



Research Paper

Various approximations of the total aggregate loss quantile function with application to operational risk

Ross Griffiths and Walid Mnif

TD Bank, 66 Wellington Street West, Toronto, ON M5K 1A2, Canada;
emails: ross.griffiths@td.com, walid.mnif@td.com

(Received June 13, 2016; revised September 14, 2016; accepted October 26, 2016)

ABSTRACT

A compound Poisson distribution is the sum of independent and identically distributed random variables over a count variable that follows a Poisson distribution. Generally, this distribution is not tractable. However, it has many practical applications that require the estimation of the quantile function at a high percentile, eg, the 99.9th percentile. Without loss of generality, this paper focuses on the application to operational risk. We assume that the support of random variables is nonnegative, discrete and finite. We investigate the mechanics of the empirical aggregate loss bootstrap distribution and suggest different approximations of its quantile function. Furthermore, we study the impact of empirical moments and large losses on the empirical bootstrap capital at the 99.9% confidence level.

Keywords: discrete compound distribution; total aggregate loss; quantile function; analytic approximation; extended Poisson; extended binomial.

1 INTRODUCTION

A compound Poisson distribution is the sum of independent and identically distributed random variables over a count variable that follows a Poisson distribution. Generally, its distribution is not tractable. However, it has many practical applications that require the estimation of the quantile function at a high percentile, eg, the 99.9th percentile. Without loss of generality, this paper focuses on its application to operational risk.

Operational risk is defined by the Basel Committee on Banking Supervision as follows:

The risk of loss resulting from inadequate and failed internal processes, people or systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.

The Basel II framework introduced the advanced measurement approach (AMA) to calculate regulatory capital for operational risk. AMA is a risk-sensitive methodology based on sophisticated mathematical tools. It offers a flexible framework to measure risk exposure and, therefore, to implement risk measurement and reporting tools.

AMA-regulated financial institutions are expected to collect and analyze loss data at a certain level of granularity to capture the idiosyncratic risks reflecting the complexity of their business activities. The most popular modeling concept by far is the loss distribution approach (LDA). This is mainly based on two parametric distributions: a single loss (or severity) distribution, and an annual frequency distribution of loss events, from which an annual aggregate loss distribution is then inferred. As operational loss data usually contains outliers, heavy-tailed distributions are often chosen to model loss severity. Calibrating these distributions to estimate the percentiles sought for regulatory capital thus becomes a challenge (Cope *et al* 2009), necessitating the use of benchmarks.

The empirical bootstrap distribution is widely used to benchmark the continuous compound distribution. It approximates the single loss distribution by the empirical distribution function arising from a sample. We investigate the mechanics of the empirical aggregate loss bootstrap distribution and suggest different approximations of its quantile function. Furthermore, we study the impact of empirical moments and large losses in the context of the empirical bootstrap measure of operational risk capital. Historical data is plagued with large outliers, which dominate the upper percentiles of the aggregate loss percentiles used to benchmark capital results. This paper offers a better understanding of the relationship between outlier losses and the empirical bootstrap distribution, particularly at high percentiles.

The paper is organized as follows. Section 2 defines the problem that we tackle, and introduces some notation. Section 3 presents a simple and intuitive analytical approximation of the quantile function of the empirical aggregate loss bootstrap distribution. Section 4 develops the multinomial representation of total loss random variable. This

representation is the key to computing various approximations (Poisson approximation, Cornish–Fisher expansion) that are developed in Section 5. We investigate their numerical efficiency in Section 6. Section 7 is devoted to the binomial approximation. Section 8 extends the analytic approximation developed in Section 3 to continuous distributions. Section 9 concludes.

2 DEFINITION OF THE PROBLEM

Consider an ordered sample of single losses,

$$L = \{x_1, x_2, \dots, x_n = M\} \quad \text{such that } x_i \leq x_j \text{ for all } i \leq j.$$

The aggregate loss is defined as

$$T = \sum_{i=1}^{f(t)} s_i, \quad (2.1)$$

where s_i is randomly sampled with replacement from L . Each individual loss x_i has a probability weight of being drawn equal to $1/n$. $f(t)$ is the frequency of loss over t years. Throughout this paper, it is assumed to follow a Poisson distribution with intensity λ .¹ Without loss of generality, we normalize the loss data sample L by its largest loss M :

$$\tilde{L} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n = 1\}, \quad \tilde{x}_i = \frac{x_i}{M}.$$

Define

$$\tilde{\mu}_k = \frac{1}{k} \sum_{i=1}^k \tilde{x}_i.$$

When $k = n$, $\tilde{\mu}_n$ is the sample mean of \tilde{L} . The normalized aggregate loss is

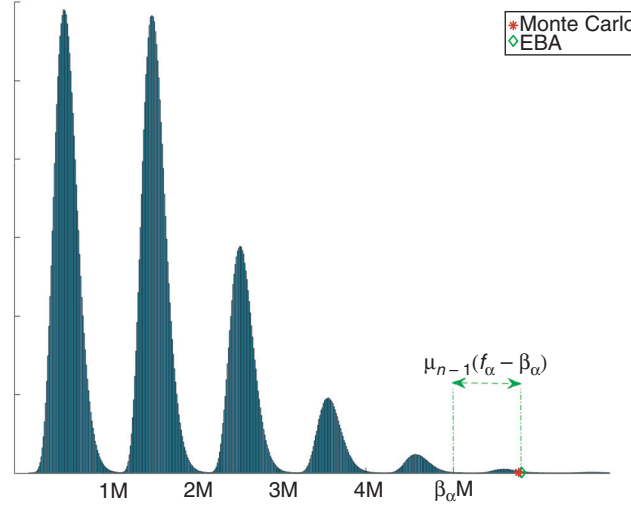
$$\tilde{T} = \frac{T}{M} = \sum_{i=1}^f \tilde{s}_i,$$

where \tilde{s}_i is randomly sampled with replacement from \tilde{L} .

Let us denote the cumulative density function (CDF) of \tilde{T} by F . We are interested in the approximation of the quantile function \tilde{T}_α of \tilde{T} , with a focus on samples with very large severity loss:

$$\tilde{T}_\alpha = \min \left\{ x = \sum_{i=1}^f \tilde{s}_i : F(x) \geq \alpha \right\}.$$

¹ The time horizon, t , is assumed to be constant. Hereafter, for clarity, we omit it.

FIGURE 1 Histogram of an empirical aggregate loss bootstrap distribution.

M is the largest single loss. β_α is defined in Section 3.3. In this example, α is equal to 99.9%. As expected, the empirical bootstrap analytic (EBA) approximation (Section 3) is efficient for type I samples. The binomial approximation (Section 7) estimates the empirical bootstrap distribution given that M appears precisely β_α times (with the remainder due to the contribution of smaller losses).

Approximating the quantile function \tilde{T}_α is challenging for a high percentile, α . In fact, \tilde{T} is a discrete distribution that exhibits clustering around multiples of large losses at the tail. The clustering is perceptible in the presence of extremely large losses (see Figure 1).

3 EMPIRICAL BOOTSTRAP ANALYTIC APPROXIMATION

We present various theoretical descriptions of the empirical bootstrap, first exploring some special boundary cases for which a precise formula exists and then developing a more general framework leading to multiple approximation formulas.

