Construction of forward looking distributions by using limited historical data and scenario assessments

PJ (Riaan) de Jongh and Helgard Raubenheimer, Centre for Business Mathematics and Informatics (BMI), North-West University (NWU).

ABSTRACT

Financial institutions are concerned about various forms of risk that might impact them. The management of these institutions have to demonstrate to shareholders and regulators that they manage these risks in a pro-active way. Often the main risks are caused by excessive claims on insurance policies or losses that occur due to defaults on loan payments or by operations failing. In an attempt to quantify these risks, the estimation of extreme quantiles of loss distributions is of interest. Since financial companies have limited historical data available, they often use scenario assessments by experts to augment the historical data and by providing a forward-looking view. This gives an exposition of a particular statistical approach that may be used to combine historical data and scenario assessments in order to estimate extreme quantiles. In particular, we will illustrate its use by means of practical examples. The statistical method has recently been selected by a major international bank as its preferred model for operational risk capital determination and is currently being implemented in SAS. Based on the experience gained in the process, we include some practical guidelines for the implementation of the method.

INTRODUCTION

Financial losses of any form need to be carefully managed and catered for by financial institutions. For example, the claims made against short-term insurance policies need to be analysed in order to enable insurance companies to determine the reserves needed to meet their obligations and determine the adequacy of their pricing strategies. Similarly, banks are forced by regulatory authorities to set aside regulatory capital to absorb unexpected losses that might occur. Of course the financial institutions are more interested in the aggregate loss or total claim that may occur one year in the future, rather than what the individual losses or claims will be. In particular, note that the interest is focused on the year ahead more so than what has happened in the past. Popular modelling methods involve the construction of annual aggregate loss or claim distributions using the so-called loss distribution approach or random sums method. Such a distribution is assumed to be an adequate reflection of the past but need to be forward looking in the sense that anticipated future losses are taken into account. The constructed distribution may be used to answer questions like 'What aggregate loss level will be exceeded once in c years?' or 'What is the expected annual aggregate loss level?' or 'If we want to guard ourselves against a one in a thousand year aggregate loss, how much capital should we hold next year?' The aggregate loss distribution and its quantiles provide the answers to the above questions and therefore it is paramount that this distribution is modelled and estimated as accurately as possible. Often it is the extreme quantiles of this distribution that is of interest, for instance, the regulator require that when dealing with operational losses, a bank should hold capital that will protect them against a one-in-a-thousand year aggregate loss. To determine the capital the 99.9% Value-at-Risk (VaR) of the distribution has to be calculated. In order to estimate a one-in-a-thousand year loss, one would hope that at least a thousand years of historical data is available, which is not the case. Usually only ten years of data is available and therefore scenario assessments by experts are used to augment the historical data and to provide a forward looking view.

In this paper we provide an exposition of a statistical method (Venter’s approach) that may be used to estimate VaR using historical data in combination with quantile assessments by experts. The proposed approach has been discussed and studied elsewhere (see de Jongh et al. 2015), but specifically in the context of operational risk and economic capital estimation. In this paper we concentrate on the estimation of the VaR of the aggregate loss or claims distribution and strive to make the approach more accessible to a wider audience. Also, based on a SAS implementation done for a major bank, we include some practical guidelines for the implementation and use of the method in practice. In the next section we discuss two approaches (Monte Carlo and Single Loss Approximation) that may be used for the approximation of VaR.
assuming known distributions and parameters. Then, in the third section (Scenario modelling), we will formulate the scenario approach and discuss how scenarios may be created and assessed by experts. In the following section (Estimating VaR), three statistical approaches for estimating VaR are compared and evaluated. In the fifth section (Implementation recommendations) some guidelines on the practical implementation of the preferred approach are given. Some concluding remarks are made in the last section.

