ABSTRACT

The objective of this paper is to use a simulation comparison to evaluate quantile approximation methods in terms of their practical usefulness and potential applicability in an operational risk context. A popular method in modelling the aggregate loss distribution in risk and insurance is the Loss Distribution Approach (LDA). Many banks currently use the LDA for estimating regulatory capital for operational risk. The aggregate loss distribution is a compound distribution resulting from a random sum of losses, where the losses are distributed according to some severity distribution and the number (of losses) distributed according to some frequency distribution. In order to estimate the regulatory capital, an extreme quantile of the aggregate loss distribution has to be estimated. A number of numerical approximation techniques have been proposed to approximate the extreme quantiles of the aggregate loss distribution. We use PROC SEVERITY to fit various severity distributions to simulated samples of individual losses from a preselected severity distribution. The accuracy of the approximations obtained, is then evaluated against a Monte Carlo approximation of the extreme quantiles of the compound distribution resulting from the preselected severity distribution.

We find that the second-order perturbative approximation, a closed-form approximation, performs very well at the extreme quantiles and over a wide range of distributions, and it is very easy to implement.

INTRODUCTION

Compound frequency distributions have many applications, particularly in insurance and risk management. For example, in insurance, they are used to model the distribution of aggregate claims on an insurance policy over a fixed period or, in operational risk, to model the annual aggregate loss distribution. An aggregate loss (or claims) distribution is a compound distribution resulting from a random sum of losses (or claims), where the individual losses (or claims) are independently and identically distributed according to some distribution (the severity distribution) and the number of losses (or claims) is independently distributed according to some frequency distribution.

In operational risk the aggregate annual loss distribution is constructed in order to determine regulatory capital, which is equivalent to determining the 0.999 quantile of this distribution. In practice this quantity is estimated by using the so-called loss distribution approach (LDA). This method is usually implemented by fitting a distribution to the individual losses and then the aggregate loss distribution is constructed through brute force Monte Carlo simulation of random sums of losses, where the number of losses are generated from a (fitted) frequency distribution (e.g. Poisson) and the individual losses generated from the (fitted) severity distribution. This approach has been studied extensively and it has been shown that the resulting capital estimates are not trustworthy and depend on many uncontrollable factors (see e.g. Cope et al. 2009). In practice the LDA boils down to modelling the loss frequency by the Poisson or negative binomial distributions and the severity distribution by a wide class of sub-exponential distributions (e.g. the Burr, g-and-h, lognig, lognormal and combinations of these with the generalised Pareto).

The quantiles of a compound distribution can typically not be calculated exactly and have to be approximated in some way. In practice, brute force Monte Carlo (MC) simulation methods are mostly used to approximate these extreme quantiles. Depending on the accuracy required, these MC simulations are computer intensive and even utilising today’s computer power can become impractical when implemented. A number of numerical approximation techniques have been proposed to overcome this difficulty. Numerical approximation recursive techniques that can be used to approximate the quantile of the compound distribution include the Panjer recursion (Panjer, 1981) and techniques using Fourier inversion and the fast Fourier transform (see e.g. Grübel and Hermesmeier, 1999). These techniques require several input parameters that have to be selected carefully to ensure convergence and are computationally intensive.
For the class of sub-exponential (severity) distributions, Böcker and Klüppelberg (2005) derived a first order single-loss approximation (SLA) which was extended by Degen (2010) to a second order single-loss approximation. Interestingly, the SLA methods for approximating the extreme quantile of the compound distribution are based on an even more extreme quantile of the severity distribution. Hernandez et al. (2013, 2014) derived closed-form approximations for high percentiles of the aggregate distribution based on a perturbative expansion, which they claim improve on the above-mentioned methods under certain conditions. The perturbative series introduced differs from previous approximations in that the terms in the series are expressed as a function of the moments of the right truncated distribution for the individual random variables in the sum. These censored moments exist even when the moments of the original distribution (without truncation) diverge. Consequently the same expression is valid for both the infinite and finite mean cases.