3.1 All losses are identical

The first extreme case is when all losses are equal to the maximum, ie, $\tilde{x}_1 = \dots = \tilde{x}_n = 1$. In this situation, the aggregate loss distribution, \tilde{T} , is identical to the frequency random variable, f . Indeed,

$$\tilde{T} = \sum_{i=1}^f \tilde{s}_i = \sum_{i=1}^f 1 = f. \quad (3.1)$$

Therefore,

$$\tilde{T}_\alpha = f_\alpha,$$

where

$$f_\alpha = F_\lambda^{-1}(\alpha) \quad (3.2)$$

and $F_\lambda^{-1}(\cdot)$ is the inverse CDF of the Poisson distribution with intensity λ .

3.2 One (or more) extremely large loss(es)

We consider the case where there is a single, isolated largest loss, with all other losses being much smaller. The extreme example corresponds to $\tilde{x}_1 = \tilde{x}_2 = \dots = \tilde{x}_{n-1} = 0$ and $\tilde{x}_n = 1$.² Let us generalize this situation slightly to cover the case of l identical largest losses and $n - l$ trivial losses, ie, $\tilde{x}_1 = \tilde{x}_2 = \dots = \tilde{x}_{n-l} = 0$ and $\tilde{x}_{n-l+1} = \dots = \tilde{x}_n = 1$. Note that variable \tilde{T} becomes a counting variable for the number of times the maximal loss is sampled in each compound draw. This is similar to (3.1); however, the expected count will be adjusted by the reduced probability of choosing a largest loss. A candidate for the resulting counting variable is

$$\tilde{T} \sim \text{Poisson}\left(\lambda \frac{l}{n}\right). \quad (3.3)$$

We now show that this is precisely the case. First note that the s_i are Bernoulli trials with probability of success equal to l/n . Therefore, we have

$$\tilde{T} = \sum_{i=1}^f \tilde{s}_i \implies \tilde{T} | f \sim \text{binom}\left(f, \frac{l}{n}\right).$$

The equivalence in (3.3) is due to the following lemma.

LEMMA 3.1 *Let $f \sim \text{Poisson}(\lambda)$ and suppose $\tilde{T} \in \{0, 1, 2, \dots\}$ is a discrete random variable such that the conditional distribution $(\tilde{T} | f = m) \sim \text{binom}(m, \pi)$.*

Then,

$$\tilde{T} \sim \text{Poisson}(\lambda\pi). \quad (3.4)$$

² We are generally interested in the problem when all losses are strictly greater than zero; however, this can be thought of as the limiting case characterizing $M \gg x_{n-1}$.

PROOF This is a well-known result that can be shown by simply computing the probability mass function for \tilde{T} at $k \in \{0, 1, 2, \dots\}$:

$$\begin{aligned}
 \Pr(\tilde{T} = k) &= \sum_{m=k}^{\infty} \Pr(\tilde{T} = k \mid f = m) \Pr(f = m) \\
 &= \sum_{m=k}^{\infty} \binom{m}{k} \pi^k (1 - \pi)^{m-k} e^{-\lambda} \frac{\lambda^m}{m!} \\
 &= e^{-\lambda} \frac{(\lambda \pi)^k}{k!} \sum_{m=k}^{\infty} \frac{[\lambda(1 - \pi)]^{m-k}}{(m - k)!} \\
 &= e^{-\lambda} \frac{(\lambda \pi)^k}{k!} \sum_{m=0}^{\infty} \frac{[\lambda(1 - \pi)]^m}{m!} \\
 &= e^{-\lambda} \frac{(\lambda \pi)^k}{k!} e^{\lambda(1 - \pi)} \\
 &= e^{-\lambda \pi} \frac{(\lambda \pi)^k}{k!}.
 \end{aligned}$$

□

The result for the first example (ie, for $l = n$) can be seen as a consequence of this when $\pi = 1$.

3.3 Analytic approximation

A key observation from the previous examples is that in each of these special cases the mean of the normalized data equals the probability of drawing a largest loss in a single trial. That is,

$$\tilde{\mu}_n = \frac{l}{n} = \Pr(\tilde{s}_i = 1).$$

$\tilde{\mu}_n$ can be written as

$$\tilde{\mu}_n = \frac{1}{n} \sum_{i=1}^{n-l} \tilde{x}_i + \frac{l}{n}, \quad 0 \leq \tilde{x}_i < 1.$$

Thus, $\tilde{\mu}_n \in [1/n, 1]$ and attains its minimum and maximum values in the extreme cases corresponding to $l = 1$ and $l = n$, respectively. Therefore, $\tilde{\mu}_n$ is a generalization of the probability of selecting a maximal loss in a single random draw. Given this, a candidate distribution for \tilde{T} in the general case, motivated by (3.4), is

$$\tilde{T} \sim \text{Poisson}(\lambda \tilde{\mu}_n).$$

This approximation has a few shortcomings: the first being that it is integer valued; the second is (as we shall see) that, while it does capture the correct mean, it overestimates the variance of the true distribution. We discuss these issues in later sections, where we also propose solutions. Meanwhile, we suggest another approximation of \tilde{T}_α , which is our main focus in the remainder of the paper.

For a fixed percentile level, α , simulation experiments (see, for example, Figure 3) suggest a linear relationship exists between \tilde{T}_α and $\tilde{\mu}_n$. Consequently, we linearly interpolate between the extreme values of $\tilde{\mu}_n$ and their corresponding known values of \tilde{T}_α (corresponding to the $l = 1$ and $l = n$ boundary cases discussed above). For the $l = n$ case, the value of \tilde{T}_α is given by (3.2); at the other extreme, $l = 1$, it is equal to the percentile, α , of the variable

$$\beta \sim \text{Poisson}\left(\frac{\lambda}{n}\right).$$

Then, the line between $(1/n, \beta_\alpha)$ and $(1, f_\alpha)$ is given by

$$\frac{n}{n-1} \left(\tilde{\mu}_n - \frac{1}{n} \right) (f_\alpha - \beta_\alpha) + \beta_\alpha.$$

We define

$$\tilde{\mu}_{n-1} = \frac{n}{n-1} \left(\tilde{\mu}_n - \frac{1}{n} \right) = \frac{1}{n-1} \sum_{i=1}^{n-1} \tilde{x}_i,$$

and then write the approximation to \tilde{T}_α as

$$\tilde{T}_\alpha \approx \tilde{\mu}_{n-1} (f_\alpha - \beta_\alpha) + \beta_\alpha.$$

Rewriting in terms of the nonnormalized distribution, L , we obtain

$$T_\alpha \approx \mu_{n-1} (f_\alpha - \beta_\alpha) + \beta_\alpha M, \quad (3.5)$$

where $\mu_{n-1} = \tilde{\mu}_{n-1} M$ (see Figure 1).

3.4 Remarks

The development of (3.5) implicitly relies on $\beta_\alpha > 0$, corresponding to

$$\alpha \geq F_{\lambda/n}(1). \quad (3.6)$$

This is generally true for high α and when the number of data points is less than $\lambda/(1 - \alpha)$.³ Otherwise, it turns out that the maximum loss is actually too rare to

³ For example, when $\alpha = 99\%$ and $\lambda = 10$, $\beta_\alpha > 0$ when the number of data points is less than $10 \times 1000 = 10\,000$.

TABLE 1 Coefficient of the maximum loss (β_α) for various percentiles (α) and numbers of time periods (N).

N	α	
	99%	99.9%
5	2	3
6	2	2
7	1	2
23	1	1
99	1	1
100	0	1
999	0	1
1000	0	0

contribute to the α percentile of the aggregate loss distribution. β_α appears as the coefficient of the maximum loss, M , in (3.5) and is defined as

$$\beta_\alpha = F_{\lambda/n}^{-1}(\alpha). \quad (3.7)$$

Note that if $\lambda = n/N$, where N is the number of time periods over which the sample was collected, then we have

$$\beta_\alpha = F_{1/N}^{-1}(\alpha)$$

and so this can be directly linked to the number of data collection periods involved in obtaining the sample. β_α is given for various values of α and N in Table 1.

