Combining scenario and historical data in the loss distribution approach: A new procedure that incorporates measures of agreement between scenarios and historical data.

PJ de Jongh[1], T de Wet[2], H Raubenheimer[3] and JH Venter[4]

Abstract

Many banks use the loss distribution approach in their advanced measurement models to estimate regulatory or economic capital. This boils down to estimating the 99.9% VaR of the aggregate loss distribution and is notoriously difficult to do accurately. Also, it is well-known that the accuracy with which the tail of the loss severity distribution is estimated is the most important driver in determining a reasonable estimate of regulatory capital. To this end, banks use internal data and external data (jointly referred to as historical data) as well as scenario assessments in their endeavour to improve the accuracy with which the severity distribution is estimated. In this paper we propose a simple new method whereby the severity distribution may be estimated using historical data and experts' scenario assessments jointly. The way in which historical data and scenario assessments are integrated incorporates measures of agreement between these data sources, which can be used to evaluate the quality of both. In particular we show that the procedure has definite advantages over traditional methods where the severity distribution is modelled and fitted separately for the body and tail parts, with the body part based only on historical data and the tail part on scenario assessments.

1 Introduction

Many banks worldwide currently use various versions of the so-called loss distribution approach (LDA) to calculate Value-at-Risk (VaR) for Operational Risk under the Basel Accord's Advanced Measurement Approach (AMA). For a description of the practical application of the LDA in a bank the interested reader is referred to Aue and Kalkbrener (2007).

[1] Director of the Centre for Business Mathematics and Informatics (BMI), North-West University (NWU).
[2] Extra-ordinary professor at the Centre for BMI, NWU
[3] Associate professor at the Centre for BMI, NWU.
[4] Professor at the Centre for BMI, NWU.

One of the important building blocks of the LDA approach is the annual aggregate loss distribution in a particular unit of measure or operational risk category (ORC). According to the Basel II Accord, the 99.9% VaR of the aggregate loss distribution is used as a measure of the regulatory capital (EC) in that particular ORC. The aggregate loss distribution is obtained by convoluting a loss event frequency distribution and a loss severity distribution by means of the random sums method. Typically the Poisson distribution is used as frequency distribution together with various alternatives for the severity distribution. Popular choices for the latter include the Lognormal, Burr, g-and-h, or mixtures of these distributions with Generalised Pareto distributions.

More formally, if the random variable N denotes the annual number of loss events and if the random variables $X_1, \dots, X_N$ denote the loss severities of these loss events, then the annual aggregate loss is $A = \sum_{n=1}^{N} X_n$ and the distribution of A is the aggregate loss distribution of the ORC. This aggregate loss distribution and its VaR are difficult to calculate analytically in most cases so that in practice Monte Carlo (MC) simulation is often used. For reference purposes later in this paper, here is a summary of the procedure for doing this:

Algorithm 1 : B**asic Monte Carlo (MC) approach**

(1) Generate N distributed according to the assumed frequency distribution;

(2) Generate $X_1, \dots, X_N$ independent and identically distributed according to the true loss severity distribution G (abbreviated $X_1, \dots, X_N \sim iid\ G$) and calculate $A = \sum_{n=1}^{N} X_n$;

(3) Repeat (1) and (2) I times independently to obtain $A_i, i = 1,2, \dots, I$ and approximate the 99.9% VaR as $A_{([0.999*I]+1)}$ where $A_{(i)}$ 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 and the frequency and loss severity distributions. The number of 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 VaR. We always use one million repetitions here; while this may sound unnecessarily high, keep in mind that a very high quantile (99.9%) is calculated which forces a large number of repetitions in order to gain accuracy.

The frequency and severity distributions assumed in Steps (1) and (2) of Algorithm 1 are typically not known in practice. Their choices need to be based on actual loss data and require suitable statistical estimation methods. More formally, assume that we have collected the loss severities of a total of K loss events spread over a years and denote these observed or historical losses by $x_1, \dots, x_K$. Then the annual frequency is estimated by $\hat{\lambda} = K/a$. Also let $F(x; \theta)$ denote a suitable family of distributions with which to model the true loss severity distribution G. One of the families mentioned above may be chosen and it is fitted (e.g. using maximum likelihood) to the historical losses obtaining the fitted distribution $F(x; \hat{\theta})$, with $\hat{\theta}$ denoting the estimate of the parameter(s) θ. Then the following adjustment of Algorithm 1 is typically used to **estimate** the required VaR:

Algorithm 2 : Standard Monte Carlo approach

(1) Assume that N is Poisson distributed and estimate the annual frequency by $\hat{\lambda} = K/a$; then generate N from the Poisson distribution with parameter $\hat{\lambda}$;

(2) Generate $X_1, \dots, X_N \sim iid\ F(.; \hat{\theta})$ and calculate $A = \sum_{n=1}^{N} X_n$;

(3) Repeat (1) and (2) I times independently to obtain $A_i, i = 1,2, \dots, I$ and estimate the 99.9% VaR by the corresponding empirical quantile of these A_i's as in Step (3) of Algorithm 1.

Algorithm 2 differs from Algorithm 1 in that actual loss data is involved as well as assumptions and parameter estimation regarding the underlying frequency and severity distributions.

The large number of simulation repetitions involved in the MC approaches above motivates the use of other numerical methods such as Panjer recursion, methods based on fast Fourier transforms (see e.g. Panjer, 2006) and the single loss approximation (SLA) method (see e.g. Böcker and Klüppelberg 2005). The latter has become very popular in the financial industry due to its simplicity. The SLA can be stated as follows: If G is the true underlying severity distribution function of the individual losses and λ the true annual frequency then the $100(1-\gamma)\%$ VaR of the compound loss distribution may be approximated by $G^{-1}(1-\gamma/\lambda)$ or, as modified by Degen (2010) for large λ , by $G^{-1}(1-\gamma/\lambda) + \lambda\mu$ where μ is the finite mean of the true underlying severity distribution. The SLA currently receives a lot of attention as judged by several recent papers on the topic (see e.g. Hernandez et al. 2013 &

2014, Peters et al. 2013). Because the limitations of the technique are not yet fully understood, we will use the MC approaches in this paper, although we did find that the SLA often provided useful insights into the results and it will be used for that purpose.

The estimation of EC using the LDA has been discussed in several books and papers (see e.g. McNeil at al. 2005, Nešlehová et al. 2006, Embrechts and Pucetti, 2006 & 2008, Gourier et al. 2009 and Shevchenko 2010). Cope et al. (2009) studied the feasibility of meeting the Basel II measurement standard of estimating the 99.9% percentile of the annual aggregate loss distribution. They examined the possibility of meeting this measurement standard given the amount of operational loss data that is currently available from either internal or external sources, and pointed out difficulties that arise in applying the LDA to computing operational risk exposures, as well as in validating the capital models. Cope et al. (2009) conclude that many of these problems are insurmountable and suggest some changes to the regulatory framework that would circumvent these difficulties. One possible change is to estimate the 95th percentile of the compound distribution in order to improve the reliability and stability of the quantile estimate. To reach the 99.9th percentile, the 95th percentile is then multiplied with a fixed factor determined by the regulator to scale the capital estimate obtained at the lower percentile. In particular Cope et al. (2009) found that the heaviness of the tail of the loss severity distribution has much more to do with a reduction in capital charges than the diversification benefit obtained through more realistic dependency modelling options (e.g. the use of copulas). Generally, heavy-tailed data sets are hard to model and require much caution when interpreting the resulting capital charges. For example, a single extreme loss can cause drastic changes in the estimate of the means and variances of severity distributions even if a large amount of loss data is available. Annual aggregate operational losses for an ORC will typically be driven by the value of the most extreme losses and the high quantiles of the aggregate annual loss distribution are primarily determined by the high quantiles of the severity distributions containing the extreme losses. Use of diverse severity distributions will result in widely varying capital charges and the use of a Generalised Pareto Distribution (GPD) for modeling the tail of the severity distribution should be regarded as conservative and can provide inaccurate estimates, especially when interpolating beyond the available data. Two different severity distributions for modelling the individual losses within the ORC may both fit the data well in terms of goodness-of-fit statistics, yet may provide capital estimates which may differ by billions. In addition to Cope et al. (2009), Embrechts and Hofert (2011) have highlighted several deficiencies and expressed areas of concern about elements of the LDA approach, in particular, the estimation of the severity distribution and the subsequent estimation of an extreme VaR of the aggregate loss distribution. Common sense suggests that if historical data are to be used, at least 1000 years of loss data for the ORC in question would be required to estimate such a quantity with reasonable accuracy. Typically banks have available between 5 and 10 years of internal loss data, and in order to circumvent this problem, and as suggested by the Basel Accord, external data may be used to improve estimation of the severity distribution of the annual individual losses. External data are available in various pooled resources like the proprietary ORX database (see ORX, 2008) and SAS OpRisk Global Data, a repository of publicly reported operational risk losses. Integrating the external data with internal data is not straight forward (see e.g. Lambrigger et al., 2007, Figini et al., 2008, Gustafsson and Nielsen, 2008, Dahen and Dionne, 2010, and Bolancé et al., 2013).