In this paper we present an extension to the study presented by de Jongh et al. (2016). de Jongh et al. (2016) found that the second order perturbative approximation performs very well at the extreme quantiles and over a wide range of distributions. We extend the simulation study by simulating samples of individual losses from a Burr distribution. We then use PROC SEVERITY to fit the Burr distribution to these samples. The accuracy of the approximation methods applied to the fitted sample distributions are then evaluated against a Monte Carlo approximation of the extreme quantiles of the compound distribution resulting from the preselected severity distribution.

APPROXIMATION TECHNIQUES

In this section a brief overview is given of the standard Monte Carlo approximation, the single-loss approximation and the perturbative approximation techniques. Before we discuss these methods let us state the problem in theoretical terms. Assume that random variables \(X_1, ..., X_N\) are independent and identically distributed according to some distribution \(F\) and that \(N\) follows the Poisson distribution with parameter \(\lambda\). We know that if \(N \sim \text{Pois}(\lambda)\) and \(X_1, ..., X_N \sim F\), then \(S = \sum_{i=1}^{N} X_n \sim \text{CoP}(F, \lambda)\) is the compound Poisson distribution with parameters \(\lambda\) and \(F\). We are interested in obtaining the \(100(1 - \gamma)\%\) quantile of this distribution and since no closed-form solution exists, approximation methods are used. The \(100(1 - \gamma)\%\) quantile of the latter distribution is the so-called VaR and for calculating regulatory capital (RC) one is particularly interested in the \(99.9\%\) VaR, i.e. where \(\gamma = 0.001\).

MONTE CARLO APPROXIMATION

As before, let the random variable \(N\) denote the annual number of loss events and the random variables \(X_1, ..., X_N\) denote the loss severities of these loss events. Then the annual aggregate loss is \(S = \sum_{n=1}^{N} X_n\) and the distribution of \(S\) is the aggregate loss distribution. This aggregate loss distribution and its \(100(1 - \gamma)\%\) are difficult to calculate analytically and Monte Carlo (MC) simulation is often used to approximate it as follows:

I. Generate \(N\) distributed according to the assumed frequency distribution;

II. Generate \(X_1, ..., X_N\) independent and identically distributed according to the severity distribution \(G\) and calculate \(S = \sum_{n=1}^{N} X_n\);

III. Repeat (i) and (ii) \(I\) times independently to obtain \(S_i, i = 1, 2, ..., I\) and approximate the \(99.9\%\) quantile as \(\hat{S}_{(0.999\%)} = \max_{i} S_i\) where \(S_i\) denotes the \(i\)-th order statistic and \([k]\) the largest integer contained in \(k\).

Note that three input items are required to perform it, namely the number of MC repetitions \(I\) as well as the frequency and loss severity distributions. The number of MC repetitions determines the accuracy of the approximation and the larger it is, the higher its accuracy. In principle infinitely many repetitions are required to get the exact true quantile. We always use one million repetitions in our simulations; while this may sound high, it should be kept in mind that a very high quantile (99.9%) is calculated which forces a large number of repetitions in order to gain accuracy. The Monte Carlo approximation method will subsequently be
referred to as MC. Below an example using a data step in SAS® to simulate one million observations from a $CoP(10, F)$ where $F \sim \log N(0,1)$:

```
data sim;
drop i;
do i = 1 to 1000000;
n = ranpoi(0,10);
output;
end;
run;
data sim;
set sim;
drop i;
aggloss = 0;
do i = 1 to n;
   aggloss = aggloss + quantile('lognorm',ranuni(0),0,1);
end;
run;
```

**SINGLE-LOSS APPROXIMATION**

Böcker and Klüppelberg (2005) derived a simple single-loss approximation that provides closed-form estimates for the class of sub-exponential distributions. If $F$ is the true underlying severity distribution function of the individual losses and $\lambda$ the true annual frequency then the $100(1 - \gamma)\%$ quantile of the compound loss distribution may be approximated by

$$F^{-1}(1 - \gamma / \lambda).$$

(1)

This will be subsequently referred to as SLA.

For heavy-tailed distributions, Degen (2010), using second order exponentiality, derived an improved single-loss approximation. Assuming that $N \sim Pois(\lambda)$, $F$ has a finite mean ($E(X) < \infty$), he showed that the $100(1 - \gamma)\%$ quantile of the compound loss distribution may be approximated by

$$F^{-1} \left(1 - \frac{\gamma}{\lambda}\right) + \lambda \mu,$$

(2)

where $\mu = E(X)$ is the finite mean of $F$.

