.9985	0.8	4.7	5.5	0.3	5.7
7	0.9976	0.4	0.3	0.7	1.4	2.1
8	0.9878	0.6	1.2	1.8	4.7	6.5
9	0.8405	36.8	196.2	233.1	2.7	235.7
10	0.9582	6.2	20.6	26.8	0.6	27.3
11	0.9834	91.7	7.0	98.7	0.4	99.1
12	0.9989	0.5	0.4	0.9	0.5	1.4
Total		827.4	340.2	1167.6	21.6	1189.2

scheme used is

$$V = wV_\mathrm{I} + (1-w)V_\mathrm{E} + w_s V_\mathrm{S}. \tag{6.1}$$

The weights w are shown in Table 4. Table 6 shows a slightly different view of the composition of the total VaR presented in Tables 4 and 5. It sets out the weighted internal and external VaRs for comparison with the weighted scenario VaR. The validator should compare these values for each UoM in order to detect any anomalies. For example, for UoMs 4 and 8, the weighted external VaR is much greater than the weighted internal VaR. This disparity might then be investigated further and possible causes explored. In the case of UoM 9, the credibility weight is relatively low. For UoM 4, the internal plus external VaR is low compared with the total VaR (for all UoMs); thus, if an underlying problem exists, it is unlikely to be material.

6.1 Comparison with results based on fitted distributions

The primary purpose of this discussion is not to formally compare capital values derived from an EB with the generality of calculation methods that use fitted distributions. It is worth reiterating the point that the EB method should provide a minimum estimate of capital. Therefore, a smaller capital value derived from any other method should be investigated and a resolution sought. Dutta and Perry (2007) give a summary comparison of the EB method with calculations based on the g-and-h distribution. They state that the 25th percentile and the median of the empirical estimates are about

TABLE 7   LDA and OpVision-fitted capital values in € (millions).

UoM	Total LDA VaR (V)	VaR using distributions fitted using OpVision
1	119.2	112.1
2	122.8	292.7
3	18.2	25.5
4	20.0	71.8
5	531.0	387.9
6	5.7	142.8
7	2.1	4.6
8	6.5	23.4
9	235.7	266.2
10	27.3	10.8
11	99.1	106.9
12	1.4	55.4
Total	1189.2	**1500.1**

one-half and one-third of the g-and-h estimates, respectively. The 75th percentiles of the empirical estimates are greater than the corresponding g-and-h estimates. Overall, they stress the dependence of the EB on the sampling method used.

Summary results obtained using distributions fitted to the twelve UoMs using the OpVision application are shown in Table 7. This application allows the user to choose a severity and a frequency distribution, fits data to those distributions and calculates a goodness-of-fit measure. It also calculates capital using the convolution method of Frachot *et al* (2001).[1]

It should be noted that the VaR values calculated by fitting a distribution to data are greater than the corresponding theoretical minimums (the LDA totals), except in the case of UoMs 1, 5 and 10. In the case of UoMs 6 and 12, the fitted VaRs are substantially greater than the minimum VaR. In both cases, the discrepancy is a signal to investigate further. Having two calculation methods for the "minimum VaR" helps the validation process. If both are greater than the VaR calculated from a fitted distribution, then the conclusion that the latter is "too low" is strengthened. Major discrepancies are often due to the influence of external data and to the modeling threshold set in the fitting process. Using high modeling thresholds has the effect

---

[1] For further details of the software used to generate these results, see OpVision's website: www.qrr.es/products/opvision.

of reducing frequency in the LDA process by cutting out a significant number of low-value losses, while retaining the bulk of the total loss. We have often observed this phenomenon. In extreme cases, removing 95% of the lowest value losses results in the retention of 95% of the total loss in the remainder (that is, the 95% that were removed comprised only 5% of the total value). Raising the modeling threshold often depresses the calculated VaR value. In cases where the fitted distribution produces an "excessively high" VaR value, very high scenarios were responsible for inflating the values due to data alone.

From a validation viewpoint, an inflated VaR after the addition of scenarios could be taken as an indicator, alerting us to the fact that a scenario value is out of line with the data. Although a scenario value can be compared directly with a maximum datum, for example, it is sometimes only after the VaR has been calculated that such problems become apparent.

6.2 Summary of validation points

In this section, we summarize positive reasons for introducing the CLT method as a validation tool. They are as follows.

- CLT calculations are quick to complete and require simple initial calculations. The Monte Carlo calculations involved in the EB can be very time consuming. A crude comparison is between an instant CLT calculation and an hour-long EB requiring more than a million iterations.