The Basel Accord (see BCBS 195 and BCBS 196, 2011) also suggests the use of scenario assessments to improve severity distribution estimation. These assessments are typically regarded as 'forward-looking' and need to be integrated with the 'backward-looking' historical data. The BCBS 196 (2011) paper refers to three types of scenarios, namely the individual scenario approach, the interval approach and the percentile approach. Unfortunately the description of these methods is not clear. Various approaches to incorporating scenario assessments into the modeling of the severity distribution have been proposed in the literature (see e.g. Agostini et al., 2010, Cope, 2012, Dutta and Babbel, 2013 and Lambrigger et al., 2007). Agostini et al. (2010) modeled the body and tail of the severity distribution separately, using the empirical distribution to model the body and the GPD to model the tail. The basic idea of their (integration) methodology is to estimate the parameters of the frequency and severity distributions based on historical losses, and correct them, via advanced credibility theory, to include the information coming from scenario assessments. They follow a 3-step approach where firstly they obtain estimates of the parameters of the severity and frequency distribution by using historical data, secondly they obtain subjective estimates of the parameters of the severity and frequency distribution by using scenario assessments, and thirdly they use credibility theory (see Buhlmann, et al. 2007) to integrate the historical and subjective estimates. One of their assumptions in the second step is that the worst case scenario assessment is equal to the 99.9% quantile of the severity distribution (GDP in their case). Cope (2012) studied the stochastic process that generates losses within an ORC and models this as a superposition of various sub-processes that characterize individual loss-generating mechanisms. He then proposed a method for identifying loss-generating mechanisms, performing scenario assessments and combining the outcomes with relevant historical loss data to compute an aggregate loss distribution for the ORC. Dutta and Babbel (2013) proposed a method that combines scenario assessments with historical loss data using the so-called Change of Measure approach, which enabled them to evaluate the impact of each scenario on the total estimate of operational risk capital. The type of scenario assessments required by their approach is different to what is studied here in that the expert is presented with a description of an event based on which a range estimate of the severity of the event and an estimate of the frequency of the occurrence of the event over a specified number of years are required. Lambrigger et al. (2007) proposed a Bayesian inference method that allows the combination of internal and external data as well as scenario assessments. Bayesian methods involve the estimation of a prior density, of which the parameter is estimated by expert opinion, which is then multiplied by the likelihood of the data in order to find a posterior distribution of the parameter.

In this paper our focus will be to investigate another approach, which will be referred to as the 1-in-c years scenario approach. We encountered it when we were involved in the validation of the AMA Operational Risk EC models of banks. In particular, we are interested in studying this methodology in terms of scenario consistency and the sensitivity of EC estimates and severity distribution fits to variation in expert scenario assessments. We will show that as it stands, this methodology has shortcomings in integrating the historical and scenario data in an ORC. We then propose an alternative methodology which addresses these shortcomings while also providing measures for assessing the agreement between the two sources of data. The methodologies are evaluated by means of a Monte Carlo simulation study in terms of its ability to provide reliable and stable capital estimates in a given ORC. Since the method proposed here differs from the previous proposals cited in the literature, we were not able to include these methods in our studies. Any attempt to compare these methods would be difficult because the scenarios and the way in which the scenarios are assessed differ significantly. It should be noted that probability assessments by experts are notoriously difficult and unreliable as discussed by Kahneman et al. (1982) and Kahneman (2011).

The layout of the paper is as follows. In the next section we describe the 1-in-c years approach and analyse its shortcomings in terms of integrating historical data and scenario assessments. In Section 3, using the same data, we propose the alternative method for integrating the historical data with the scenario assessments. Then, in Section 4, we analyse the SAS OpRisk Global data in order to establish which distribution is suitable as a model for the true underlying severity distribution in our simulation study. It turns out that the Burr distribution is a reasonable model for this purpose. In Section 5, we compare the performance of the two methods by means of a Monte Carlo simulation study. Through this study we demonstrate that the new method outperforms the traditional 1-in-c years scenario method, especially when experts provide assessments that differ from the true quantiles, and the degree of outperformance increases with increasing unreliability in the experts' assessments and greater amounts of historical data used in the integration process. Section 6 contains some concluding remarks and suggestions for future research.

2 Description of the 1-in- c year method

In the 1-in-c years scenario approach the scenario makers are asked to answer the question: **'What loss level q_c is expected to be exceeded once every c years?'**. Our validation of the AMA models of banks suggested that popular choices for c vary between 5 and 100 and often 3 values for c are used. As an example, one bank used $c = 7, 20 \text{ and } 100$ and motivated the first choice as the number of years of historical data available to them. In this case the largest loss in the historical data may serve as a guide for choosing q_7 since this loss level has been reached once in 7 years. If the scenario makers judge that the future will be better than the past, they may want to provide a lower assessment for q_7 than the largest loss experienced so far. If they foresee deterioration they may judge that a higher assessment is more appropriate. The other choices of c are selected in order to obtain a scenario spread within the range that one can expect reasonable improvement in accuracy from the experts' inputs. Of course the choice of $c = 100$ may be questionable because judgments on a 1-in-100 years loss level are likely to fall outside many of the experts' experience. They may however, take additional guidance from external data of similar banks which in effect amplifies the number of years for which historical data are available. Of course requiring that the other banks are similar to the bank in question may be a difficult issue and the scaling of external data in an effort to make it comparable to the bank's own internal data raises further problems (see e.g. Figini et al., 2008, Dahen and Dionne, 2010, Embrechts and Hofert, 2011). We will not dwell on this issue here and henceforth assume that we do have the 1-in-c years scenario assessments for a range of c-values, but have to keep in mind that subjective elements may have affected the reliability of the assessments.

If the annual loss frequency is $Poi(\lambda)$ distributed and the true underlying severity distribution is G, and if the scenario makers are of oracle quality in the sense of actually knowing λ and G, then the assessments provided should be

$$q_c = G^{-1}(1 - \tfrac{1}{c\lambda}). \tag{1}$$

To see this, let N_c denote the number of loss events experienced in c years and let M_c denote the number of these that are actually greater than q_c. Then $N_c \sim Poi(c\lambda)$ and the conditional distribution of M_c given N_c is binomial with parameters N_c and $p_c = P(\mathrm{X} \geq q_c) = 1 - G(q_c)$ with $X \sim G$. Therefore $EM_c = E[E(M_c|N_c)] = E[N_c p_c] = c\lambda(1 - G(q_c))$. Requiring that $EM_c = 1$ leads to (1). Notice that (1) involves the annual frequency λ and scenario makers should be made aware of and take this feature into account when they produce their assessments. Of course real scenario makers are not oracles and their assessments will not actually be equal to q_c in (1). To acknowledge this in the notation we denote their assessments by $\tilde{q}_c$.

In the existing form of the 1-in-c years method, the backward-looking historical information is 'integrated' with the forward-looking scenario information to obtain one severity distribution by splicing the distribution in an 'expected' (or 'body') part and an 'unexpected' (or 'tail') part. Estimation of the expected part is based on the historical data and estimation of the unexpected part on the scenario assessments. To formulate and explain the procedure in more detail, select a number b and let q_b be the corresponding quantile given by (1), i.e. $q_b = G^{-1}(1 - \frac{1}{b\lambda})$. We consider the interval below q_b as the expected part and the interval above q_b as the unexpected part of the severity distribution. Define two distribution functions

$$G_e(x) = G(x)/G(q_b) \text{ for } x \le q_b \text{ and}$$
$$G_u(x) = [G(x) - G(q_b)]/[1 - G(q_b)] \text{ for } x > q_b, \quad (2)$$

i.e. $G_e(x)$ is the conditional distribution function of a random loss $X \sim G$ given that $X \le q_b$ and $G_u(x)$ is the conditional distribution function given that $X > q_b$. Notice that we then have the identity

$$G(x) = (1 - \frac{1}{b\lambda})G_e(x) + \frac{1}{b\lambda}G_u(x) \text{ for all } x \, . \quad (3)$$

This formula can be used as a basis for estimating G as we now explain.

As in the previous section, suppose we have available a years of historical loss data $x_1, x_2, \ldots, x_K$. Then the annual frequency λ can be estimated as $\hat{\lambda} = K/a$ as noted before. Next b and the threshold q_b must be specified. One possibility is to take b as the smallest of the scenario c-year multiples and to estimate q_b as the corresponding smallest of the scenario assessments $\tilde{q}_b$ provided by the scenario makers. This can be motivated by the considerations explained above. Next $G_e(x)$ can be estimated by fitting a parametric family (such as the Burr) to the data $x_1, x_2, \ldots, x_K$ or by calculating the empirical distribution and then conditioning it to the interval $(0, \tilde{q}_b]$. Either of these estimates is a reasonable choice, especially if K is large and the parametric family is well chosen. Whichever estimate we use, denote it by $\tilde{G}_e(x)$. For the sake of future notational consistency, we shall also put tildes on all estimates of distribution functions which involve use of the scenario assessments.