Such information could be used to adjust empirical bootstrap estimates. For example, after collecting five periods of data, we might have reason to believe that the largest observed loss has an actual frequency of one in ten years rather than the observed one in five. From this, we conclude that a better empirical bootstrap analytic (EBA) estimate of the 99th percentile is obtained by adjusting the β_α parameter (ie, from 2 to 1).

4 MULTINOMIAL REPRESENTATION OF THE TOTAL LOSS DISTRIBUTION

In this section, we look at the distribution that governs the count, f^i , of a particular point, \tilde{x}_i , conditional on the total annual frequency, f . We have

$$\sum_{i=1}^n f^i = f, \quad f \sim \text{Poisson}(\lambda).$$

The experiment can be seen as drawing f independent trials from n categories. Each trial consists of choosing a loss with probability $1/n$. f^i is the number of trials in which x_i is drawn. Therefore, (f^1, \dots, f^n) has a multinomial distribution with index f and parameter $(1/n, \dots, 1/n)$. Steel (1953) shows that, conditional on f , a multinomial distribution is equivalent to the number of successes observed in n independent Poisson distributions with intensities λ_i , $i = 1, \dots, n$. λ_i is defined as the probability weight assigned to \tilde{x}_i multiplied by the total intensity, λ . Consequently,

$$f^i \sim \text{Poisson}(\lambda_i), \quad \lambda_i = \frac{\lambda}{n}.$$

The total loss random variable can be written as

$$\tilde{T} = \sum_{i=1}^n f^i \tilde{x}_i. \quad (4.1)$$

COROLLARY 4.1 *\tilde{T} is a linear combination of n independent Poisson distributions with intensities $\lambda_i = \lambda/n$, $i = 1, \dots, n$. It has mean $\bar{\mu}_1$, variance $\bar{\mu}_2$ and third central moment $\bar{\mu}_3$, where*

$$\begin{aligned} \bar{\mu}_1 &= \lambda \bar{\mu}_n, \\ \bar{\mu}_2 &= E[(\tilde{T} - \bar{\mu}_1)^2] = \frac{\lambda}{n} \sum_{i=1}^n \tilde{x}_i^2, \\ \bar{\mu}_3 &= E[(\tilde{T} - \bar{\mu}_1)^3] = \frac{\lambda}{n} \sum_{i=1}^n \tilde{x}_i^3. \end{aligned}$$

Corollary 4.1 uses the fact that a Poisson distribution has its first three central moments equal to its intensity. As $\tilde{x} \leq 1$, we have

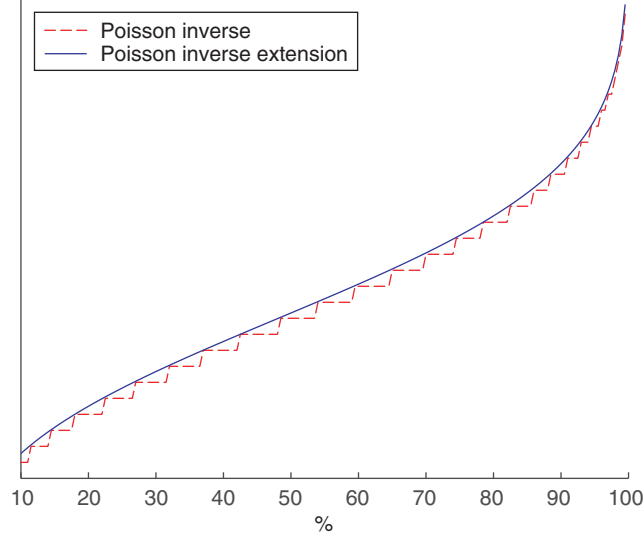
$$\bar{\mu}_2 \leq \bar{\mu}_1.$$

5 APPROXIMATION OF THE TOTAL LOSS DISTRIBUTION

5.1 Poisson approximation

Corollary 4.1 shows that \tilde{T} is a linear combination of independent Poisson distributions. \tilde{T} follows a Poisson distribution if and only if $\tilde{x}_i = 1$ for all $i = 1, \dots, n$, ie, $\bar{\mu}_n = 1$. Otherwise, $\bar{\mu}_1 > \bar{\mu}_2$, as there exists at least a normalized loss that is strictly less than 1. We suggest linearly mapping \tilde{T} into another random variable, Y , which has equal first and second central moments. Define

$$Y = \frac{\bar{\mu}_1}{\bar{\mu}_2} \tilde{T}.$$

FIGURE 2 Extension of the Poisson distribution over the positive real line.

Y has the same mean and variance, equal to

$$\lambda^* = \frac{\bar{\mu}_1^2}{\bar{\mu}_2}.$$

We approximate the Y distribution by an extension over the positive real line of a Poisson distribution with parameter λ^* . The Poisson distribution CDF is a stepwise function, which can be defined as

$$F_{\lambda^*}(x) = \frac{\Gamma(\lfloor x + 1 \rfloor, \lambda^*)}{\lfloor x \rfloor!} \quad \text{for all } x > 0, \quad (5.1)$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function and $\lfloor \cdot \rfloor$ is the floor function. We extend F_{λ^*} by a smooth envelope that matches its values over the natural numbers (see Figure 2). A straightforward generalization is to remove the floor function from (5.1) and replace the factorial by the gamma function, $\Gamma(\cdot)$:

$$x \mapsto \frac{\Gamma(x + 1, \lambda^*)}{\Gamma(x + 1)} \quad \text{for all } x > 0, \quad (5.2)$$

However, CDF (5.2) can be inaccurate on the corners. Therefore, we suggest the following generalization:

$$x \mapsto 1 - \frac{\gamma(x + 1, \lambda^*)}{\Gamma(x + 1)},$$

where $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function. In fact, CDF (5.1) can be written as a function of the χ^2 -distribution:

$$x \mapsto \Pr(\chi_{2(x+1)}^2 > 2\lambda^*)$$

(see Johnson *et al* (2005) for more details).

PROPOSITION 5.1 *The quantile function of \tilde{T} can be approximated as follows:*

$$\tilde{T}_\alpha \approx \frac{\bar{\mu}_2}{\bar{\mu}_1} (G_P^{-1}(\alpha, \lambda^*) + 1),$$

where

$$\begin{aligned} G_P: \mathbb{R}_+ \times \mathbb{R}_+ &\rightarrow [0, 1], \\ (x, y) &\mapsto 1 - \frac{\gamma(x+1, y)}{\Gamma(x+1)}. \end{aligned}$$

5.2 Cornish–Fisher expansion

The Poisson approximation matches the first two moments of \tilde{T} . We suggest approximating the discrete quantile function of the aggregate distribution by a continuous function using the Cornish–Fisher expansion. The latter is based on the availability of cumulants. We present the formula using the first three moments provided in Corollary 4.1. The quantile function is approximated by

$$\begin{aligned} \tilde{T}_\alpha &\approx \bar{\mu}_1 + \sqrt{\bar{\mu}_2} \delta(\alpha), \\ \delta(\alpha) &= \Phi^{-1}(\alpha) + \frac{\bar{\mu}_3}{\bar{\mu}_2^{3/2}} \frac{\Phi^{-1}(\alpha)^2 - 1}{6}. \end{aligned}$$

We provide a higher-order approximation of the Cornish–Fisher expansion in Appendix A online.