- The CLT method can be used to investigate the sensitivity of VaR to outliers, weights, system parameters and statistics of the data. In the past, we have used many weeks of overnight runs to produce heat maps as part of an extensive validation process. Using the CLT, that work could have been done in hours.

- With increased regulatory interest in the calculation of operational risk capital, a method allowing us to validate a bootstrap calculation is a useful weapon.

- The CLT method can be used as a tool to detect fitted distributions that produce VaR results that are either too small (because they are less than a theoretical minimum) or too large (because they are from a particular class of distributions that is known to produce excessively high VaRs).

7 CONCLUSION

The aim of this paper was to formulate a realistic process that would be an alternative to the EB method of estimating minimal capital value. The proposed normal approximation agrees well with the EB results and has certain advantages over it. They are as follows.

- The normal approximation is an independent method of validation.

- Sampling is not required as the method is not stochastic. The normal approximation is fast: the rate-determining step is the data read.

- It is a unified methodology: the same methods may be used to estimate a component of VaR based on internal data, external data and scenarios.

- The normal approximation method does not depend on modeling a loss distribution. In particular, the problem of grossly inflated VaR does not arise if a GP (or similar) distribution is used.

- The normal approximation produces a more conservative (that is, a higher) minimal capital than the EB in most cases.

- The normal approximation allows for fast approximation of VaR at different percentiles to inform on an institution's risk appetite considerations.

We stress that the processes described in this paper – either the EB or the normal approximation – tell us nothing about what the actual capital values should be. They only provide minimum values. The normal approximation is capable of being configured to estimate the capital value itself (as opposed to the minimal capital value) by choosing the appropriate parameters, which is a further advantage of using it.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

Abad, P., Muela, S. B., and Martín, C. L. (2015). The role of the loss function in value-at-risk comparisons. *The Journal of Risk Model Validation* **9**(1), 1–19 (https://doi.org/10.21314/JRMV.2015.132).

Balta, E., Menéndez, S. C., de Fontnouvelle, P., Embrechts, P., Ergashev, B., Hamidieh, K., Kiefer, N. M., and Swenson, K. (2009). Operational risk: modeling the extreme. White Paper, Office of the Comptroller of the Currency–National Institute of Statistical Sciences. URL: http://bit.ly/2HxAahZ.

Basel Committee on Banking Supervision (2011). BCBS 196 – Operational risk: supervisory guidelines for the advanced measurement approaches. Report, June, Bank for International Settlements.

Board of Governors of the Federal Reserve System (2014). Supervisory guidance for data, modeling, and model risk management under the operational risk advanced measurement approaches. Basel Coordination Committee Bulletin 14–1, June, Board of Governors of the Federal Reserve System.

Böcker, K., and Klüppelberg, C. (2005). Operational VaR: a closed-form approximation. *Risk* **18**(12), 90–93.

Böcker, K., and Sprittulla, J. (2006). Operational VaR: meaningful means. *Risk* **19**(12), 96–98.

Bühlmann, H., and Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*, pp. 111–133

Cruz, M. G., Peters, G. W., and Shevchenko, P. V. (2015). *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook*, Section 7.1.4. Wiley (https://doi.org/10.1002/9781118573013).

Curti, F., Ergen, I., Le, M., Migueis, M., and Stewart, R. (2016). Benchmarking operational risk models. Finance and Economics Discussion Series 2016–070, Federal Reserve Board (https://doi.org/10.17016/FEDS.2016.070).

Danielsson, J., Jorgensen, B. N., Mandira, S., Samorodnitsky, G., and de Vries, C. G. (2005). Subadditivity re-examined: the case for value-at-risk. Working Paper 549, Financial Markets Group, London School of Economics and Political Science. URL: http://bit.ly/2CFXo1t.

Desmedt, S., and Walhin, J.-F. (2009). On the subadditivity of tail value at risk: an investigation with copulas. *Casualty Actuarial Society: Variance* **2**(2), 231–252.

Dixon, W. J. (1950). Analysis of extreme values. *Annals of Mathematical Statistics* **21**(4), 488–506 (https://doi.org/10.1214/aoms/1177729747).

Dutta, K., and Babbel, D. (2014). Scenario analysis in the measurement of operational risk capital: a change of measure approach. *Journal of Risk & Insurance* **81**(2), 303–334 (https://doi.org/10.1111/j.1539-6975.2012.01506.x).

Dutta, K., and Perry, J. (2007). A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. Working Paper 06–13, Federal Reserve Bank of Boston. URL: http://bit.ly/2HuaLWa.

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**(1), 54–77 (https://doi.org/10.1214/ss/1177013815).

Esseen, C. G. (1956). A moment inequality with an application to the central limit theorem. *Skandinavisk Aktuarietidskrift* **39**, 160–170 (https://doi.org/10.1080/03461238.1956.10414946).