Next, $G_u(x)$ can be modelled by the GPD given by

$$G_u(x) = GPD(x; \sigma, \xi, q_b) = \begin{cases} 1 - \left[1 + \frac{\xi}{\sigma}(x - q_b)\right]^{\frac{-1}{\xi}} & \xi > 0 \\ 1 - \exp(-\frac{x - q_b}{\sigma}) & \xi = 0, \end{cases} \quad (4)$$

with $x \ge q_b$, thus taking q_b as the so-called EVT threshold and with σ and ξ respectively scale and shape parameters. That this can be done follows from the Pickands-Balkema-de Haan limit theorem (see e.g. McNeil et al., 2005), which states that the conditional tail of all distributions in the domain of attraction of the Generalised Extreme Value distribution (GEV), tends to a GPD distribution. The distributions in the domain of attraction of the GEV are a wide class of distributions, which includes most distributions of interest to us. Although one could consider alternative distributions to the GPD for modelling the tail of a severity distribution, this theorem, and the limiting conditions that we are interested in, suggest that the GPD is a good choice. In our case we will fit this distribution using the scenario assessments as explained next.

Notice that from (1) and (2) it follows that

$$G_u(q_c) = \frac{G(q_c) - G(q_b)}{1 - G(q_b)} = \frac{\left(1 - \frac{1}{c\lambda}\right) - (1 - \frac{1}{b\lambda})}{\frac{1}{b\lambda}} = 1 - \frac{b}{c} \qquad (5)$$

For ease of explanation, suppose we have actual scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$ and thus take $b = 7$ and estimate q_b by $\tilde{q}_7$. Substituting these scenario assessments into (5) with $b = 7$, $c = 20$, 100 yields two equations

$$G_u(\tilde{q}_{20}) = GPD(\tilde{q}_{20}; \sigma, \xi, \tilde{q}_7) = 0.65 \text{ and } G_u(\tilde{q}_{100}) = GPD(\tilde{q}_{100}; \sigma, \xi, \tilde{q}_7) = 0.93 \qquad (6)$$

that can be solved to obtain estimates of the parameters σ and ξ in the GPD that are based on the scenario assessments. With more than three scenario assessments, fitting techniques can be based on (5) which links the quantiles of the GPD to the scenario assessments. An example would be to minimize $\sum_c |GPD(\tilde{q}_c; \sigma, \xi, \tilde{q}_7) - (1 - b/c)|$. Other possibilities include a weighted version of the sum of deviations in this expression or deviation measures comparing the GPD quantiles directly to the q_c assessments. Whichever route we follow, we denote the final estimate of $G_u(x)$ by $\tilde{G}_u(x)$. All these ingredients can now be substituted into (3) to yield the estimate of $G(x)$, namely

$$\tilde{G}(x) = (1 - \frac{1}{b\hat{\lambda}})\tilde{G}_e(x) + \frac{1}{b\hat{\lambda}}\tilde{G}_u(x) \text{ for all } x\,. \qquad (7)$$

Note in passing that (5) does not depend on the annual frequency λ and has interesting suggestions about the formulation of the basic question of the 1-in-c years approach. For example, if we take $b = 1$ then q_1 would be the scenario makers' answer to the question 'What loss level is expected to be exceeded once annually?'. Unless we are dealing with an ORC with only rare loss events, a reasonably accurate assessment of q_1 should be possible.

Then (5) reads $G_u(q_c) = 1 - 1/c$ or $1 - G_u(q_c) = 1/c$. Keeping in mind the conditional probability meaning of G_u this tells us that q_c would be the answer to the question: 'Among those losses that are larger than q_1 what level is expected to be exceeded only once in c years?'. Conditioning on the losses larger than q_1 has the effect that the annual frequency of all losses drops out of consideration when an answer is sought.

Returning now to practical use of equation (7), Algorithm 3 below summarizes the integration of the historical data with the 1-in-c years scenarios following the MC estimation approach. For simplicity of description we assume that we have scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$ and take the threshold between the body and tail parts at $\tilde{q}_7$. Let $\tilde{G}_e(x)$ be the empirical distribution function of the observed loss data conditioned to the body interval $(0, \tilde{q}_7]$. Further, fit a GPD to the scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$ as described above and denote the fitted tail distribution by $\tilde{G}_u(x)$. Substitute these estimates into (7) to obtain an estimate $\tilde{G}$ of G. Note also that we have

$$\hat{\lambda}\tilde{G}(x) = \left(\hat{\lambda} - \frac{1}{7}\right)\tilde{G}_e(x) + \frac{1}{7}\tilde{G}_u(x).$$

Algorithm 3 : GPD Monte Carlo integration approach

(1) Generate $N_e \sim Poi\left(\hat{\lambda} - \frac{1}{7}\right)$ and $N_u \sim Poi\left(\frac{1}{7}\right)$;

(2) Generate $X_1, \ldots, Y_{N_e} \sim \text{iid } \tilde{G}_e$ and $X_{N_e+1}, \ldots, X_{N_e+N_u} \sim \text{iid } \tilde{G}_u$ and calculate $A = \sum_{n=1}^{N} X_n$ where $N = N_u + N_e$. Using the identity above it easily follows that A is distributed as a random sum of N iid losses from $\tilde{G}$.

(3) Repeat (1) and (2) I times independently to obtain $A_i, i = 1,2, \ldots, I$ and estimate the 99.9% VaR by the corresponding empirical quantile of these A_i's as before.

Variations of Algorithm 3 include basing $\tilde{G}_e$ on a distribution fitted to the observed data contained in the body interval and basing $\tilde{G}_u$ on some other heavy tailed distribution instead of the GPD, to fit the scenario assessments.

To close this section some remarks on the SLA are in order. The SLA states that if G is the true underlying severity distribution function assumed to belong to the sub-exponential class and λ is the true annual expected frequency then, under certain limiting conditions, the $100(1-\gamma)\%$ VaR of the aggregate loss distribution may be approximated by $G^{-1}(1-\gamma/\lambda)$. With $\gamma = 0.001 = 1/1000$ this becomes $G^{-1}\left(1-\frac{1}{1000\lambda}\right) = q_{1000}$ in the notation above. Thus the SLA suggests that we may as well ask the scenario maker directly what loss level is exceeded once in 1000 years and use this answer to estimate the EC. The counter arguments are twofold: (a) the SLA gives only an approximation to the EC we are trying to estimate, and (b) experts are very unlikely to have the experience or the information at their disposal to assess a 1-in-1000 year event reliably. One can realistically only expect them to assess events occurring more frequently such as once in 100 years or even more often as once in 30 year. It is interesting to note that Agostini et al. (2010) also make use of expert assessments, which fully determine their expert scenario subjective distributions. They use the GPD to model this distribution and the shape parameter is calibrated by severity worst-case estimation, where the severity worst-case is identified with the 99.9th-quantile of the severity distribution. In our context, this assumption seems unrealistic. For example, if $\lambda = 50$ then $q_7 = G^{-1}(0.99714)$, $q_{20} = G^{-1}(0.999)$ and $q_{100} = G^{-1}(0.9998)$ which implies that the quantiles of interest are much more extreme than the 'worst cases' mentioned above.

3 A new approach to integrating scenario assessments and historical data

When using the GPD 1-in-c years integration approach to model the severity distribution in the LDA, we realised that the regulatory capital or 99.9% VaR of the aggregate distribution is almost exclusively determined by the scenario assessments and their reliability greatly affects the reliability of the resulting regulatory capital. The research around the SLA further supports this conclusion, since it is known that the SLA provides a good approximation of the VaR of the aggregate distribution when considering very high quantiles (such as 99.9%) and a heavy-tailed severity distribution with Extreme Value Index (EVI) >1 from the sub-exponential class. The challenge is therefore to find a way of integrating the historical data and scenario assessments such that both sets of information are adequately utilised in the process. In particular, it would be nice to have measures indicating whether the scenario experts' assessments are in line with the observed historical data and if not, to require them to produce reasons why their assessments are so different. Below we describe an estimation method that will meet these aims.