5.3 Panjer recursion

Panjer (1981) suggests an algorithm to compute the compound Poisson distribution using a recursive relationship between the probabilities of the number of losses that occur in the same time interval. The interval $[\tilde{x}_1, \tilde{x}_n]$ is split into m equispaced points. Let us denote the step size by $\Delta\tilde{x}$, and grid points by x_i^* , $i = 1, \dots, m$, with $x_1^* = \tilde{x}_1$, $x_m^* = \tilde{x}_n$ and $x_i^* = \tilde{x}_1 + (i-1)\Delta\tilde{x}$. Each point x_i^* is assigned a probability weight, p_i . The normalized aggregate loss \tilde{T} is approximated by

$$\begin{aligned} \tilde{T}_n &\approx x_0^* + D_m \Delta\tilde{x}, \\ D_m &= \sum_{i=1}^m i f_i^*, \quad x_0^* = \tilde{x}_1 - \Delta\tilde{x}, \end{aligned}$$

where f_i^* has a Poisson distribution with intensity λp_i . A simple probability scheme is to assign

$$p_1 = \frac{1}{n},$$

$$p_i = \frac{|\{\tilde{x}_j \in (x_{i-1}^*, x_i^*], j = 1, \dots, n\}|}{n}, \quad i = 2, \dots, m.$$

The probability mass function $g(\cdot)$ of D_m is computed using the following recursion:

$$g(0) = e^{-\lambda},$$

$$g(y) = \frac{\lambda}{y} \sum_{i=1}^m i p_i g(y-i) \quad \text{for all } y = 1, 2, \dots$$

The quantile function of D_m , $F_{D_m}^{-1}(\cdot)$, can be computed as follows:

$$F_{D_m}^{-1}(p) = \min \left\{ x, x = 1, \dots, \sum_{j=0}^x g(j) \geq p \right\}.$$

Therefore,

$$\tilde{T}_\alpha \approx x_0^* + F_{D_m}^{-1}(\alpha) \Delta \tilde{x}.$$

Different probability schemes are proposed in the literature, eg, methods of matching moments (Gerber 1982). The choice of $\Delta \tilde{x}$ and $p_i, i = 1, \dots, m$, is of paramount importance to obtain an efficient Panjer approximation.

6 NUMERICAL RESULTS

We compare the efficiency of each approximation with respect to a Monte Carlo estimate using 1000 000 simulations. Even though the Monte Carlo simulation embeds statistical errors, it remains a popular and reliable benchmark to estimate the quantile function.

We generate three sets of 5000 samples with sample size 100, 500 and 1000, respectively, from a lognormal distribution with an associated normal distribution of mean 0 and standard deviation 2. The losses are assumed to be observed in a five-year ($\lambda = n/5$) or ten-year ($\lambda = n/10$) time window.

We focus on estimating the 99.9th percentile for different frequencies and sample sizes. We classify samples into three categories.

Type I: sample has an extremely large loss (ie, $\tilde{\mu}_n \approx 1/n$).

Type II: all losses have the same order of magnitude (ie, $\tilde{\mu}_n \approx 1$; see Section 3.1).

Type III: otherwise.

TABLE 2 Average percentage deviation from Monte Carlo estimates for the 15 000 generated samples.

α (%)	EBA (%)	Poisson (%)	Fisher 3 (%)	Fisher 5 (%)	Panjer (%)
95	−1(8)	3(6)	1(3)	0(3)	5(4)
97.5	−10(7)	2(7)	4(6)	3(6)	5(4)
99.5	−13(11)	−4(7)	4(8)	4(13)	4(3)
99.9	1(7)	−9(5)	4(4)	4(13)	3(2)

Standard deviations are reported in parentheses. The collection periods are five and ten years.

Operational risk samples are usually of type I or III, as they generally contain at least one outlier. Therefore, we hereafter focus our numerical investigation on these two types.

Figure 3 compares EBA to Monte Carlo. By construction, EBA is efficient for sample types I and II. It is close to Monte Carlo elsewhere. Figure 4 depicts the Poisson approximation results. It behaves poorly for category I. Otherwise, it provides relatively efficient estimates. Cornish–Fisher approximates well the quantile function using three cumulants (see Figure 5). Expansion to five cumulants overestimates the 99.9th percentile for sample type I (see Figure 6). The empirical aggregate loss distribution is a discrete distribution, and it is fairly approximated with a continuous distribution by matching the first three cumulants. Figure 7 depicts the Panjer recursion for $m = 1000$. Panjer recursion is a numerical procedure that provides efficient results if it is well parameterized.

Table 2 reports the average percentage deviation from the Monte Carlo estimates for the 15 000 samples as well as their standard deviations for the 95th, 97.5th, 99.5th and 99.9th percentiles. On average, EBA seems to slightly underestimate the quantile function for low percentiles. Otherwise, other approximations behave well on average, except for the Poisson when estimating the 99.9th percentile.

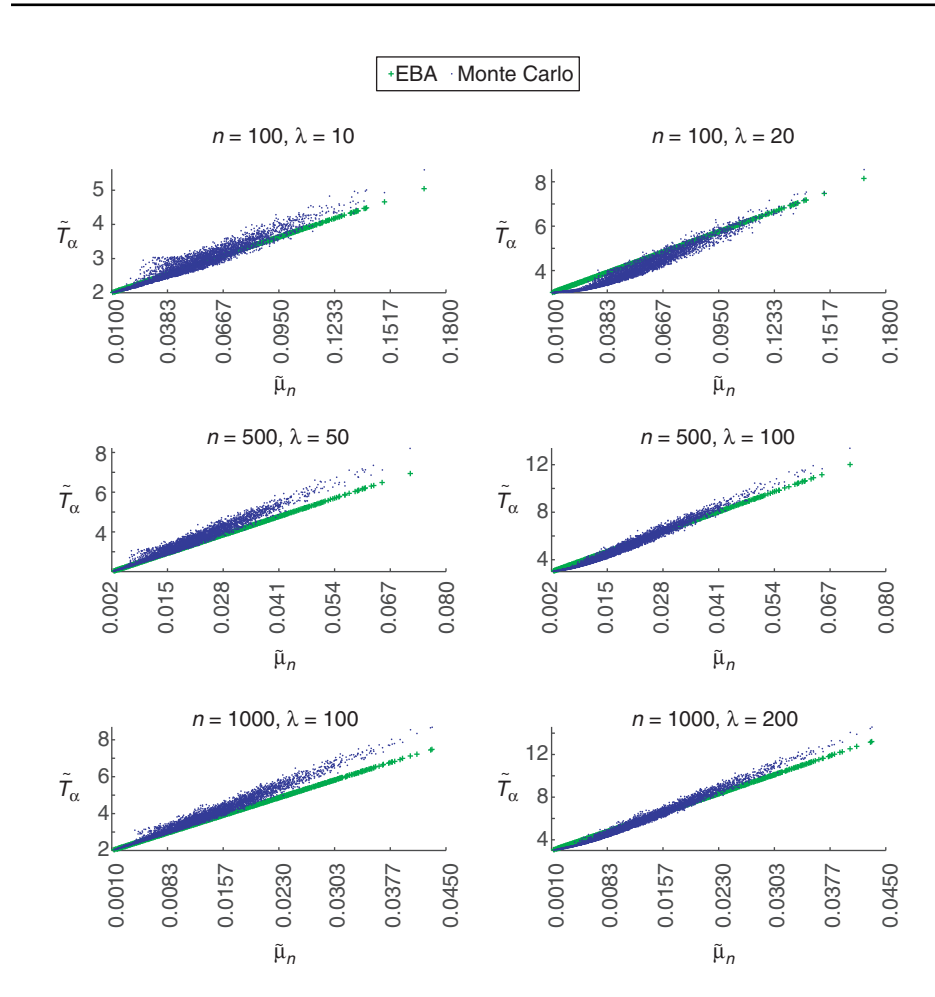
7 BINOMIAL APPROXIMATION FOR SAMPLES WITH A SINGLE LARGE LOSS

We focus on the distribution of \tilde{T} in the presence of an extremely large loss in the sample. We isolate the largest loss in (4.1):

$$\tilde{T} = \sum_{i=1}^{n-1} f^i \tilde{x}_i + f^n \tilde{x}_n. \quad (7.1)$$

At a high α , the largest loss is expected to appear at least β_α times (see Figure 1). Therefore, f^n is left-truncated to exclude $0, \dots, \beta_\alpha - 1$ as a possible outcome. Let