For infinite mean models ($\mu = E(X) = \infty$) the $100(1 - \gamma)\%$ quantile of the compound loss distribution may be approximated by

$$F^{-1} \left(1 - \frac{\gamma}{\lambda}\right) + \gamma F^{-1} \left(1 - \frac{\gamma}{\lambda}\right) \frac{C_\kappa}{1 - \kappa},$$

(3)

where $C_\kappa = (1 - \kappa) r^2 \frac{\Gamma(1 - \frac{2}{\kappa})}{2 \Gamma(1 - \frac{1}{\kappa})}$, $\Gamma$ the gamma function and $\kappa$ the extreme value index (EVI of tail index, see Embrechts et al. 1997 for more on EVI) of the distribution $F$. See Degen (2010) for more information on special cases. The approximation suggested by Degen will be referred to as SLAD.

Hannah and Puza (2015) extends the SLA adding another correction term to account for the effects of the two largest losses. Assuming that $N \sim Pois(\lambda)$, $G$ has a finite mean ($E(X) < \infty$), they showed that the $100(1 - \gamma)\%$ quantile may be approximated by solving for $c$ in

$$\gamma = \lambda \left(1 - F(c - \lambda \mu)\right) + \frac{1}{2} \lambda^2 \left(1 - F \left(\frac{c - \lambda \mu}{\lambda}\right)\right)^2.$$

(5)

In the case of infinite mean models ($\mu = E(X) = \infty$) we use Degen’s approximation for $\lambda \mu = \gamma G^{-1} \left(1 - \frac{\gamma}{\lambda}\right) \frac{C_\kappa}{1 - \kappa}$ (for $1 < \kappa < \infty$) and $\lambda \mu = \lambda \mu G^{-1} \left(1 - \frac{\gamma}{\lambda}\right)$ (for $\kappa = 1$) in equation (5). The approximation suggested by Hannah and Puza (2015) subsequently will be referred to as SLAH.
PERTURBATIVE APPROXIMATION

Hernandez et al. (2014) introduced k-th order perturbative approximations for calculating the 100(1 − γ)% quantile of the compound loss distribution. Assume \( X \sim F \) and the frequency distribution is \( \text{Pois}(\lambda) \) then the 0th, 1st and 2nd order approximations (subsequently denoted by PA0, PA1 and PA2 respectively) are given by \( Q_0, Q_0 + Q_1 \) and \( Q_0 + Q_1 + \frac{Q_2}{2} \) respectively, where

\[
Q_0 = F^{-1}\left(\frac{\lambda + \ln(1-\gamma)}{\lambda}\right)
\]

\[
Q_1 = (\lambda + \ln(1-\gamma))E(X|X < Q_0);
\]

\[
Q_2 = -\left(\lambda f(Q_0) + \frac{f'(Q_0)}{f(Q_0)}\right)(\lambda + \ln(1-\gamma))E(X^2|X < Q_0) - \lambda f(Q_0) Q_0^2
\]

with \( f \) the density of \( F \).

METHODOLOGY FOR EVALUATION OF APPROXIMATION METHODS

In order to test the accuracy of the above-mentioned closed-form approximation techniques we designed a Monte Carlo (MC) study. Whereas de Jongh et al. (2016) measures the accuracy of the approximation methods on the preselected distributions to the MC approximation we measure the accuracy of the approximations methods when estimating the quantiles on samples of these distributions. We restrict our study only to the Burr distribution. As stated before, we assumed a Poisson frequency distribution throughout and selected the severity distribution as the Burr distribution. The density of the Burr distribution is given below as well as the parameter sets that were used to generate the compound distributions.