We continue with the previous notation and also need the following items. As before let $F(x;\theta)$ denote a suitable family of distribution functions to model the true severity distribution G and assume that these distributions have support on the full interval $[0,\infty)$. Also as before fit $F(x;\theta)$ to the historical data and denote the parameter estimate by $\hat{\theta}$. For ease of explanation we assume that scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$ are available and let $\hat{p}_7 = 1 - \frac{1}{7\hat{\lambda}}$, $\hat{p}_{20} = 1 - \frac{1}{20\hat{\lambda}}$ and $\hat{p}_{100} = 1 - \frac{1}{100\hat{\lambda}}$ be estimates of the corresponding true probabilities $p_7 = 1 - \frac{1}{7\lambda}$, $p_{20} = 1 - \frac{1}{20\lambda}$ and $p_{100} = 1 - \frac{1}{100\lambda}$ occurring in (1). Further introduce the ratios

$$R(7) = \frac{\hat{p}_7}{F(\tilde{q}_7;\hat{\theta})}\ ,\ \ R(7,20) = \frac{\hat{p}_{20}-\hat{p}_7}{F(\tilde{q}_{20};\hat{\theta})-F(\tilde{q}_7;\hat{\theta})},$$
$$R(20,100) = \frac{\hat{p}_{100}-\hat{p}_{20}}{F(\tilde{q}_{100};\hat{\theta})-F(\tilde{q}_{20};\hat{\theta})} \text{ and } R(100) = \frac{1-\hat{p}_{100}}{1-F(\tilde{q}_{100};\hat{\theta})}\ . \tag{8}$$

Notice that if our estimates were actually exactly equal to what they are estimating, these ratios would all be equal to 1. For example, we would then have $R(7) = p_7/G(q_7) = 1$ by (1), and similarly for the others. Our new method is to estimate the true severity distribution function G by an adjusted form of $F\left(x,\hat{\theta}\right)$, denoted by $\widetilde{H}$ and defined as follows:

$$\widetilde{H}(x) = \begin{cases} R(7)F(x;\hat{\theta}) & for\ \ x \leq \tilde{q}_7 \\ \hat{p}_7 + R(7,20)[F(x;\hat{\theta}) - F(\tilde{q}_7;\hat{\theta})] & for\ \ \tilde{q}_7 < x \leq \tilde{q}_{20} \\ \hat{p}_{20} + R(20,100)[F(x;\hat{\theta}) - F(\tilde{q}_{20};\hat{\theta})] & for\ \ \tilde{q}_{20} < x \leq \tilde{q}_{100} \\ \hat{p}_{100} + R(100)[F(x;\hat{\theta}) - F(\tilde{q}_{100};\hat{\theta})] & for\ \ \tilde{q}_{100} < x < \infty. \end{cases} \tag{9}$$

Notice again that this estimate is consistent in the sense that it actually reduces to G if all estimators are exactly equal to what they are estimating. Also note that $\widetilde{H}(\tilde{q}_7) = \hat{p}_7$, $\widetilde{H}(\tilde{q}_{20}) = \hat{p}_{20}$ and $\widetilde{H}(\tilde{q}_{100}) = \hat{p}_{100}$, i.e. the equivalents of (1) hold for the scenario assessments when estimates are substituted for the true unknowns. Hence at the estimation level the scenario assessments are consistent with the probability requirements expressed by (1). Thus this new estimated severity distribution estimate $\widetilde{H}$ 'believes' the scenario quantile information, but follows the distribution fitted on the historical data to the left of, within and to the right of the scenario intervals. The ratios $R(7)$, $R(7{,}20)$, $R(20{,}100)$ and $R(100)$ in (8) can be viewed as measures of agreement between the historical data and the scenario assessments and could be useful for assessing their validities and qualities. The steps required to implement the new method are as follows:

Algorithm 4 : New method (NM) Monte Carlo integration approach

(1) Generate $N \sim Poi(\hat{\lambda})$

(2) Generate $X_1, \dots, X_N \sim \text{iid } \widetilde{H}$ and calculate $A = \sum_{n=1}^{N} X_n$;

(3) Repeat (1) and (2) I times independently to obtain $A_i, i = 1,2, \dots, I$ and estimate the 99.9% VaR by the corresponding empirical quantile of these A_i's as before.

Some variations on these steps will be discussed later and a simulation based evaluation of the new methodology is given in Section 5 below.

Before proceeding some remarks on the properties of $\widetilde{H}$ may be helpful. In Table 1 and Figure 1 below, we illustrate the behaviour of the new method with an artificially constructed example. Taking the expected annual Poisson frequency as $\lambda = 50$ and assuming that the true severity distribution is a Burr XII with parameters all equal to 1 (see (10) below), we obtain $q_7 = 349$, $q_{20} = 999$ and $q_{100} = 4999$. We now adjust these quantiles by multiplying them individually by factors u_7, u_{20} and u_{100} to get $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$. Then we calculate the R's from (8) but replacing the estimated $\hat{p}$'s and $F(\tilde{q}, \hat{\theta})$'s by their assumed true values. For Case 1 of Table 1 no adjustment was done while the other cases were obtained by varying the perturbing factors over the values 0.7, 1 and 1.3. We see that $R(7)$ stays close to 1 in all cases. This happens because numerator and denominator in its definition are both close to 1. However, the other R's respond upward or downward in the direction corresponding to the relevant perturbation. They therefore carry information regarding the direction and extent to which the scenario assessments differ from what the true underlying severity distribution would predict.

In order to study the tail behaviour of the true underlying distribution in comparison to $\widetilde{H}$ and the GDP fitted to a set of perturbed quantiles, we plot the log of 1 minus the fitted distribution functions for case 10 and case 15 in Figure 1. From Figure 1 it is clear that the new method $\widetilde{H}$ treats the perturbations in the quantiles in such a way that its tail behaviour is closer to the true underlying distribution than the fitted GPD.

Table 1: Measures of agreement for different combinations of perturbed quantiles in illustrative example.

Case	u_7	u_{20}	u_{100}	R(7)	R(7,20)	R(20,100)	R(100)
1	1	1	1	1.000	1.000	1.000	1.000
2	0.7	0.7	0.7	1.001	0.701	0.700	0.700
3	0.7	0.7	1	1.001	0.701	0.651	1.000
4	0.7	0.7	1.3	1.001	0.701	0.628	1.300
5	0.7	1	0.7	1.001	0.604	1.120	0.700
6	0.7	1	1	1.001	0.604	1.000	1.000
7	0.7	1	1.3	1.001	0.604	0.945	1.300
8	0.7	1.3	0.7	1.001	0.562	1.654	0.700
9	0.7	1.3	1	1.001	0.562	1.405	1.000
10	0.7	1.3	1.3	1.001	0.562	1.300	1.300
11	1	0.7	0.7	1.000	1.299	0.700	0.700
12	1	0.7	1	1.000	1.299	0.651	1.000
13	1	0.7	1.3	1.000	1.299	0.628	1.300
14	1	1	0.7	1.000	1.000	1.120	0.700
15	1	1	1.3	1.000	1.000	0.945	1.300
16	1	1.3	0.7	1.000	0.890	1.654	0.700
17	1	1.3	1	1.000	0.890	1.405	1.000
18	1	1.3	1.3	1.000	0.890	1.300	1.300
19	1.3	0.7	0.7	0.999	2.408	0.700	0.700
20	1.3	0.7	1	0.999	2.408	0.651	1.000
21	1.3	0.7	1.3	0.999	2.408	0.628	1.300
22	1.3	1	0.7	0.999	1.549	1.120	0.700
23	1.3	1	1	0.999	1.549	1.000	1.000
24	1.3	1	1.3	0.999	1.549	0.945	1.300
25	1.3	1.3	0.7	0.999	1.299	1.654	0.700
26	1.3	1.3	1	0.999	1.299	1.405	1.000
27	1.3	1.3	1.3	0.999	1.299	1.300	1.300

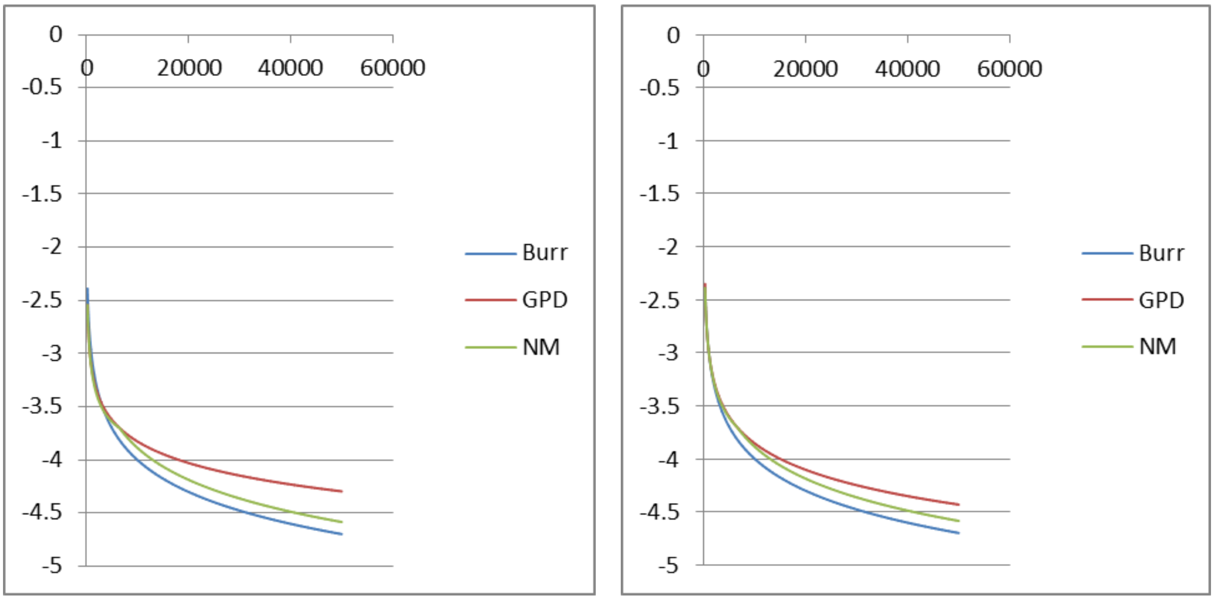


Figure 1: Tail behaviour of distributions for Case 10 on the left and Case 15 on the right. The blue line represents the tail of the true underlying Burr and the red and green lines the fitted GPD and NM respectively.