FIGURE 3 Empirical bootstrap analytic (EBA) approximation compared with Monte Carlo simulation at the 99.9th percentile.



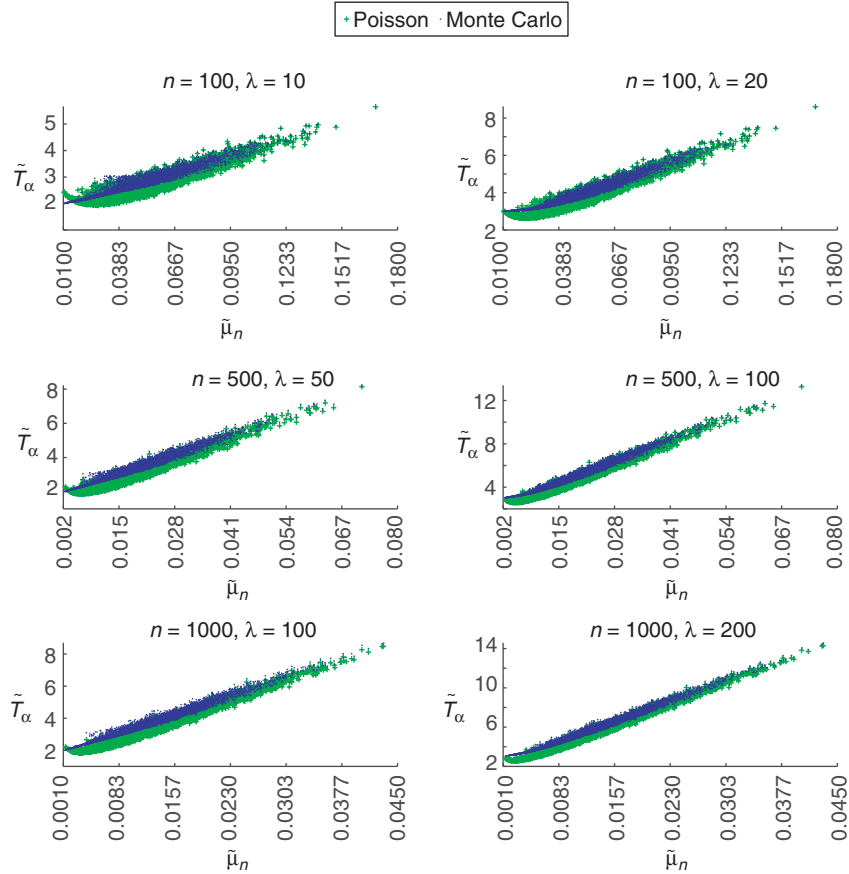
us denote its left-truncated distribution by f^{n*} . The probability mass function is

$$\Pr(f^{n*} = j) = h(j) \left(1 - \sum_{i=0}^{\beta_\alpha - 1} h(i) \right)^{-1},$$

where

$$h(i) = \frac{e^{-\lambda/n} (\lambda/n)^i}{i!}.$$

FIGURE 4 Poisson approximation compared with Monte Carlo simulation at the 99.9th percentile.



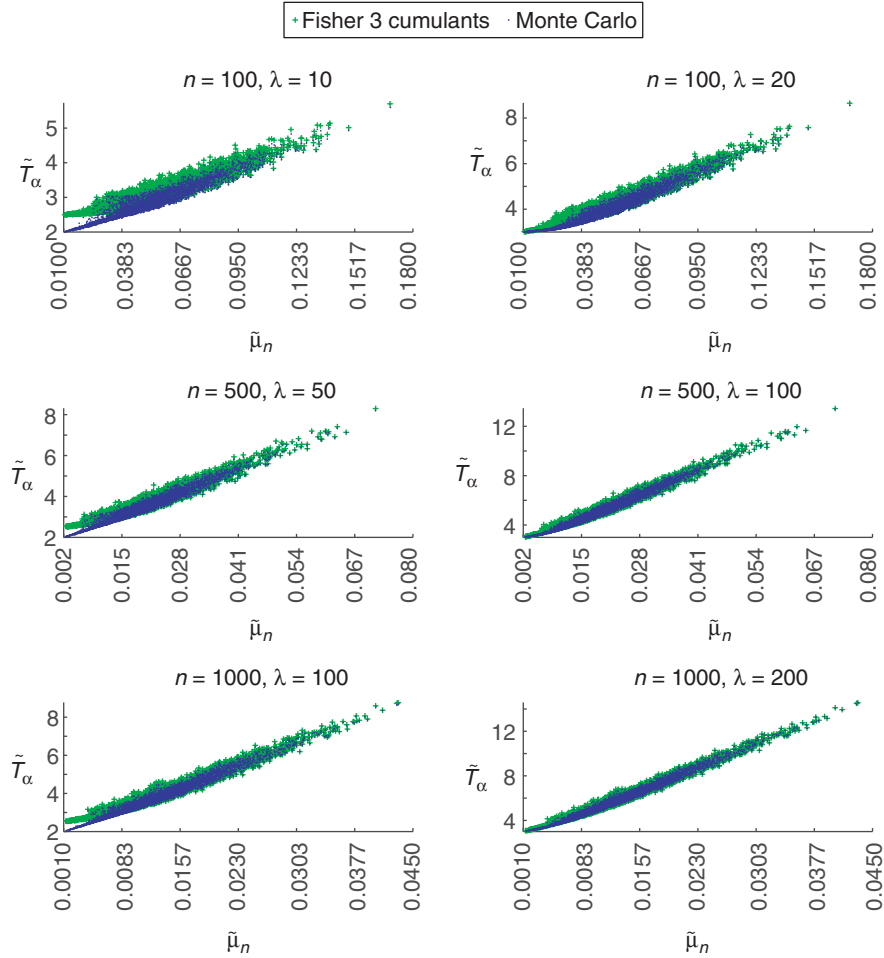
We can show that it has a mean

$$\theta_1 = \left(\frac{\lambda}{n} - \sum_{i=0}^{\beta_\alpha-1} i h(i) \right) \left(1 - \sum_{i=0}^{\beta_\alpha-1} h(i) \right)^{-1}$$

and variance

$$\theta_2 = \left(\frac{\lambda}{n} \left(\frac{\lambda}{n} + 1 \right) - \sum_{i=0}^{\beta_\alpha-1} i^2 h(i) \right) \left(1 - \sum_{i=0}^{\beta_\alpha-1} h(i) \right)^{-1} - \theta_1^2.$$

FIGURE 5 Cornish–Fisher approximation using three cumulants compared with Monte Carlo simulation at the 99.9th percentile.