**APPROXIMATING VAR**

Let the random variable \( N \) denote the annual number of loss events and assume that \( N \) is distributed according to a Poisson distribution with parameter \( \lambda \), i.e. \( N \sim \text{Poi}(\lambda) \). Note that one could use other frequency distributions like the negative binomial, but the Poisson is the most popular in practice (see e.g. Embrechts et al. 2015). Furthermore assume that the random variables \( X_1, \ldots, X_N \) denote the loss severities of these loss events. Further assume that these loss severities are independently and identically distributed according to a severity distribution \( T \), i.e. \( X_1, \ldots, X_N \sim \text{iid } T \). Then the annual aggregate loss is \( A = \sum_{n=1}^{N} X_n \) and the distribution of \( A \) is a compound Poisson distribution that depends on \( \lambda \) and \( T \) and denoted by \( \text{CoP}(T, \lambda) \). Of course, in practice we do not know \( T \) and \( \lambda \) and have to estimate it. First we have to decide on a model for \( T \), for example a class of distributions \( F(\chi, \theta) \). Then \( \theta \) and \( \lambda \) have to be estimated by using statistical estimates.

The compound Poisson distribution \( \text{CoP}(T, \lambda) \) and its VaR are difficult to calculate analytically so that in practice Monte Carlo (MC) simulation is often used. This is done by generating \( N \) according to the assumed frequency distribution and then by generating \( X_1, \ldots, X_N \) independent and identically distributed according to the true severity distribution \( T \). Then the annual aggregate loss is \( A = \sum_{n=1}^{N} X_n \). The previous process is repeated \( I \) times independently to obtain \( A_i, i = 1, 2, \ldots, I \) and then the 99.9% VaR is approximated by \( A_{(0.999*I+1)} \) 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 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 order to illustrate the Monte Carlo approximation method, we assume that the Burr is the true underlying severity distribution and we use six parameter sets corresponding to an extreme value index (EVI) of 0.33, 0.83, 1.0, 1.33, 1.85 and 2.35 as indicated in Table 1 below. See Beirlant et al. (2004) or refer to Appendix A for a discussion of EVI and the characteristics of this distribution and its properties.

We take the number of repetitions as \( I = 1\,000\,000 \) and repeat the calculation of VaR 1000 times. The 90% band containing the VaR values are shown in Figure 1 below. Here the lower (upper) bound has been determined as the 5% (95%) percentile of the 1000 VaR values, divided by its median, and by subtracting 1. In mathematical terms the 90% band is defined as \( \left[ \frac{\text{VaR}_{(0.951)}}{\text{Median(VaR}_1, \ldots, \text{VaR}_{1000})} - 1, \frac{\text{VaR}_{(0.051)}}{\text{Median(VaR}_1, \ldots, \text{VaR}_{1000})} - 1 \right] \), where \( \text{VaR}_{(k)} \) denotes the \( k \)-th order statistic. From Figure 1 it is clear that the spread, as measured by the 90% band, declines with increasing lambda, but increases with increasing EVI.

**Table 1:** Parameter sets of Burr distribution

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( \alpha )</th>
<th>( \tau )</th>
<th>EVI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>5.00</td>
<td>0.60</td>
<td>0.33</td>
</tr>
<tr>
<td>1.00</td>
<td>2.00</td>
<td>0.60</td>
<td>0.83</td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1.00</td>
<td>1.50</td>
<td>0.50</td>
<td>1.33</td>
</tr>
<tr>
<td>1.00</td>
<td>0.30</td>
<td>1.80</td>
<td>1.85</td>
</tr>
<tr>
<td>1.00</td>
<td>0.17</td>
<td>2.50</td>
<td>2.35</td>
</tr>
</tbody>
</table>
In principle infinitely many repetitions are required to get the exact true VaR. 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). For a detailed comparison of numerical approximation methods, the interested reader is referred to de Jongh et al. (2016). The SLA has become very popular in the financial industry due to its simplicity and can be stated as follows: If $T$ is the true underlying severity distribution function of the individual losses and $\lambda$ the true annual frequency then the $100(1-\gamma)\%$ VaR of the compound loss distribution may be approximated by $T^{-1}(1-\gamma/\lambda)$ or, as modified by Degen (2010) for large $\lambda$, by $T^{-1}(1-\gamma/\lambda) + \lambda\mu$, where $\mu$ is the finite mean of the true underlying severity distribution. The first order approximation by Böcker and Klüppelberg

$$
\text{CoP}^{-1}(1 - \gamma) \approx T^{-1}(1 - \gamma/\lambda)
$$

states that the $100(1 - \gamma)\%$ VaR of the aggregate loss distribution may be approximated by the $100(1 - \gamma/\lambda)\%$ VaR of the severity distribution, if the latter is part of the sub-exponential class of distributions. This follows from a theorem from extreme value theory (EVT) which states that $P(A = \sum_{n=1}^{N} X_n > x)$ as $x \to \infty$ (see e.g. Embrechts et al. 1997). The result is remarkable in that a quantile of the aggregate loss distribution may be approximated by a more extreme quantile (if $\lambda > 1$) of the underlying severity distribution. EVT is all about modelling extremal events and is especially concerned about modelling the tail of a distribution (see e.g. Beirlant et al. 2004), i.e. that part of the distribution we are most interested in. Bearing this in mind we might consider modelling the body and tail of the severity distribution separately as follows.