The SLA again sheds some light on the new method. As noted above the SLA implies that we need to estimate $q_{1000} = G^{-1}\left(1 - \frac{1}{1000\lambda}\right)$ and its estimate would be $\hat{q}_{1000} = \widetilde{H}^{-1}\left(1 - \frac{1}{1000\hat{\lambda}}\right) = \widetilde{H}^{-1}(\hat{p}_{1000})$. Some algebra shows that the equation $F(\hat{q}_{1000}; \hat{\theta}) = F(\tilde{q}_{100}; \hat{\theta}) + (\hat{p}_{1000} - \hat{p}_{100})/R(100)$ needs to be solved for $\hat{q}_{1000}$. Depending on the choice of the family of distributions (x, θ) , this may be easy (e.g. when we use the Burr family for which we have an explicit expression for its quantile function). This clearly shows that a combination of the historical data and scenario assessments is involved, and not exclusively the latter. In as much as the SLA provides an approximate to the actual VaR of the aggregate loss distribution, we may expect the same to hold for the NM integration approach.

4 Analysis of Operational Risk data base

To motivate the choice of models and input parameters for our simulation study, we analysed the repository of information on publicly reported operational losses, SAS OpRisk Global Data, using SAS PROC SEVERITY. In particular we were interested in determining the typical frequencies observed in ORC's, appropriate choices of severity distributions as models for the severity loss distribution and the typical parameter estimates of the most popular distribution. SAS OpRisk Global data contains losses collected over a 23-year period in 70 business line/event categories (10 business lines and 7 events). This data base contains publicly reported operational losses in excess of USD 100 000. The annual frequencies of each ORC vary between 0 and 321 with an average frequency of about 20. Of the 70 categories we identified 53 as suitable for our analysis based on data availability. For each of the 53 ORC's we used SAS PROC SEVERITY to fit the loss severity data by the range of available distributions in this PROC (Burr, Exponential, Gamma, Pareto, Lognormal, Weibull, GPD and Inverse-Gaussian). Each of these distributions was fitted conditional on the left truncation of the data at USD 100 000.

We then evaluated the best fitting distributions by three goodness-of-fit test statistics, namely Anderson-Darling (AD), Kolmogorov-Smirnoff (KS) and Cramer-von Mises (CM) and ranked the distributions in terms of their relative goodness-of-fit performance in each ORC. The results are given in Table 2 below and should be interpreted as follows. In the second column the Burr fit achieved the best fit (rank 1) according to the KS statistic in 19 of the 53 ORC's, the second best fit (rank 2) in 13 of the ORC's and third best fit (rank 3) in 12 of the ORC's, etc. The results in the remaining columns should be interpreted in the same way. Note that we did not consider separate models for the body and the tail of the distribution, for example a Lognormal body with a GPD tail, mainly due to the fact that our focus is on finding a distribution which may be a good model for the unknown true underlying distribution. This distribution can then be used in the simulation study as a model for the true underlying distribution. According to the results displayed in Table 2 below, the Burr distribution was the clear overall best choice, obtaining a top three position in 83% (KS), 96% (AD) and 92% (CM) of the ORC's. The closest contender of the distributions considered here, was the Lognormal.

Table 2: Number of ranks obtained by different distributions for the Anderson-Darling (AD), Kolmogorov-Smirnoff (KS) and Cramer-von Mises (CM) statistic in the 53 ORC's.

Rank	Burr			Exponential			Gamma			Pareto			Lognormal			Weibull			GPD			Igauss		
	KS	AD	CM	KS	AD	CM	KS	AD	CM	KS	AD	CM	KS	AD	CM	KS	AD	CM	KS	AD	CM	KS	AD	CM
1	19	29	26	0	0	0	5	7	7	7	1	3	11	12	9	6	3	4	4	1	4	1	0	0
2	13	12	10	0	0	0	1	0	0	8	17	12	10	6	6	3	11	10	15	5	11	3	2	4
3	12	10	13	0	0	0	2	1	3	14	12	13	6	5	5	7	5	6	12	19	13	0	1	0
4	7	2	3	0	0	0	2	3	2	10	12	12	20	25	26	0	0	1	10	8	7	4	3	2
5	0	0	1	0	0	0	1	0	0	10	8	8	4	5	5	28	22	22	5	12	12	5	6	5
6	2	0	0	0	0	0	17	18	19	3	3	5	1	0	2	9	12	10	7	8	6	14	12	11
7	0	0	0	2	1	3	25	24	22	1	0	0	1	0	0	0	0	0	0	0	0	24	28	28
8	0	0	0	51	52	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	3

Based on the good performance of the Burr observed in each of the 53 ORC's in the SAS OpRisk Global data, we will focus our subsequent studies on the three parameter Burr type XII distribution function

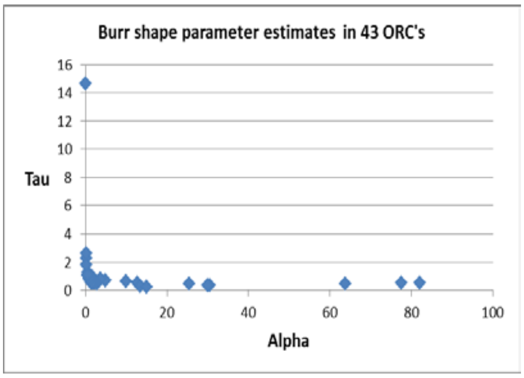
$$B(x; \eta, \tau, \alpha) = 1 - (1 + (x/\eta)^{\tau})^{-\alpha}, \text{ for } x > 0 \qquad (10)$$

with parameters $\eta, \tau, \alpha > 0$ (see e.g. Beirlant et al., 2004). Here η is a scale parameter and τ and α shape parameters. Note the EVI of the Burr distribution is given by $EVI = \zeta = 1/\tau\alpha$ 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, having a distribution function of the form $1 - F(x) = x^{-1/\zeta}\ell_F(x)$, with $\ell_F(x)$ a slowly varying function at infinity (see e.g. Embrechts et al., 1997). For Pareto-type, when the EVI>1, the expected value does not exist, and when EVI>0.5, the variance is infinite. Note also that the Burr distribution is regularly varying with index $-\tau\alpha$ and therefore belongs to the class of sub-exponential distributions (see Fasen and Kluppelberg, 2006). In Table 3 the Burr fits for 43 of the 53 ORC's are provided, because PROC SEVERITY indicated that the Burr fit 'converged' in only 43 ORC's. Clearly, the shape parameters play an important role in determining heaviness of the tail. Because of this, we plot the estimates of the shape parameters obtained in the 43 ORC'S in Figure 2 below.