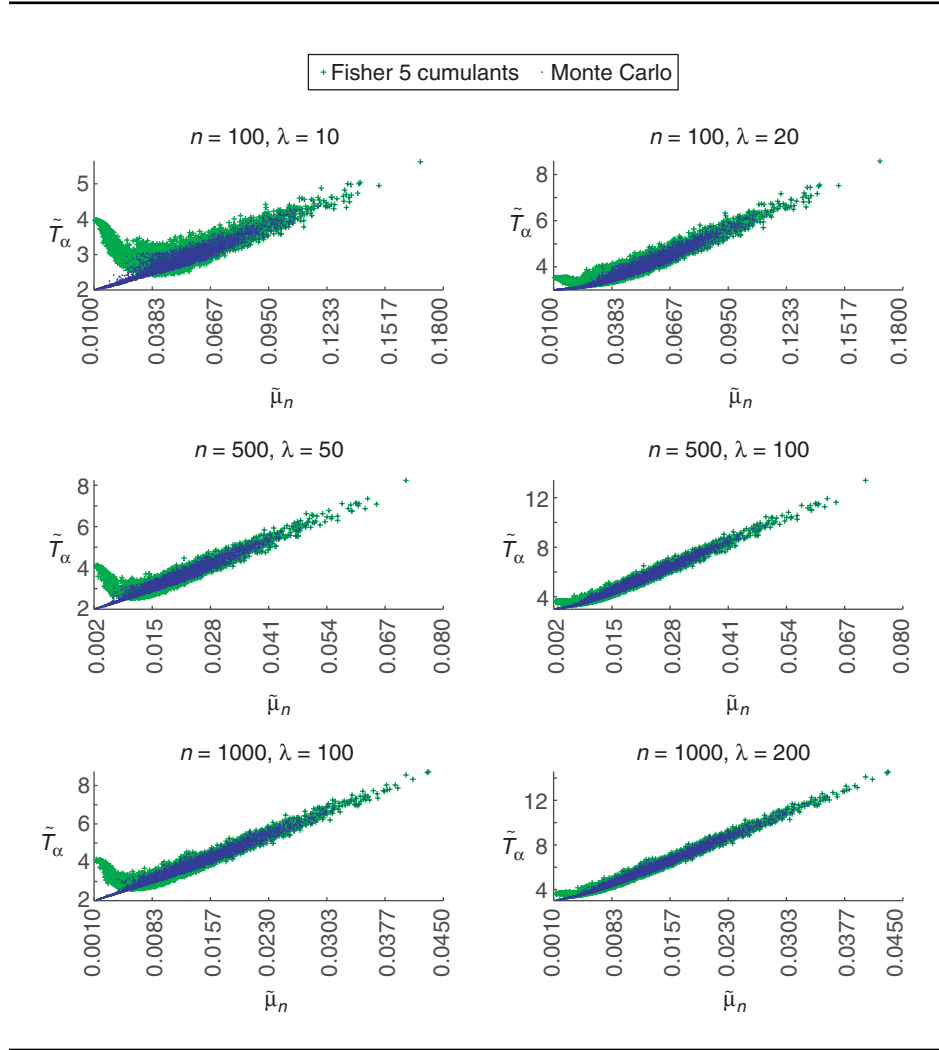


COROLLARY 7.1 Conditional on M being drawn β_α times, \tilde{T} has mean $\bar{\mu}'_1$ and variance $\bar{\mu}'_2$, where

$$\bar{\mu}'_1 = \frac{n-1}{n} \lambda \tilde{\mu}_{n-1} + \theta_1 \tilde{x}_n,$$

$$\bar{\mu}'_2 = \frac{\lambda}{n} \sum_{i=1}^{n-1} \tilde{x}_i^2 + \theta_2 \tilde{x}_n^2.$$

FIGURE 6 Cornish–Fisher approximation using five cumulants compared with Monte Carlo simulation at the 99.9th percentile.

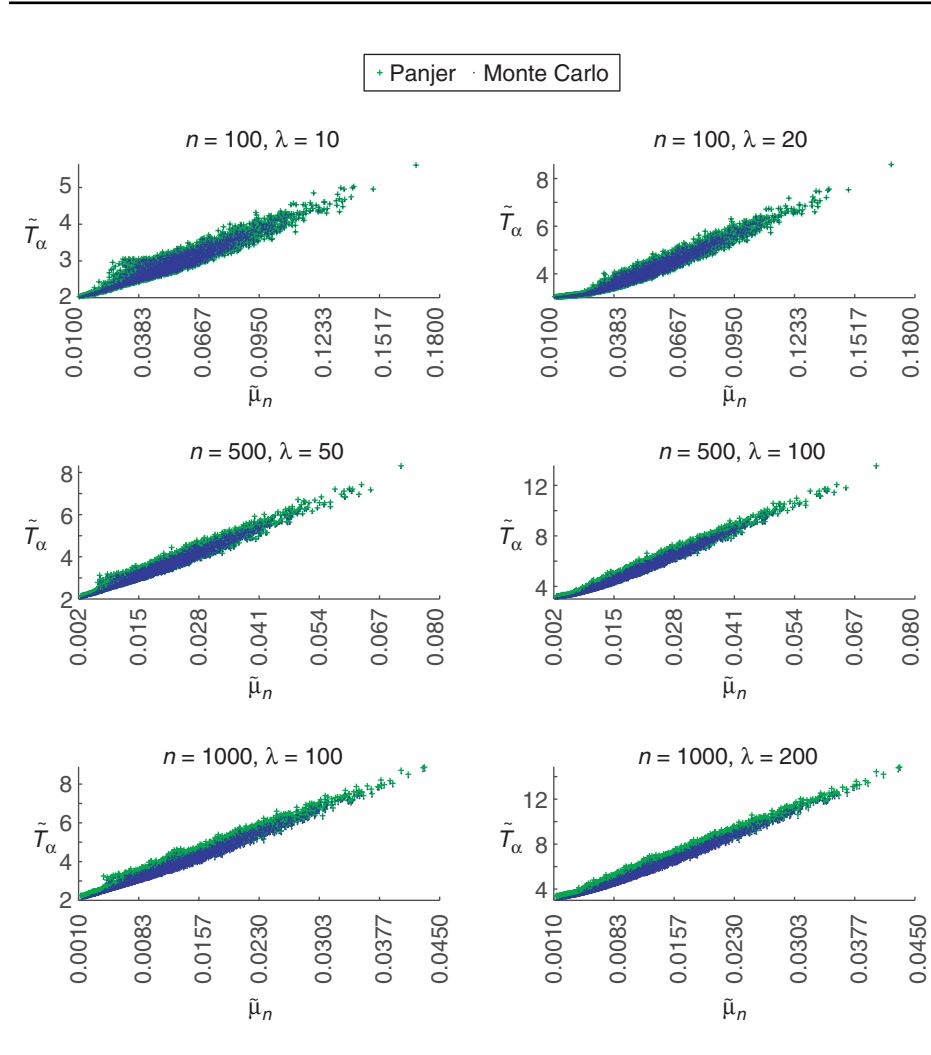


Left-truncation increases the mean and reduces the variance:

$$\theta_2 \leq \frac{\lambda}{n} \leq \theta_1.$$

Therefore, we have

$$\bar{\mu}'_2 \leq \bar{\mu}_2 \leq \bar{\mu}_1 \leq \bar{\mu}'_1.$$

FIGURE 7 Panjer recursion compared with Monte Carlo simulation at the 99.9th percentile.