Let $q$ be a quantile of the severity distribution $T$. We use $q$ as a threshold that splice $T$ in such a way that the interval below $q$ is the expected part and the interval above $q$ the unexpected part of the severity distribution. Define two distribution functions

$$
T_e(x) = T(x)/T(q) \text{ for } x \leq q \text{ and } T_u(x) = [T(x) - T(q)]/[1 - T(q)] \text{ for } x > q,
$$

(2)
i.e. \( T_e(x) \) is the conditional distribution function of a random loss \( X \sim T \) given that \( X \leq q \) and \( T_u(x) \) is the conditional distribution function given that \( X > q \). Note that we then have the identity
\[
T(x) = T(q)T_e(x) + [1 - T(q)]T_u(x) \quad \text{for all } x.
\]
This identity represents \( T(x) \) as a mixture of the two conditional distributions. Instead of modelling \( T(x) \) with a class of distributions \( F(x, \theta) \) we may now consider modelling \( T_e(x) \) with \( F_e(x, \theta) \) and \( T_u(x) \), with \( F_u(x, \theta) \).

From EVT we know that a popular choice for \( F_u(x, \theta) \) is the generalised Pareto distribution (GPD), while a host of choices are available for \( F_e(x, \theta) \), the obvious being the empirical distribution. Note that the Pickands-Balkema-de Haan limit theorem (see e.g. McNeil et al., 2015), 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 the fourth section (Estimating VaR) we will discuss this in more detail.

**SCENARIO MODELLING**

In operational risk management, the Basel Accord (see BCBS 195 and BCBS 196, 2011) suggests the use of scenario assessments to improve severity distribution estimation. BCBS refers to three types of scenarios namely the individual scenario approach, the interval approach and the percentile approach. Unfortunately the description of these methods are not clear, and in the remainder of the paper we discuss a percentile approach suggested by de Jongh et al. (2015) which we believe is the most practical of the existing approaches available in the literature. That being said, it should be noted that probability assessments by experts are notoriously difficult and can be very unreliable as discussed in Kahneman et al. (1982). We mentioned previously that often an extreme quantile of the aggregate loss distribution is of interest. In the case of operational risk, the regulator requires that the one-in-a-thousand year quantile of this distribution be estimated, in other words the aggregate loss level that will be exceeded once in a thousand years. Considering that banks’ have available limited historical data, i.e. maximum 10 years of data, the estimation of such a quantile, using historical data only, is a near impossible task. So modellers have suggested the use of scenarios and experts’ assessments thereof.

As discussed in de Jongh et al. (2015), we advocate the use of the so-called one-in-c year scenario approach. In the one-in-c years scenario approach the experts are asked to answer the question: *What loss level \( q_c \) is expected to be exceeded once every \( c \) years?*. Popular choices for \( c \) vary between 5 and 100 and often 3 values for \( c \) are used. As an example, one bank used \( c = 7, 20 \) 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 \( q_c \) since this loss level has been reached once in 7 years. If the experts judge that the future will be better than the past, they may want to provide a lower assessment for \( q_c \) 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 one-in-hundred years loss level are likely to fall outside many of the experts’ experience. In the banking environment, they may 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. Embrechts and Hofert, 2011). We will not dwell on this issue here and henceforth assume that we do have the one-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 Poisson \((\lambda)\) distributed and the true underlying severity distribution is \( T \), and if the experts are of *oracle quality* in the sense of actually knowing \( \lambda \) and \( T \), then the assessments provided should be
\[ q_c = T^{-1}(1 - \frac{1}{\gamma}). \]  

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 \( N_c \sim \text{Poi}(\lambda c) \) and the conditional distribution of \( M_c \) given \( N_c \) is binomial with parameters \( N_c \) and \( 1 - p_c = P(X \ge q_c) = 1 - T(q_c) \) with \( X \sim T \) and \( p_c = T(q_c) \). Therefore \( EM_c = E[E(M_c|N_c)] = E[N_c(1 - p_c)] = c\lambda(1 - T(q_c)) \). Requiring that \( EM_c = 1 \) yields \( 4 \) and \( p_c = 1 - \frac{1}{\gamma} \).