Table 3: Parameter estimates of Burr converged fits to 43 ORC's

Number	Event type	Business Line	N	$\hat{\lambda}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\tau}$	EVI
1	External Fraud	Payment and Settlement	26	1.13	28803.16	82.146	0.547	0.022
2	Employment Practices and Workplace Safety	Asset Management	13	0.57	21377.51	77.533	0.542	0.024
3	Execution, Delivery & Process Management	Commercial Banking	58	2.52	38473.25	63.854	0.446	0.035
4	External Fraud	Asset Management	25	1.09	24293.89	25.444	0.478	0.082
5	Execution, Delivery & Process Management	Payment and Settlement	29	1.26	24953.96	30.537	0.362	0.090
6	Execution, Delivery & Process Management	Insurance	52	2.26	26703.86	30.022	0.367	0.091
7	External Fraud	Retail Brokerage	44	1.91	86.88	10.033	0.655	0.152
8	External Fraud	Corporate Finance	24	1.04	7320.73	12.796	0.511	0.153
9	Internal Fraud	Trading & Sales	146	6.35	35289.11	13.522	0.312	0.237
10	Damage to Physical Assets	Retail Banking	52	2.26	16317.41	15.000	0.237	0.281
11	Clients, Products & Business Practices	Agency Services	62	2.70	246.91	4.879	0.676	0.303
12	Internal Fraud	Agency Services	14	0.61	73.53	3.677	0.821	0.331
13	Employment Practices and Workplace Safety	Commercial Banking	28	1.22	6.81	1.709	0.956	0.612
14	Clients, Products & Business Practices	Corporate Finance	433	18.83	168.02	2.809	0.491	0.725
15	Employment Practices and Workplace Safety	Trading & Sales	36	1.57	4.88	1.316	0.933	0.815
16	Clients, Products & Business Practices	Insurance	1228	53.39	22.19	1.971	0.614	0.827
17	Execution, Delivery & Process Management	Retail Banking	126	5.48	2.75	2.260	0.533	0.830
18	Clients, Products & Business Practices	Asset Management	419	18.22	37.85	2.299	0.518	0.841
19	External Fraud	Retail Banking	1648	71.65	0.67	1.441	0.756	0.918
20	Execution, Delivery & Process Management	Non-FS	242	10.52	2.82	1.269	0.826	0.954
21	Internal Fraud	Commercial Banking	487	21.17	26.21	2.054	0.509	0.957
22	External Fraud	Agency Services	14	0.61	14.24	1.026	0.993	0.981
23	Internal Fraud	Retail Brokerage	194	8.43	2.33	1.323	0.754	1.003
24	Internal Fraud	Insurance	297	12.91	2.53	1.342	0.718	1.038
25	Internal Fraud	Asset Management	116	5.04	5.29	1.293	0.737	1.050
26	Clients, Products & Business Practices	Non-FS	7386	321.13	10.56	1.223	0.754	1.084
27	Clients, Products & Business Practices	Retail Brokerage	606	26.35	2.55	1.745	0.516	1.110
28	External Fraud	Commercial Banking	836	36.35	2.44	0.821	1.073	1.136
29	External Fraud	Non-FS	436	18.96	3.18	1.013	0.805	1.226
30	Employment Practices and Workplace Safety	Non-FS	923	40.13	2.43	0.840	0.967	1.231
31	Clients, Products & Business Practices	Commercial Banking	377	16.39	10.90	1.367	0.589	1.242
32	Employment Practices and Workplace Safety	Retail Brokerage	52	2.26	3.56	0.897	0.893	1.248
33	Internal Fraud	Retail Banking	1442	62.70	0.45	0.971	0.816	1.263
34	Clients, Products & Business Practices	Retail Banking	1046	45.48	10.50	1.464	0.527	1.296
35	Business Disruption and System Failures	Non-FS	184	8.00	3.93	1.069	0.697	1.342
36	Internal Fraud	Corporate Finance	44	1.91	2.34	0.868	0.850	1.355
37	Execution, Delivery & Process Management	Retail Brokerage	107	4.65	0.12	0.049	14.673	1.395
38	Employment Practices and Workplace Safety	Retail Banking	68	2.96	1.25	0.638	1.071	1.464
39	Damage to Physical Assets	Non-FS	2966	128.96	2.18	0.510	1.255	1.562
40	Employment Practices and Workplace Safety	Corporate Finance	36	1.57	1.49	0.284	2.234	1.573
41	Internal Fraud	Non-FS	995	43.26	1.46	0.525	1.112	1.712
42	Employment Practices and Workplace Safety	Insurance	54	2.35	1.34	0.308	1.807	1.798
43	Execution, Delivery & Process Management	Agency Services	16	0.70	0.35	0.169	2.595	2.280

The graph on the left hand side in Figure 2 contains a scatter plot of the Burr shape parameter estimates obtained in 43 ORC's, and in the graph on the right hand side the scatter plot of the cluster of points in the lower corner of the left hand graph is enlarged. The plotted estimates exhibit an interesting inverse relationship induced by the formula $EVI = \frac{1}{\tau\alpha}$. In order to illustrate this relationship, we have plotted the average EVI obtained over the 43 ORC's ($EVI = 0.899$) as the red curve on the graph on the right hand side of Figure 2.



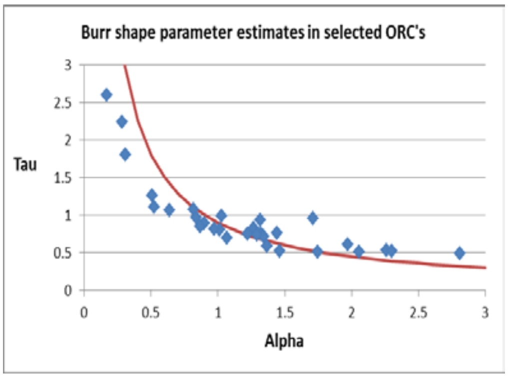


Figure 2: Estimates of the Burr shape parameters for 43 ORC's (left) and selected ORC's (right).The red curve represents the function $EVI = 0.899 = \frac{1}{\tau\alpha}$.

Due to this relationship we decided to base the simulation study on the set of parameters given in Table 4 below. In order to conduct a simulation study we need, apart from a true underlying severity distribution function G, an annual frequency λ and scenario assessments q_7, q_{20} and q_{100}, with $q_c = G^{-1}(1 - \frac{1}{c\lambda})$ for $c = 7, 20$ and 100. The frequencies observed in the ORC's suggest that the choices $\lambda = 10, 30, 50, 100, 200, 300$ would provide a reasonable range of values for the simulation study.

Table 4: Burr parameters suggested for the simulation study

η	α	τ	EVI
1.00	5.00	0.60	0.33
1.00	2.00	0.60	0.83
1.00	1.00	1.00	1.00
1.00	1.50	0.50	1.33
1.00	0.30	1.80	1.85
1.00	0.17	2.50	2.35

Note that the γ-th quantile of the Burr distribution is $q(\gamma) = B^{-1}(\gamma; \eta, \tau, \alpha) = \eta\left((1 - \gamma)^{-1/\alpha} - 1\right)^{1/\tau}$, so that $q_c = \eta\left((c\lambda)^{1/\alpha} - 1\right)^{1/\tau}$ from (1). If all the quantiles are changed proportionately by the same scale factor, the shape parameters of the Burr will be unaffected and only the scale parameter will be changed proportionately. For example, in our case, if the oracle incorrectly adds 50% to all three quantiles, the shape parameters will be unaffected, but the scale parameter will change to 1.5η and the capital amount or VaR will increase by 50%.

5 A simulation based study of the new methodology

The simulation study assumes that the true underlying severity distribution is Burr Type XII as given in (10) and the parameters will be varied across the values given in Table 4. The annual Poisson frequencies (λ) listed just above Table 4 are used and throughout we base the study on 7 years of simulated "historical loss" data. We also need a way to simulate the scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$ that may be provided by the "experts", but it is not immediately evident how this can be done realistically.

Below we report mainly on the following **basic scenario simulation scheme**. For each parameter configuration calculate the three true quantiles q_7, q_{20} and q_{100} from (1) applied to the distribution (10). For each simulation run, choose three perturbation factors u_7, u_{20} and u_{100} independently and uniformly distributed over the interval $[1-\epsilon, 1+\epsilon]$ and then tentatively take $\tilde{q}_7 = u_7 q_7, \tilde{q}_{20} = u_{20} q_{20}$ and $\tilde{q}_{100} = u_{100} q_{100}$ but truncate these so that the final values are increasing, i.e. $\tilde{q}_7 \leq \tilde{q}_{20} \leq \tilde{q}_{100}$. Here the fraction ϵ expresses the size or extent of the possible deviations (or mistakes) inherent in the scenario assessments. If $\epsilon = 0$ then the assessments are completely correct (within the simulation context) and the scenario makers are in effect oracles. In practice choosing $\epsilon > 0$ is more realistic, but how large the choice should be is not clear and the best we can do is to vary ϵ over a range of values. We chose the values 0, 0.1, 0.2, 0.3 and 0.4 for this purpose in the results below. Choosing the perturbation factors to be uniformly distributed over the interval $[1-\epsilon, 1+\epsilon]$ implies that on average they have the value 1, i.e. the scenario assessments are about unbiased. This may not be realistic and other choices are possible, e.g. we could mimic a pessimistic scenario maker by taking the perturbations to be distributed on the interval $[1, 1+\epsilon]$ and an optimistic scenario maker by taking them on the interval $[1-\epsilon, 1]$. There are many other possible distributions for the perturbations that could be assumed and we shall comment briefly below on the differences such other scenario simulation schemes make to the results reported for the basic scheme. Ultimately fully realistic scenario simulation methodology can only be established by doing an empirical study first on the properties of real scenario data but this is outside the scope of the present paper.

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 and for each choice of the perturbation size parameter ϵ work through the following items:

(a) Use Algorithm 1 in Section 1 to determine the 99.9% VaR as described in Steps (1)-(3). Note that the value obtained here approximately equals the true 99.9% VaR; the only approximation involved is that it is based on 1 million repetitions rather than infinitely many. We refer to this value as the **approximately true (AT)** VaR.

(b) Generate a data set of historical losses, i.e. generate $K \sim Poi(7\lambda)$ and then generate $x_1, x_2, \dots, x_K \sim iid$ Burr Type XII with the current choice of parameters. Apply Algorithm 2 to estimate the 99.9% VaR. Here the family $F(x, \theta)$ is chosen as the Burr Type XII but it is refitted to the generated historical data to estimate the parameters as required. The resulting estimate is based on historical data only, and uses no scenario information (NSI) and we refer to it below as the **NSI** VaR estimate.

(c) Add to the historical losses of item (2) three scenarios $\tilde{q}_7, \tilde{q}_{20}, \tilde{q}_{100}$ generated by the basic scenario simulation scheme explained above. Apply Algorithm 3 to estimate the 99.9% VaR. This estimate uses both the historical and the scenario data and we refer to it as the **GPDI** VaR estimate.