We approximate the distribution of \tilde{T} by an extension over the positive real line of the binomial distribution with parameters (r^*, p^*) ,

$$r^* = \left\lceil \frac{\bar{\mu}'_1}{p^*} \right\rceil,$$

$$p^* = 1 - \frac{\bar{\mu}'_2}{\bar{\mu}'_1},$$

where $\lceil \cdot \rceil$ is the ceiling function. The Camp–Paulson approximation links the binomial distribution CDF, F_{p^*, r^*} , to the standard normal distribution CDF, Φ :

$$F_{p^*, r^*}(x) \approx \Phi(W), \quad (7.2)$$

$$W = -\frac{Y}{3\sqrt{(Z)}}, \quad (7.3)$$

$$Y = \left[\frac{\eta - xp^*}{(x+1)(1-p^*)} \right]^{1/3} \left(9 - \frac{p^*}{\eta - p^*x} \right) - 9 + \frac{1}{x+1}, \quad (7.4)$$

$$Z = \left[\frac{\eta - xp^*}{(x+1)(1-p^*)} \right]^{2/3} \left(\frac{p^*}{\eta - p^*x} \right) + \frac{1}{x+1}, \quad (7.5)$$

$$\eta = \frac{\bar{\mu}'_1}{p^*}. \quad (7.6)$$

Based on Johnson *et al* (2005), its maximum absolute error is less than $0.007(\bar{\mu}'_1(1-p^*))^{-1/2}$.

PROPOSITION 7.2 *The quantile function of \tilde{T} can be approximated as follows:*

$$\tilde{T}_\alpha \approx \Phi^{-1}(W),$$

where W is given through (7.3)–(7.6). $\Phi^{-1}(\cdot)$ is the standard normal inverse CDF.

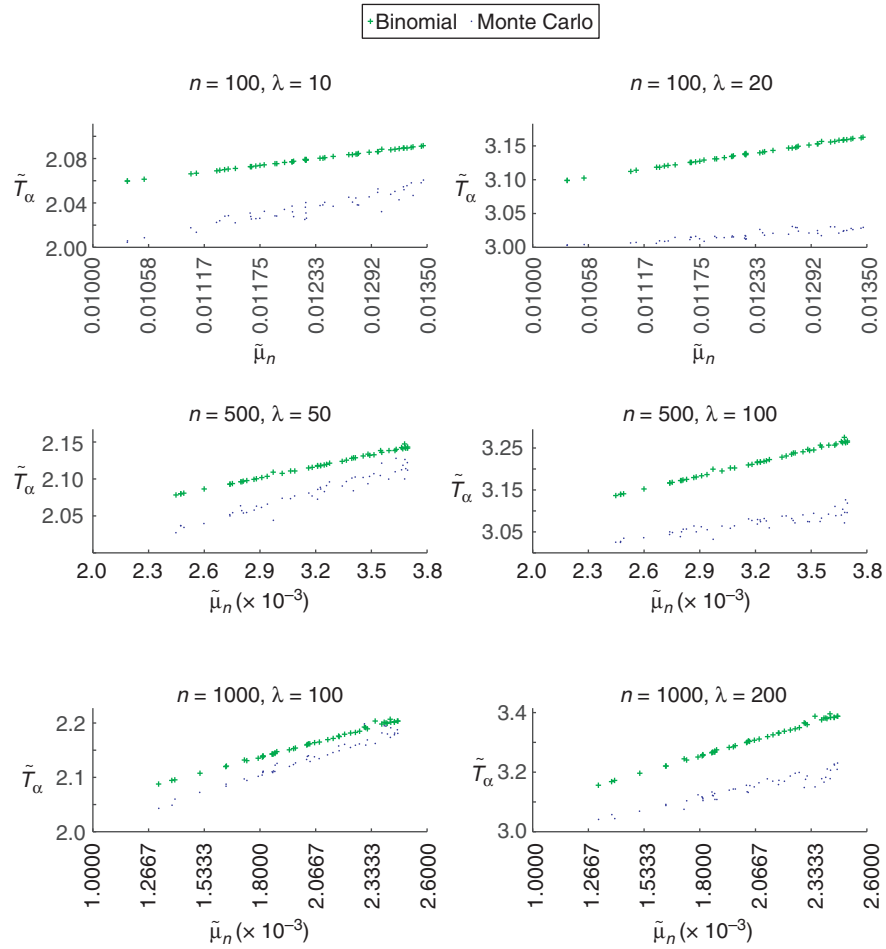
Figure 8 compares the approximation with Monte Carlo when $\tilde{\mu}_n$ is in the neighborhood of $1/n$ (ie, type I). The numerical investigation shows that the binomial approximation of Proposition 7.2 is decent at a high α .

8 EXTENSION TO CONTINUOUS DISTRIBUTIONS

The extension of the EBA to the continuous case is motivated by the fact that continuous distributions may be approximated by discretizations, to which the EBA is then applicable. However, better approximations correspond to a larger sample size, n , and for large enough n the constraint on α given by (3.6) is violated; moreover, the largest loss is unbounded as n increases. To remedy this, we replace the largest loss with that expected to occur at least once in $t_\alpha = 1/(1-\alpha)$, or on average once in $\lambda t_\alpha = \lambda/(1-\alpha)$ losses. We define

$$\begin{aligned} M_{\alpha^*} &= \sup \left\{ x \mid P(X > x) > \frac{1}{\lambda/(1-\alpha)} \right\} \\ &= F^{-1} \left(1 - \frac{1-\alpha}{\lambda} \right) \\ &= F^{-1}(\alpha^*), \end{aligned}$$

FIGURE 8 Binomial approximation compared with Monte Carlo simulation at the 99.9th percentile when the sample \tilde{L} contains a very large loss.



where

$$\alpha^\star = 1 - \frac{1 - \alpha}{\lambda}.$$

In our context, M_{α^\star} is the analog of the largest loss M in the discrete case. We propose the following formula as an approximation of the quantile function in the continuous case:

$$\text{EBA}^c = \mu_{\alpha^\star}(f_\alpha^c - \beta_\alpha^c) + \beta_\alpha^c M_{\alpha^\star}, \quad (8.1)$$

TABLE 3 Minimum, average and maximum percentage deviation of the extension formula, EBA^c , from the Monte Carlo estimate for $\lambda = 100$.

	α				
	95%	97.5%	99%	99.5%	99.9%
Minimum (%)	2.0	1.2	-1.4	-3.4	-1.4
Average (%)	3.4	2.1	0.7	0.0	2.0
Maximum (%)	4.8	3.6	4.8	6.1	9.6

TABLE 4 Minimum, average and maximum percentage deviation of the extension formula, EBA^c , from the Monte Carlo estimate for $\lambda = 1000$.

	α				
	95%	97.5%	99%	99.5%	99.9%
Minimum (%)	-1.5	-1.4	-1.2	-1.2	-5.0
Average (%)	0.3	-0.2	0.1	0.8	-2.4
Maximum (%)	2.7	1.3	1.1	2.0	1.0

where μ_{α^*} is the conditional mean below M_{α^*} ,

$$\mu_{\alpha^*} = \frac{1}{\alpha^*} \int_0^{M_{\alpha^*}} x dF(x).$$

f_{α}^c (respectively, β_{α}^c) is the equivalent of f_{α} (respectively, β_{α}) in (3.5) and is computed using the extension over the positive line of the Poisson distribution:

$$f_{\alpha}^c = G_P^{-1}(\alpha, \lambda) + 1, \quad (8.2)$$

$$\beta_{\alpha}^c = G_P^{-1}(\alpha, 1 - \alpha) + 1, \quad (8.3)$$

where $G_P(\cdot, \cdot)$ is given in Proposition 5.1. As previously mentioned, β_{α}^c represents the number of times the large loss, in our case M_{α^*} , appears in t_{α} years. Thus, its intensity in (8.3) is $1 - \alpha$.