As illustration of the complexity of the experts’ task, take \( \lambda = 50 \) then \( q_1 = T^{-1}(0.99714) \), \( q_{20} = T^{-1}(0.999) \) and \( q_{100} = T^{-1}(0.9998) \) which implies that the quantities that have to be estimated are very extreme.

Returning to the SLA i.e. \( CaP^{-1}(1 - \gamma) \approx T^{-1}(1 - \gamma/\lambda) \), and by taking \( \gamma = 0.001 \), which implies \( \epsilon = 1000 \), we could ask the oracle the question ‘What loss level \( q_{1000} \) is expected to be exceeded once every 1000 years?’ The oracle will then produce an answer that can be used directly as an approximation for the 99.9% VaR of the aggregate loss distribution. Of course, the experts’ we are dealing with are not of oracle quality.

In the light of the above arguments one has to take in consideration: (a) the SLA gives only an approximation to the VaR we are trying to estimate, and (b) experts are very unlikely to have the experience or the information at their disposal to assess a one-in-thousand year event reliably. One can realistically only expect them to assess events occurring more frequently such as once in 30 years.

Returning to the oracle’s answer in \( 4 \) the expert have to consider both the true severity distribution and the annual frequency when an assessment is provided. In order to simplify the task of the expert, consider the mixed model \( 3 \) discussed in the previous section. This model will assist us in formulating an easier question for the expert to answer. Note that the oracle’s answer to the question in the previous setting can be stated as \( T(q_c) = p_c = 1 - \frac{1}{\gamma} \) (from \( 4 \)) and therefore depends on the annual frequency. However using the definition of \( T_u \) and taking \( q = q_b, b < c \) it follows that \( T_u(q_c) = 1 - \frac{b}{\gamma} \) which does not depend on the annual frequency. The fact that \( q_c = T^{-1}(1 - \frac{1}{\gamma}) = T_u^{-1}(1 - \frac{b}{\gamma}) \) has interesting suggestions about the formulation of the basic question of the one-in-\( c \) years approach. For example, if we take \( b = 1 \) then \( q_1 \) would be the experts’ answer to the question ‘What loss level is expected to be exceeded once annually?’.

Unless we are dealing with only rare loss events, a reasonably accurate assessment of \( q_1 \) should be possible. Then \( T_u(q_c) = 1 - 1/c \) or \( 1 - T_u(q_c) = 1/c \). Keeping in mind the conditional probability meaning of \( T_u \) 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?’.

ESTIMATING VAR

Suppose we have available \( a \) years of historical loss data \( x_1, x_2, \ldots, x_k \) and scenario assessments \( \tilde{q}_7, \tilde{q}_{20} \) and \( \tilde{q}_{100} \) provided by the experts. 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.

In the previous sections two modelling options have been suggested for modelling the true severity distribution \( T \) and a third will follow below. The estimation of the 99.9% VaR of the aggregate loss distribution is of interest and we will consider three approaches to estimate it namely the naïve approach, the GPD approach and Venter’s approach. The naïve approach will make use of historical data only, the GPD approach and Venter’s approach will make use of both historical data and scenario assessments.

Below we demonstrate that, as far as estimating VaR is concerned, Venter’s approach has several advantages when compared to the GPD and naïve approaches.

NAÏVE APPROACH

Assume that we have available only historical data and that we collected the loss severities of a total of \( K \) loss events spread over \( a \) years and denote these observed or historical losses by \( x_1, \ldots, x_k \). Then the annual frequency is estimated by \( \tilde{\lambda} = K/a \). Let \( F(x; \theta) \) denote a suitable family of distributions to model the true loss severity distribution \( T \). The fitted distribution is denoted by \( F(x; \tilde{\theta}) \), with \( \tilde{\theta} \) denoting the (maximum
likelihood) estimate of the parameter(s) \( \theta \). In order to estimate VaR a small adjustment of the Monte Carlo approximation approach, discussed earlier, is necessary.