THE BURR DISTRIBUTION

The three parameter Burr type XII distribution function is given by

\[
\text{Burr}(x; \eta, \tau, \alpha) = 1 - \left(1 + \left(\frac{x}{\eta}\right)^\tau\right)^{-\alpha}, \text{ for } x > 0
\] (6)

with parameters \( \eta, \tau, \alpha > 0 \) (see e.g. Beirlant et al., 2004). Here \( \eta \) is a scale parameter and \( \tau \) and \( \alpha \) shape parameters. Note the extreme value index of the Burr distribution is given by \( \text{EVI} = \kappa = 1/\tau \alpha \) and that heavy-tailed distributions have a positive EVI and larger EVI implies heavier tails. This follows (also) from the fact that for positive EVI the Burr distribution belongs to the Pareto-type class of distributions. For Pareto-type, when the EVI ≥ 1 the expected value does not exist, and when EVI > 0.5, the variance is infinite. The density of the Burr type XII is given by

\[
burr(x; \eta, \tau, \alpha) = \frac{\tau x^{\tau-1}}{\eta \tau} \left[1 + \left(\frac{x}{\eta}\right)^\tau\right]^{-(\alpha+1)}, \text{ for } x > 0
\] (7)

with parameters \( \eta, \tau, \alpha > 0 \).

PARAMETER SETS FOR THE SIMULATION STUDY

The parameter values considered in the simulation study are given in Tables 1 below for the Burr distribution. The annual Poisson frequencies are taken as \( \lambda = 10, 20, 50, 100, 200, 500 \), the probability levels \( \gamma = 0.001, 0.005, 0.01, 0.025, 0.05 \) and we base the study on seven years of simulated historical loss data throughout.
Table 1: Burr parameter sets selected for the Monte Carlo study

<table>
<thead>
<tr>
<th>η</th>
<th>τ</th>
<th>α</th>
<th>κ</th>
<th>𝐸(𝑋)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>5</td>
<td>0.333333</td>
<td>&lt; ∞</td>
</tr>
<tr>
<td>1</td>
<td>0.6</td>
<td>2</td>
<td>0.833333</td>
<td>&lt; ∞</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>∞</td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1.5</td>
<td>1.333333</td>
<td>∞</td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>0.3</td>
<td>1.851852</td>
<td>∞</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td>0.17</td>
<td>2.352941</td>
<td>∞</td>
</tr>
</tbody>
</table>

The overall simulation study entailed the following work program. For each combination of parameters of the assumed true underlying Poisson frequency and Burr severity distributions:

a) Use the MC algorithm in Section 2 to determine the 100(1 − 𝜉)% quantile as described in i–iii. Note that the value obtained here approximately equals the true 100(1 − 𝜉)% quantile. The only approximation involved is that it is based on 1 million repetitions, rather than an infinite amount. We refer to this value as the approximately true (AT) 100(1 − 𝜉)% quantile.

b) Generate a data set of historical losses, i.e., generate \( K \sim \text{Pois}(7, \lambda) \), and then generate \( x_1, x_2, \ldots, x_K \sim \text{iid Burr type XII} \) with the current choice of parameters. Refit the Burr type XII distribution to the generated historical data to estimate the parameters.

c) Calculate the closed-form approximations (i.e., SLA, SLAD, SLAH, PA0, PA1 and PA2) of the 100(1 − 𝜉)% quantile for the estimated Burr type XII distribution.

d) Repeat (a)–(c) \( J \) times, and then summarize and compare the resulting 100(1 − 𝜉)% quantile estimates.

We have carried out this program with the choice \( J = 1000 \). For each combination of parameter values the output of the program detailed above results in 1000 AT 100(1 − 𝜉)% quantile values as well as 1000 SLA, SLAD, SLAH, PA0, PA1 and PA2 estimates. As mentioned in (a), those 1000 AT values each approximate the true 100(1 − 𝜉)% quantile, and a summary measure such as their mean or median is very close to the true 100(1 − 𝜉)% quantile. Because we are generally dealing with positively skewed data here, we shall use the median as the principal summary measure. Denote the median of the 1000 AT values by \( \text{MedAT} \). Then, the 1000 repeated SLA, SLAD, SLAH, PA0, PA1 and PA2 values may be taken as estimates of \( \text{MedAT} \); we wish to decide which of the methods is best in the sense of coming closest to \( \text{MedAT} \). To express the quality of the methods, we shall use their median absolute relative deviation (MARD) from \( \text{MedAT} \). For example, if \( \text{SLAJ} \) denotes the SLA value of the \( j \)th repetition, then \( \text{MARD(SLA)} \) is

\[
\text{MARD(SLA)} = \left\{ \left| \frac{\text{SLAJ}}{\text{MedAT}} - 1 \right|, j = 1, \ldots, 1000 \right\},
\]

and similarly for SLAD, SLAH, PA0, PA1 and PA2.