We investigate the efficiency of this extension to the lognormal distribution with an associated normal distribution of mean 0 and a standard deviation varying from 1 to 6 with a 0.05 step increase. Tables 3 and 4 report the minimum, average and maximum percentage deviations from Monte Carlo simulation estimates for frequencies 100 and 1000. The same statistics are reported for the single loss approximation with mean correction (Böcker and Sprittulla 2006) in Tables 5 and 6:

$$SLA(\alpha) = F^{-1}(\alpha^*) + \mu(\lambda - 1), \quad (8.4)$$

TABLE 5 Minimum, average and maximum percentage deviation of SLA from the Monte Carlo estimate for $\lambda = 100$.

	α				
	95%	97.5%	99%	99.5%	99.9%
Minimum (%)	-10.3	-11.7	-12.7	-13.2	-12.6
Average (%)	155.7	52.8	11.8	2.2	-0.0
Maximum (%)	1707.5	541.8	126.1	40.9	6.8

TABLE 6 Minimum, average and maximum percentage deviation of SLA from the Monte Carlo estimate for $\lambda = 1000$.

	α				
	95%	97.5%	99%	99.5%	99.9%
Minimum (%)	-5.5	-6.6	-7.4	-7.8	-8.8
Average (%)	52.1	20.2	5.5	2.2	-3.6
Maximum (%)	449	168.5	49.0	20.9	-0.3

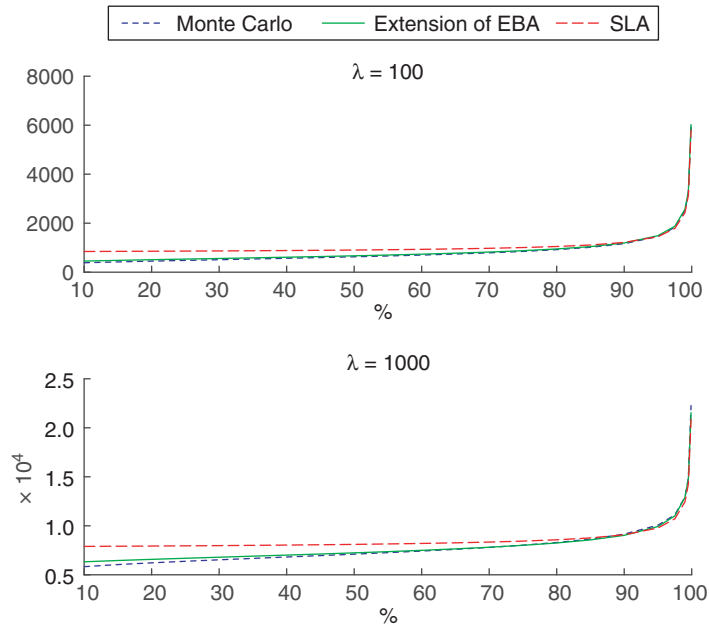
where μ is the unconditional mean. The extension is more accurate than the SLA for the 99th percentile and below. Its efficiency is comparable with the SLA for higher percentiles. Figure 9 compares the EBA extension to Monte Carlo and SLA for a lognormal distribution with mean 0 and standard variation 2 at various percentiles. The EBA^c is efficient at low percentiles, and comparable with Monte Carlo estimates. As expected, the SLA formula behaves poorly at low percentiles. We recall that SLA is asymptotically efficient for subexponential distributions.

9 CONCLUSIONS

We investigated the mechanics of the empirical aggregate loss bootstrap distribution and show that it can be represented as a linear combination of independent Poisson distributions. We suggest different approximations of its quantile function for a high percentile.

- (a) Empirical bootstrap analytic (EBA): this is an intuitive analytical approximation derived based on extreme cases. By construction, it is efficient when the sample has an extremely large loss (type I) or all losses have the same order of magnitude (type II). It remains close to the Monte Carlo estimates for samples of type III.

FIGURE 9 EBA^c formula (8.1) compared with SLA and Monte Carlo simulation for a lognormal distribution with mean 0 and standard deviation 2 at various percentiles.



- (b) Poisson approximation: this matches the first and second central moments of the total aggregate loss distribution and smooths out the Poisson cumulative density function. It behaves poorly in type I, and provides relatively efficient estimates for other samples.
- (c) Cornish–Fisher expansion: this approximates the aggregate loss bootstrap distribution, which is a discrete distribution, by a continuous one using its cumulants. It approximates well the quantile function using three cumulants in all three categories.
- (d) Panjer recursion: this is a numerical procedure whose efficiency depends on its parameterization. An appropriate choice of equispaced points and their probability mass weights results in an efficient approximation.
- (e) Binomial approximation: this approximates the aggregate loss bootstrap distribution in the presence of an extremely large loss in the sample (type I). The distribution is approximated by an extension over the real line of a binomial distribution. It provides an efficient estimate of the quantile function for a high percentile.

We also presented a possible extension of the EBA to the continuous case. At high percentiles, numerical investigation shows that its efficiency is comparable with the single loss approximation, when the severity distribution can be efficiently approximated by the empirical distribution of a finite sample. It is also efficient at approximating low percentiles (see Figure 9).

The empirical bootstrap distribution has been of great interest to operational risk practitioners for benchmarking capital results and stress testing exercises. A deeper understanding of how outlier losses and simple statistics of the loss sample affect the bootstrap distribution, particularly at high percentiles, enhances its utility as a benchmark.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. The views expressed here, if any, are those of the authors only and not necessarily of the institution with which they are affiliated. Any remaining inadequacies are the authors' alone.

ACKNOWLEDGEMENTS

The authors thank Eric Cope for his useful discussions, and the participants of the ORX 2015 Analytics Working Group forum for their feedback.

REFERENCES

- Abramowitz, M., and Stegun, I. S. (1972). *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. National Bureau of Standards, Applied Mathematics Series. Government Printing Office, Washington, DC.
- Böcker, K., and Sprittulla, S. (2006). Operational VAR: meaningful means. *Risk* **19**(12), 96–98.
- Cope, E. W., Mignola, G., Antonini, G., and Ugoccioni, R. (2009). Challenges and pitfalls in measuring operational risk from loss data. *The Journal of Operational Risk* **4**(4), 3–27 (<http://doi.org/bpqx>).
- DasGupta, A. (1998). On a differential equation and one step recursion for Poisson moments. Technical Report 98-02, Purdue University (<http://bit.ly/2qSlkcC>).
- Gerber, H. U. (1982). On the numerical evaluation of the distribution of aggregate claims and its stop-loss premiums. *Insurance: Mathematics and Economics* **1**(1), 13–18 (<http://doi.org/cq4kwn>).
- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*, 3rd edn. Wiley Interscience (<http://doi.org/ct345v>).
- Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**(1), 22–26 (<http://doi.org/b6t4>).
- Steel, R. G. D. (1953). Relation between Poisson and multinomial distributions. Biometrics Unit Technical Report BU-39-M, Cornell University (<http://bit.ly/2qJcYaY>).