**Naïve VaR estimation algorithm**

Generate \( N \) from the Poisson distribution with parameter \( \lambda \) and \( X_1, \ldots, X_N \sim iid \ F(x; \hat{\theta}) \) and calculate \( A = \sum_{i=1}^{N} X_i \). Then repeat the process \( I \) times independently to obtain \( A_i, i = 1, 2, \ldots, I \). Then the 99.9% VaR is estimated by \( A_{\left[0.999 I + 1\right]} \) where \( A_{(i)} \) denotes the \( i \)-th order statistic and \( [k] \) the largest integer contained in \( k \).

**Remarks**

The estimation of VaR using the above-mentioned naïve approach has been discussed in several books and papers (see e.g. McNeil at al. 2015). Cope et al. (2009) stated that heavy-tailed data sets are hard to model and require much caution when interpreting the resulting VaR estimates. 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 losses 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. Two different severity distributions for modelling the individual losses may both fit the data well in terms of goodness-of-fit statistics, yet may provide capital estimates which may differ by billions. Embrechts and Hofert (2011) have highlighted several deficiencies and expressed areas of concern about elements of the naïve estimation approach, in particular, the estimation of the severity distribution and the subsequent estimation of an extreme VaR of the aggregate loss distribution.

In Figure 2 below we used the naïve approach to illustrate the effect of some of the above-mentioned claims. In Figure 2 (a) we assumed a Burr distribution, i.e. \( T_{\text{Burr}}(1, 0.6, 2) \), as our true underlying severity distribution. In the top panel we show the distribution function and in the middle the log of 1 minus the distribution function. This gives us a better view of the tail behavior of the distribution. Then in the bottom panel the Monte Carlo results of the VaR approximations are given by means of a box plot using the 5% and 95% percentiles for the box. As before, one million simulations were used to approximate VaR and the VaR calculations were repeated a 1000 times. The top panel in Figure 2(b) depicts a sample of 100 losses from the \( T_{\text{Burr}}(1, 0.6, 2) \) distribution (we assumed \( \lambda = 10, \alpha = 10 \)). In the middle panel of Figure 2 (b) the fitted distribution and the maximum likelihood estimates of the parameters are given as \( F_{\text{Burr}}(1.07, 0.56, 2.2) \). In the bottom panel the results of the VaR estimates using the naïve approach is provided. Note how the distribution of the VaR estimates differ from those obtained using the true underlying severity distribution. Of course, sampling error is present and the generation of another sample will result in a different box plot. Let us illustrate this by studying the effect of extreme observations. In order to do this we moved the maximum value further into the tail of the distribution and repeat the fitting process. The data set is depicted in the top panel of Figure 2(c) and the fitted distribution in the middle as \( F_{\text{Burr}}(1.01, 0.52, 2.26) \). Again the resulting VaR estimates are shown in the bottom panel. In this case the introduction of the extreme loss has a profound boosting effect on the resulting VaR estimates.

In practice, and due to imprecise loss definitions, risk managers may incorrectly group two losses into one extreme loss that has a profound boosting effect on VaR estimates. In the light of this, it is important that the manager is made aware of the process generating the data and the importance of clear definitions in the process of recording loss events.
Figure 2: Illustration of the effects of VaR estimation using the naïve approach.

THE GPD APPROACH

This modelling approach is based on the mixed model formulation (3). As before, we have available $a$ years of historical loss data $x_1, x_2, ..., x_K$ and scenario assessments $\tilde{q}_7, \tilde{q}_{20}$ and $\tilde{q}_{100}$. Then the annual frequency $\lambda$ can again be estimated as $\hat{\lambda} = K/a$. Next the threshold $q$ must be specified, which we take as $b$ the smallest of the scenario $c$-year multiples, i.e. $q = q_b$. We then estimate $q_b$ as the corresponding smallest of the scenario assessments $\tilde{q}_b$ provided by the experts, in this case $\tilde{q}_7$. $T_n(x)$ can be estimated by fitting a parametric family $F_n(x, \theta)$ (such as the Burr) to the data $x_1, x_2, ..., x_K$ 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 $F_n(x)$. 