(d) Using the historical losses and the three scenarios of item (c), calculate the severity distribution estimate $\widetilde{H}$ and apply Algorithm 4 to estimate the 99.9% VaR. This estimate integrates the historical and the scenario data by the new method and we refer to it below as the **NMI** VaR estimate.

(e) Repeat items (a)-(d) J times and then summarise and compare the resulting VaR estimates.

We have carried out this program with the choice $J = 1000$. Note that each of the items (a) to (d) embed 1 million simulation repetitions and the overall program is carried out for many different choices of parameters and perturbation sizes. So the total workload is quite demanding. While it would be desirable to take $J > 1000$ we also have to keep this workload demand in mind. We found that 1000 repetitions were reasonably manageable and lead to clearly interpretable results which we present and discuss next.

For each combination of parameter values and perturbation sizes the output of the program detailed above results in 1000 AT VaR values as well as 1000 NSI, GPDI and NMI VaR estimates. As mentioned under item (a), those 1000 AT values each approximate the true 99.9% VaR and a summary measure such as their mean or median is very close to the true 99.9% VaR. 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 MedAT. Then the 1000 repeated NSI, GPDI and NMI values may be taken as estimates of MedAT and we wish to decide which of the three methods is best in the sense of coming closest to MedAT. To express the quality of the methods we shall use their median absolute relative deviation (MARD) from MedAT. For example, if NSI_j denotes the NSI value of the j-th repetition, then MARD(NSI) is $Median\{\left|NSI_j/MedAT-1\right|, j=1,2,\ldots,1000\}$. Similarly for GPDI and NMI. Note that the relative standard error of MedAT (standard error of MedAT divided by MedAT) is always below 0.3% for the different choices of parameters considered here. As expected the standard error of MedAT decreases as frequency increases and increases as the EVI increases.

Table 5 shows the MARD's for the oracle case (with scenario perturbation size equal to 0) and for all combinations of parameters. First some comments on the NSI method. It is immediately clear that it is poorer than the other two methods in all parameter configurations. As its name implies, NSI uses no scenario information whereas GPDI and NMI both do use it so that this finding is to be expected. Also evident is that NSI improves with increasing λ and again this is to be expected since increasing λ implies increasing amounts of historical data which leads to better distributional fit and therefore to more accurate high quantile estimation. However, this improvement never catches up with the other two methods. The NSI tends to have smaller MARD at low EVI than at high EVI and again this is in line with the shorter severity tails associated with smaller EVI

Looking next at the GPDI method, it performs best in all parameter combinations. It is only slightly sensitive to variation in λ confirming that it depends mostly on the scenario assessments and very little on the amount of historical data available. It also performs better at shorter tailed low EVI cases than at heavier tailed cases. Next the NMI method performs poorer than the GPDI method but its MARD improves substantially with increasing λ, confirming that it does take advantage from the larger amounts of historical data becoming available when λ increases. It too tends to do better in shorter tailed as compared to heavier tailed cases, but the effect is not quite consistent since increasing amounts of historical data seems to help it to handle heavier tail effects well.

Table 5: Median absolute relative deviations of the NSI, GPDI and NMI 99.9% VaR estimators integrating simulated historical data using the Burr XII severity distribution with varying parameter combinations and oracle scenario assessments

Parameters		Method		
EVI	λ	NSI	GPDI	NMI
0.33	10	0.590	0.024	0.291
	30	0.453	0.019	0.186
	50	0.392	0.020	0.149
	100	0.288	0.019	0.098
	200	0.213	0.021	0.072
	300	0.154	0.020	0.050
0.83	10	0.920	0.028	0.553
	30	0.791	0.020	0.339
	50	0.721	0.019	0.279
	100	0.620	0.017	0.202
	200	0.480	0.017	0.134
	300	0.427	0.019	0.111
1.00	10	0.907	0.022	0.458
	30	0.785	0.021	0.269
	50	0.711	0.022	0.218
	100	0.541	0.020	0.150
	200	0.417	0.022	0.102
	300	0.376	0.020	0.086
1.33	10	0.989	0.035	0.732
	30	0.930	0.030	0.444
	50	0.875	0.030	0.365
	100	0.769	0.028	0.262
	200	0.654	0.028	0.184
	300	0.558	0.028	0.151
1.85	10	0.976	0.040	0.504
	30	0.838	0.038	0.294
	50	0.748	0.040	0.225
	100	0.654	0.039	0.179
	200	0.486	0.040	0.112
	300	0.416	0.041	0.093
2.35	10	0.988	0.050	0.552
	30	0.906	0.049	0.323
	50	0.829	0.052	0.249
	100	0.670	0.045	0.178
	200	0.572	0.050	0.132
	300	0.482	0.049	0.111

The oracle case is ideal for the GPDI method since this case provides it with flawless scenario assessments on which to focus regardless of the historical loss data. However, the oracle case is not practically realistic and we next look at more realistic cases in which the scenario perturbation size $\epsilon > 0$, implying that the scenarios are no longer flawless and tend to become more flawed as ϵ increases. Table 6 shows the MARD values for $\epsilon = 0.1, 0.2, 0.3$ and 0.4 and for the same parameter combinations as before. We left out the NSI MARD's since they do not depend on ϵ and have the values already shown in Table 5 and in none of the cases competes with the GPDI and NMI methods. The cases where the GPDI performs better are shown in blue and the cases where NMI is better are shown in green. It is evident that the GPDI performs better than the NMI in low λ combined with small ϵ cases (i.e. smaller amounts of historical data combined with more accurate scenario assessments) whereas the NMI performs better in the opposite cases. Thus the NMI method achieves what it was designed to do, namely when larger amounts of historical data is available it makes effective use of it to ameliorate possible deleterious effects of inaccurate scenario assessments.

Table 6: Median absolute relative deviations of the GPDI and NMI 99.9% VaR estimators integrating simulated historical data using the Burr XII severity distribution with varying parameter combinations and simulated scenario assessments with varying perturbation sizes

Parameters		Perturbation size and method							
		$\epsilon = 0.1$		$\epsilon = 0.2$		$\epsilon = 0.3$		$\epsilon = 0.4$	
EVI	λ	GPDI	NMI	GPDI	NMI	GPDI	NMI	GPDI	NMI
0.33	10	0.182	0.288	0.284	0.301	0.360	0.307	0.415	0.312
	30	0.173	0.189	0.232	0.196	0.322	0.196	0.335	0.210
	50	0.145	0.156	0.213	0.159	0.290	0.164	0.297	0.183
	100	0.135	0.101	0.183	0.108	0.244	0.126	0.246	0.145
	200	0.119	0.070	0.151	0.081	0.192	0.092	0.197	0.113
	300	0.097	0.054	0.126	0.065	0.164	0.079	0.164	0.094
0.83	10	0.176	0.550	0.318	0.549	0.484	0.554	0.610	0.562
	30	0.181	0.341	0.302	0.343	0.517	0.362	0.608	0.355
	50	0.171	0.282	0.307	0.284	0.501	0.313	0.597	0.323
	100	0.165	0.205	0.318	0.214	0.511	0.232	0.631	0.254
	200	0.177	0.143	0.319	0.154	0.491	0.185	0.636	0.214
	300	0.182	0.120	0.316	0.134	0.472	0.165	0.611	0.197
1.00	10	0.171	0.463	0.300	0.459	0.473	0.471	0.611	0.464
	30	0.165	0.275	0.312	0.284	0.478	0.292	0.606	0.312
	50	0.161	0.228	0.308	0.234	0.458	0.252	0.633	0.288
	100	0.160	0.161	0.311	0.169	0.476	0.197	0.598	0.238
	200	0.168	0.108	0.307	0.136	0.480	0.169	0.613	0.209
	300	0.160	0.096	0.296	0.117	0.460	0.158	0.588	0.204
1.33	10	0.139	0.736	0.294	0.739	0.443	0.726	0.591	0.746
	30	0.150	0.453	0.275	0.447	0.453	0.470	0.550	0.475
	50	0.153	0.363	0.287	0.367	0.446	0.367	0.567	0.396
	100	0.152	0.263	0.282	0.272	0.448	0.297	0.587	0.307
	200	0.144	0.186	0.281	0.206	0.416	0.218	0.562	0.248
	300	0.155	0.156	0.284	0.170	0.453	0.198	0.555	0.234
1.85	10	0.139	0.499	0.261	0.501	0.396	0.512	0.504	0.530
	30	0.139	0.298	0.262	0.303	0.384	0.311	0.527	0.324
	50	0.146	0.225	0.250	0.238	0.364	0.253	0.511	0.279
	100	0.133	0.182	0.255	0.190	0.394	0.204	0.503	0.237
	200	0.141	0.119	0.265	0.135	0.416	0.176	0.509	0.220
	300	0.145	0.101	0.258	0.120	0.404	0.165	0.519	0.207
2.35	10	0.143	0.551	0.240	0.542	0.383	0.552	0.499	0.536
	30	0.146	0.324	0.234	0.328	0.364	0.334	0.496	0.370
	50	0.143	0.261	0.262	0.254	0.392	0.272	0.494	0.325
	100	0.128	0.176	0.242	0.194	0.387	0.218	0.489	0.250
	200	0.137	0.132	0.267	0.155	0.375	0.196	0.486	0.232
	300	0.133	0.118	0.249	0.131	0.359	0.174	0.505	0.192