Farkas, A. (2015). Stress test 2016. Lecture, Neue Entwicklungen der Bankenaufsicht: 16th Handelsblatt Jahrestagung, European Banking Regulation, Frankfurt.

Frachot, A., Georges, P., and Roncalli, T. (2001). Loss distribution approach for operational risk. Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais (https://doi.org/10.2139/ssrn.1032523).

Goldie, C. M., and Klüppelberg, C. (1998). Subexponential distributions. In *A Practical Guide to Heavy Tails*, Adler, R., Feldman, R. E., and Taqqa, M. S. (eds), pp. 435–459. Birkhauser, Boston, MA.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Annals of Mathematical Statistics* **21**(1), 27–58 (https://doi.org/10.1214/aoms/1177729885).

Guegan, D., Hassani, B. K., and Li, K. (2015). The spectral stress VaR (SSVaR). Working Paper, Centre d'Économie de la Sorbonne. URL: http://bit.ly/2oayuTy.

Hall, P., and Presnell, B. (1999). Biased bootstrap methods for reducing the effects of contamination. *Journal of the Royal Statistical Society* B **61**(3), 661–680 (https://doi.org/10.1111/1467-9868.00199).

Hannah, L., and Puza, B. (2015). Approximations of VaR as an extreme quantile of a random sum of heavy-tailed random variables. *The Journal of Operational Risk* **10**(2), 1–21 (https://doi.org/10.21314/JOP.2015.154).

Homolka, L. (2012). Extreme value theory and bootstrap approach for robust value at risk estimation under Basel II. In *Advances in Finance and Accounting: Proceedings of 1st WSEAS Conference on Finance, Accounting and Auditing*, Pavelkova, D., Strouhal, J., and Pasekova, M. (eds). World Scientific and Engineering Academy and Society Press.

Isserlis, L. (1918). On the value of a mean as calculated from a sample. *Journal of the Royal Statistical Society* **81**(1), 75–81 (https://doi.org/10.2307/2340569).

Jarque, C. M., and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters* **6**(3), 255–259 (https://doi.org/10.1016/0165-1765(80)90024-5).

Jarque, C. M., and Bera, A. K. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economic Letters* **7**(4), 313–318 (https://doi.org/10.1016/0165-1765(81)90035-5).

Jiménez, J. A., and Arunachalam, V. (2011). Using Tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk. *The Journal of Risk* **13**(4), 95–116 (https://doi.org/10.21314/JOR.2011.230).

Manzi, G., and Mecatti, F. (2008). Bootstrap algorithms for risk models with auxiliary variable and complex samples. *Methodology and Computing in Applied Probability* **11**(1), 21–27 (https://doi.org/10.1007/s11009-008-9072-8).

Mitic, P. (2017). Conduct risk: distribution models with very thin tails. In *SYMCOMP 2017: 3rd International Conference on Numerical and Symbolic Computation*, Loja, M. A., Barbosa, J. I., and Rodrigues, J. A. (eds). ECOMAS, Guimarães. URL: http://arxiv.org/abs/1705.06868.

Pfeffermann, D., and Correa, S. (2012). Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation. *Biometrika* **99**(2), 457–472 (https://doi.org/10.1093/biomet/ass010).

Pyle, D. (2003). *Business Modeling and Data Mining*. Elsevier.

Rao, C. R. (2001). *Linear Statistical Inference and Its Applications*, 2nd edn. Wiley.

Reiss, R. D. (1974). On the accuracy of normal approximation for quantiles. *Annals of Probability* **2**(4), 741–744 (https://doi.org/10.1214/aop/1176996617).

Shanmugam, R., and Chattamvelli, R. (2015). *Statistics for Scientists and Engineers*. Wiley (https://doi.org/10.1002/9781119047063).

Shevchenko, P. V. (2011). Loss distribution approach. In *Modelling Operational Risk Using Bayesian Inference*, Chapter 2, Section 2.8.2. Springer (https://doi.org/10.1007/978-3-642-15923-7_2).

Thompson, S. K. (2012). *Sampling*, 3rd edn. Wiley (https://doi.org/10.1002/9781118162934).

Wang, Z., Wang, W., Jin, Y., and Zhou, Y. (2012). Using BS–PSD–LDA approach to measure operational risk of Chinese commercial banks. *Economic Modelling* **29**(6), 2094–2103 (https://doi.org/10.1016/j.econmod.2012.06.031).

Weisstein, E. W. (1999). Curvature. *MathWorld: A Wolfram Web Resource*. URL: http://bit.ly/2sD2D2i.