Next, \( F_u(x) \) can be modelled by the GPD distribution. See Appendix A for the characteristics of this distribution. For ease of explanation, suppose we have actual scenario assessments \( q_7, q_{20} \) and \( q_{100} \) and thus take \( b = 7 \) and estimate \( q_b \) by \( \hat{q}_7 \). Substituting these scenario assessments into \( F_u(q_c) = 1 - \frac{b}{c} \); with \( b = 7 \), \( c = 20 \), 100 yields two equations

\[
F_u(q_{20}) = GPD(\hat{q}_{20}; \sigma, \xi, \hat{q}_7) = 0.65 \quad \text{and} \quad F_u(q_{100}) = GPD(\hat{q}_{100}; \sigma, \xi, \hat{q}_7) = 0.93
\]

(5)

that can be solved to obtain estimates \( \bar{\sigma} \) and \( \bar{\xi} \) of the parameters \( \sigma \) and \( \xi \) in the GPD that are based on the scenario assessments. Some algebra shows that a solution exists only if \( \frac{\hat{q}_{100} - \hat{q}_7}{\hat{q}_{20} - \hat{q}_7} > 2.533 \). This fact should be borne in mind when the experts do their 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 \( \sum_i [GPD(\hat{q}_i; \sigma, \xi, \hat{q}_7) - (1 - b/c)] \). Other possibilities include a weighted version of the sum of deviations in this expression or deviation measures comparing the GPD quantiles directly to the scenario assessments. Some algebra shows that a solution exists only if \( \frac{\hat{q}_{100} - \hat{q}_7}{\hat{q}_{20} - \hat{q}_7} > 2.533 \). This fact should be borne in mind when the experts do their assessments.

Returning now to practical use of equation (6), the algorithm below summarizes the integration of the GPD VaR estimation algorithm

1) Generate \( N_e \sim Pois \left( \hat{\lambda} - \frac{1}{2} \right) \) and \( N_u \sim Pois \left( \frac{1}{2} \right) \);

2) Generate \( X_1, \ldots, X_{N_e} \sim i.d. \hat{F}_e \) and \( X_{N_e+1}, \ldots, X_{N_e+N_u} \sim i.d. \hat{F}_u \) and calculate \( A = \sum_{i=1}^{N_u} X_n \) 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 \( \hat{F} \).

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.

**Remarks**

When using the GPD one-in-\( c \) years integration approach to model the severity distribution, we realised that the 99.9% VaR of the aggregate distribution is almost exclusively determined by the scenario assessments and their reliability greatly affects the reliability of the VaR estimate. The SLA supports this conclusion. As noted above the SLA implies that we need to estimate \( q_{1000} = T^{-1} \left( 1 - \frac{1}{10000} \right) \) and its estimate would be \( \hat{q}_{1000} = GPD^{-1} \left( \frac{1}{10000}, \sigma, \xi, \hat{q}_b \right) \). Therefore 99.9% VaR largely depends on the GPD fitted with the scenario assessments. In Figure 3 below we depict the VaR estimation results by fitting \( \hat{F}_v \) assuming a Burr distribution and \( \hat{F}_u \) assuming a GPD. The top panel in Figure 3 (a) depicts the tail behavior of the true severity distribution which is assumed to be a Burr and denoted as \( T_{Burr}(1,0.6,2) \). Using the VaR approximation technique discussed in the second section (Approximating VaR) and assuming \( \lambda = 10 \), \( I = 1\ 000\ 000 \) and 1 000 repetitions, the VaR approximations are depicted in the bottom panel in the form of a box plot as before. Assuming that we were supplied with quantile assessments by the oracle we use the two samples discussed in Figure 2 and apply the GDP approach. The results are displayed in Figure 3 (b) and (c) below.
The GPD fit to the oracle quantiles produce similar box plots, which in turn is very similar to the box plot of the VaR approximations. Clearly the fitted Burr has little effect on the VaR estimates. The VaR estimates obtained through the GPD approach is clearly dominated by the oracle quantiles. Of course, if the assessments are supplied by experts and not oracles the results would differ significantly. This is illustrated when we compare the GPD with Venter’s approach.

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 Venter’s estimation method that will meet these aims.