Furthermore, the MARD of the GPDI varies only slightly with increasing λ, again confirming that it is largely affected only by the quality of the scenario assessments and does not really use the historical data, i.e. it does not really integrate the two sources of information. This contrasts sharply with the NMI which improves strongly as more historical data becomes available. These findings continue to hold if we change the basic scenario generation scheme to reflect other possible features. For example, if we lower the c from 100 to 40 (say) because we expect the corresponding scenario assessment to be more reliable, we find that the performance of GPDI deteriorates relative to the NMI. Again, if we choose the perturbation factors to mimic a pessimistic or an optimistic scenario maker, the GPDI still outperforms the NMI in low λ , small ϵ cases and underperforms in the opposite cases, although the boundary between the two regions shift somewhat. Ultimately, the choice between the two integration methods depends on the extent to which the provided scenario assessments are really trustworthy. In our experience and given the complexities associated with assigning probabilities and costs to events that are simultaneously extremely unlikely but also possibly extremely costly, we think that a scenario simulation scheme with even higher perturbation sizes than those used in the reported results are probably more realistic. This sentiment favours the use of the new integration method.

6 Conclusion and recommendation for future research

Although the existing GPD 1-in-c years scenario approach claims that the scenario quantiles augment the historical data, we find that this is not really true. In line with suggestions from the single loss approximation, the 99.9% VaR of the aggregate losses is determined largely by the scenario assessments and the historical loss data plays little effective role. Therefore, the reliability of the capital estimates will be determined by the reliability with which the experts assess these quantiles. In this paper we proposed a simple new method whereby the severity distribution may be estimated using historical data and experts' scenario assessments jointly. The way in which historical data and scenario assessments are integrated incorporates measures of agreement between these data sources, which can be used to evaluate the quality of both. In particular, we show that the procedure has definite advantages over traditional methods where the severity distribution is modelled and fitted separately for the body and tail parts, with the body part based only on historical data and the tail part on scenario assessments. In a Monte Carlo simulation study we have demonstrated that the new method outperforms the traditional method, especially when the experts provide inconsistent assessments of the true quantiles, and the degree of outperformance increases with increasing amounts of historical data and increasing uncertainty as modelled by the perturbation levels in the assessments.

As far as future research is concerned, the new method should be tested on empirical scenario data in order to take better account of typical uncertainties involved in experts' scenario assessments. For example, assuming that the three perturbation factors involved in the basic scenario simulation scheme are independent may not be entirely realistic and it may be better to assume some form of dependence between them since a scenario maker that over (under) assesses one quantile may be prone to over (under) assessing the others as well. Furthermore, the effects of severity distributions other than the Burr (such as the g-and-h) and frequency distributions other than the Poisson should be investigated. Lastly, the use of the ratios produced during application of the new method requires further work. For example, how much can they deviate from 1 before concerns with the scenario makers should be raised? Among others, this will depend on the amount of historical data used in the analysis as well as on its relevance to the future period for which the regulatory capital is being estimated.

Acknowledgements

We want to thank Paul Embrechts and the referee for comments which improved the presentation of the paper, and acknowledge the grants received by the National Research Foundation (NRF), the Department of Science and Technology (DST) and the Department of Trade and Industry (DTI). *Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF does not accept any liability in regard thereto*.

REFERENCES

Agostini, A; Talamo, P. and Vecchione, V. (2010). Combining operational loss data with expert opinions through advanced credibility theory. Journal of Operational Risk, Vol 5, No 1, 3-28

Aue, F., Kalkbrener, M. (2007). LDA at work: Deutsche Bank's approach to quantifying operational risk. Journal of Operational Risk, 1(4) 49–93

BCBS 195 (2011). Principles for the sound management of Operational Risk; Issued June 2011. Basel Committee on Banking Supervision, Bank of International Settlements.

BCBS 196 (2011). Operational Risk: Supervisory guidelines for the advanced measurement approaches; Issued June 2011. Basel Committee on Banking Supervision, Bank of International Settlements.

Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). Statistics of Extremes: Theory and Applications. John Wiley and Sons.

Bolancé, C., Guillén, M., Gustafsson, J., Nielsen, J.P. (2013), Adding prior knowledge to quantitative operational risk models, Journal of Operational Risk 8(1), 17–32

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

Bühlmann, H., Shevchenko, P.V. & Wüthrich, M.V. (2007). A 'toy' model for operational risk quantification using credibility theory. The Journal of Operational Risk 2(1), 3-19.

Cope, E.W. (2012). Combining scenario analysis with loss data in operational risk quantification. The Journal of Operational Risk, 7(1), 39-56.

Cope, E.W., Mignola, G. & Ugoccioni, R. (2009). Challenges and pitfalls in measuring operational risk from loss data. The Journal of Operational Risk, 4(4), 3-27.

Dahen, H. and Dionne, G. (2010). Scaling models for the severity and frequency of external operational loss data. Journal of Banking & Finance, 34(7), 1484–1496

Degen, M. (2010). The calculation of minimum regulatory capital using single-loss approximations. The Journal of Operational Risk 5(4), 3–17.

Dutta, K.K. and Babbel, D.F. (2013). Scenario Analysis in the Measurement of Operational Risk Capital: A Change of Measure Approach, The Journal of Risk and Insurance, Vol. 81, No. 2, 303–334

Embrechts, P. and Hofert, M. (2011). Practices and issues in operational risk modelling under Basel II, Lithuanian Mathematical Journal, 51(2), 180–193

Embrechts, P., Kluppelberg, C. & Mikosch, T. (1997). Modelling extremal events for Insurance and Finance. Springer.

Embrechts, P. and Puccetti, G. (2006). Aggregating risk capital with an application to operational risk. The Geneva Risk and Insurance Review, 31(2), 71-90..

Embrechts, P. and Puccetti, G. (2008). Aggregating operational risk across matrix structured loss data. The Journal of Operational Risk, 3(2), 29-44.

Fasen, V. and Kluppelberg, C. (2006) Large insurance losses distributions. In: Everitt, B. and Melnick, E. (Eds.) Encyclopedia of Quantitative Risk Assessment, pp. 961-969.Wiley, Chichester.

Figini, S., Giudici, P., Uberti, P., and Sanyal, A. (2008). A statistical method to optimize the combination of internal and external data in operational risk measurement. The Journal of Operational Risk 2(4), 69–78.

Foss, S.; Korshunov, D. and Zachary, S. (2011). An Introduction to Heavy-Tailed and Subexponential Distributions. Springer Series in Operations Research and Financial Engineering. Springer.

Gourier, E.; Farkas, W. & Abbate, D. (2009). Quantification of Operational Risk using Extreme-Value Theory and Copulas: From Theory to Practice. Journal of Operational Risk, 4(3), 3-26.

Gustafsson, J. and Nielsen, J.P. (2008). A mixing model for operational risk. Journal of Opereational Risk, 3(3), 25-37.

Hernández, L., Tejero, J., Suárez, A. & Carillo-Menéndez, S. (2013). Percentiles of sums of heavy-tailed random variables: Beyond the single-loss approximation. Statistics and Computing (February), 1-21

Hernández, L., Tejero, J., Suárez, A. & Carillo-Menéndez, S. (2014). Closed-form approximations for operational value-at-risk. The Journal of Operational Risk, 8(4), 39-54

Kahneman, D., 2011, Thinking Fast and Slow. New York: Farrar Strauss and Giroux.

Kahneman, D., Slovic, P. and Tversky, A. (1982). Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press.

Lambrigger, D. D., Shevchenko, P. V., and Wüthrich, M. V. (2007). The quantification of operational risk using internal data, relevant external data and expert opinion. The Journal of Operational Risk 2(3), 3–27.

McNeil, A.J., Frey R. and Embrechts, P. (2005). Quantitative risk management: Concepts, techniques and tools. Princeton University Press, Oxford, 538p.

Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V. (2006). Infinite mean models and the LDA for operational risk. The Journal of Operational Risk 1(1), 3–25.

ORX. 2008. Dependence and Correlation Analysis. ORX Analytic Agent: IBM Research Report to ORX Analytics Working Group

Panjer, H. (2006). Operational Risk: Modeling Analytics. Wiley, Chichester.

Peters, G.W.; Targino,R.S. & Shevchenko,P.V. (2013). Understanding Operational Risk Capital Approximations: First and Second Orders, preprint submitted to Elsevier, http://arxiv.org/pdf/1303.2910.pdf [Date Accessed: 20 November 2013]

Shevchenko, P. V. (2010). Calculation of aggregate loss distributions. The Journal of Operational Risk 5(2), 3–40.