Furthermore we report the relative error as

\[
\text{RE} = |\text{Median}_{\text{SLA}} - \text{MedAT}|/\text{MedAT},
\]

the absolute relative deviation of the median of the estimates for a particular approximation technique from the MedAT expressed as a percentage of the MedAT. Similarly for SLAD, SLAH, PA0, PA1 and PA2.

For ease of graphical presentation we replace the probability levels \( \gamma = 0.001, 0.005, 0.01, 0.025, 0.05 \) by their logodds (\( \ln\left(\frac{1}{\gamma}\right) = 6.9, 5.3, 4.6, 3.7, 2.9 \)) and subsequently, where applicable, the horizontal axis of graphs are constructed using the logodds rather than the probability scale.

For ease of graphical presentation we replace the probability levels \( \gamma = 0.001, 0.005, 0.01, 0.025, 0.05 \) by their logodds (\( \ln\left(\frac{1}{\gamma}\right) = 6.9, 5.3, 4.6, 3.7, 2.9 \)) and subsequently, where applicable, the horizontal axis of graphs are constructed using the logodds rather than the probability scale.
USING PROC SEVERITY

The SEVERITY procedure estimates parameters of any arbitrary continuous probability distribution that is used to model the magnitude (severity) of a continuous-valued event of interest. For this reason we use PROC SEVERITY provides a default set of probability distribution models that includes the Burr, exponential, gamma, generalized Pareto, lognormal, and other loss distributions typically used in Operational Risk (see SAS, 2013). By using PROC SEVERITY you can estimate the parameters of these distributions by using a list of severity values that are recorded in a SAS data set. PROC SEVERITY computes the estimates of the model parameters, their standard errors, and their covariance structure by using the maximum likelihood method (see SAS, 2013).

With PROC SEVERITY it is possible to fit multiple distributions at the same time and choose the best distribution according to a selection criterion that you specify. Seven different statistics of fit as selection criteria can be used, including log likelihood, Akaike’s information criterion (AIC), Schwarz Bayesian information criterion (BIC), Kolmogorov-Smirnov statistic (KS), Anderson-Darling statistic (AD), and Cramér-von Mises statistic (CvM) (see SAS, 2013).

Below an example of fitting a Burr distribution in SAS® using PROC SEVERITY.

```
proc severity
  data    = losses
  OUTTEST = est
  OUTSTAT = stat
  CRITERION = AD;
  loss    loss;
  dist    burr;
run;
```

In our study using the predefined Burr parameterization (in term of logs) in PROC SEVERITY result in suboptimal estimates for the parameters. PROC SEVERITY enables you to define any arbitrary continuous parametric distribution model and to estimate its parameters. To do this the key components of the distribution need to be defined, such as its probability density function (PDF) and cumulative distribution function (CDF), as a set of functions and subroutines written with the FCMP procedure.

Below the PROC FCMP code used to define the alternative Burr distribution.

```
proc fcmp library=sashelp.svrtdist outlib = func.svrtdist2.seldef;

/**** Alternative Burr ****/

    function burr_alt_DESCRIPTION() $256;
        length desc $256;
        desc1 = "Alternative parameterisation for the Burr distribution.";
        desc2 = " eta, alpha and tau are free parameters.";
        desc = desc1 || desc2;
        return(desc);
    endsub;

    function burr_alt_PDF(x,eta,tau,alpha);
        f = (alpha*tau*eta**(-1))*((x/eta)**(tau-1))/(1 + (x/eta)**(tau))**(alpha+1);
        return(f);
    endsub;

    function burr_alt_CDF(x,eta,tau,alpha);
        F = 1 - (1 + (x/eta)**(tau))**(-alpha);
    endsub;

```
return(F);
endsub;

subroutine burr_alt_LOWERBOUNDS(eta,tau,alpha);
  outargs eta, tau, alpha;
  eta = 0; /* eta > 0 */
  alpha = 0; /* alpha > 0 */
  tau = 0; /* tau > 0 */
endsub;
quit;

The user defined distribution can then be used in PROC SEVERITY by specifying the cmplib option statement as demonstrated below. Furthermore we make use of the nloptions statement to control different aspects of this optimization process.

```plaintext
option cmplib = func.svrtdist2;
proc severity
data = losses
OUTEST = est
outstat = stat
CRIT = A ;
loss loss ;
dist burr_alt;
nloptions tech=tr maxiter=10000 maxfu = 10000 ;
run;
```

RESULTS

In order to study the performance of the approximation techniques, we present in Figures 1 to 3 below, the relative error plots for three cases ($EVI = 0.333, 1$ and $2.35$). In each case the relative error is plotted for each approximation technique against the log-odds of the probability levels and this is done for intensities 20, 100 and 500. The 5% and 95% MC confidence bands are also depicted as lines in each plot. In Figures 1 and 2 PA0 and SLA perform poorly, while the other measures struggle at the lower probability levels, especially at 0.05. For Burr $EVI = 0.33$ the performance of the approximation methods are particularly poor and it is only PA1 and PA2 that performs reasonably, but only at the lower probability levels and higher frequency. The performance of the approximations (SLAD, SLAH, PA1 and PA2) improves significantly for heavier tails ($EVI = 1$ and $2.35$) with PA2 clearly the preferred one. The results show that PA1 and PA2 are almost ‘spot-on’ for heavier tails while SLAD and SLAH tend to overestimate slightly and SLA and PA0 underestimate slightly at especially the higher probability levels. Note how the MC CI increases as the probability level decreases.

We are especially interested in the performance of the approximation methods at a probability level of 0.001 when estimating regulatory capital. From Figure 1 ($EVI = 0.33$) it is clear that PA1 and PA2 performs better, especially the higher frequencies. And for heavier tails in Figures 2 and 3 ($EVI = 1$ and $2.35$) it is clear that all the approximation methods, with the possible exception of SLA, SLAD and SLAH in Figure 2 perform well.
Figure 1: Relative error plots for the semi-heavy tailed Burr distribution (EVI=0.33)

Figure 2: Relative error plots for the heavy tailed Burr distribution (EVI=1)

Figure 3: Relative error plots for the very-heavy tailed Burr distribution (EVI=2.35)
In Figures 4 to 6 below the MARD plots for three cases ($EVI = 0.333, 1 \text{ and } 2.35$) are presented. In each case the MARD is plotted for each approximation technique against the log-odds of the probability levels and this is done for intensities 20, 100 and 500. In Figures 1 and 2 PA0 and SLA perform poorly. We also observe that the MARD increase with lower probability levels (as seen from Figure 1 to 3 above that the MC CI increases) and that the MARD decreases with higher frequencies. Form Figure 1 and 2 ($EVI = 0.33 \text{ and } 1$) PA1 and PA2 clearly performs better. And for heavier tails in Figure 3 ($EVI = 2.35$) it is clear that all the approximation methods perform well.

Figure 4: MARD plots for the semi-heavy tailed Burr distribution ($EVI=0.33$)

Figure 5: MARD plots for the heavy tailed Burr distribution ($EVI=1$)
From the discussion above we conclude that for finite and infinite mean distributions the PA2 approximation technique is the preferred choice at all quantiles and intensities considered.

CONCLUSION

We have analysed the performance of the approximation methods for the quantiles of the compound distribution. We found that the PA2 perturbative closed-form approximation method of Hernandez et al. (2014) performed very well in most cases considered (see also de Jongh et al., 2016). If the goal is to approximate an extreme quantile of the compound distribution (such as at a probability level of 0.001), then this method is recommended. We have also illustrated how to use PROC SEVERITY and incorporating a user defined distribution.

REFERENCES


**CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Helgard Raubenheimer  
Centre for BMI, North-West University  
Helgard.Raubenheimer@nwu.ac.za  
[www.nwu.ac.za/bmi](http://www.nwu.ac.za/bmi)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.