**VENTER’S APPROACH**

A retired colleague, Hennie Venter suggested that, given the quantiles \(q_7, q_{20}, q_{100}\) one may write the distribution function \(T\) as the following identity:

\[
T(x) = p_7 T_e(x) + [p_{20} - p_7] T_{u1}(x) + [p_{100} - p_{20}] T_{u2}(x) + [1 - p_{100}] T_{u3}(x) \quad \text{for all } x, \tag{7}
\]

where

\[
T_e(x) = T(x) / T(q_7) \quad \text{for } x \leq q_7,
\]

\[
T_{u1}(x) = [T(x) - T(q_7)] / [T(q_{20}) - T(q_7)] \quad \text{for } q_7 < x \leq q_{20},
\]

\[
T_{u2}(x) = [T(x) - T(q_{20})] / [T(q_{100}) - T(q_{20})] \quad \text{for } q_{20} < x \leq q_{100}, \text{ and}
\]

\[
T_{u3}(x) = [T(x) - T(q_{100})] / [1 - T(q_{100})] \quad \text{for } q_{20} < x \leq q_{100}.
\]
As was the case in (3), this is clearly a mixed distribution where \( T_e(x) \) is the conditional distribution function of a random loss \( X \sim T \) given that \( X \leq q \), \( T_{u_i}(x) \) the conditional distribution function given that \( q < X \leq q_{20} \), \( T_{u_2}(x) \) the conditional distribution function given that \( q_{20} < X \leq q_{100} \), and \( T_{u_3}(x) \) is the conditional distribution function given that \( X > q_{100} \). Equivalently, we can write (7) as follows:

\[
T(x) = \begin{cases} 
R(7)T(x) & \text{for } x \leq q_7 \\
p_7 + R(7,20)[T(x) - T(q_7)] & \text{for } q_7 < x \leq q_{20} \\
p_{20} + R(20,100)[T(x) - T(q_{20})] & \text{for } q_{20} < x \leq q_{100} \\
p_{100} + R(100)[T(x) - T(q_{100})] & \text{for } q_{100} < x < \infty.
\end{cases}
\]

where

\[
R(7) = p_7/T(q_7), \quad R(7,20) = [p_{20} - p_7]/[T(q_{20}) - T(q_7)],
\]

\[
R(20,100) = [p_{100} - p_{20}]/[T(q_{100}) - T(q_{20})], \quad \text{and } R(100) = [1 - p_{100}]/[1 - T(q_{100})].
\]

Again \( T(q_c) = p_c = 1 - \frac{1}{\alpha} \) and it should be clear that the expressions on the right reduces to \( T(x) \) and all the \( R \) ratios are equal to 1. Also, the definition of \( T(x) \) could easily be extended for more quantiles. Given the previous discussion we can model \( T(x) \) by \( F(x, \theta) \) and estimate it by \( F(\hat{x}, \hat{\theta}) \) using the historical data and maximum likelihood, and estimate the annual frequency by \( \hat{\lambda} = K/\alpha \). Given scenario assessments \( q_7, q_{20} \) and \( q_{100} \), then \( T(q_c) \) can be estimated by \( F(\hat{q}_c, \hat{\theta}) \) and \( p_c \) by \( \hat{p}_c = 1 - \frac{1}{\alpha} \). The estimated \( \hat{R} \) ratios are then

\[
\tilde{R}(7) = \frac{\hat{p}_7}{F(\hat{q}_7; \hat{\theta})}, \quad \tilde{R}(7,20) = \frac{\hat{p}_{20} - \hat{p}_7}{F(\hat{q}_{20}; \hat{\theta}) - F(\hat{q}_7; \hat{\theta})},
\]

\[
\tilde{R}(20,100) = \frac{\hat{p}_{100} - \hat{p}_{20}}{F(\hat{q}_{100}; \hat{\theta}) - F(\hat{q}_{20}; \hat{\theta})} \quad \text{and } \tilde{R}(100) = \frac{1 - \hat{p}_{100}}{1 - F(\hat{q}_{100}; \hat{\theta})}.
\]

Notice that if our estimates were actually exactly equal to what they are estimating, these ratios would all be equal to 1. With the formulation in (8) the true severity distribution function \( T \) may now be estimated by \( \tilde{H} \) as follows (see de Jongh et al. 2015):

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

Notice again that this estimate is consistent in the sense that it actually reduces to \( T \) if all estimators are exactly equal to what they are estimating.

Also note that \( \tilde{H}(\hat{q}_7) = \tilde{p}_7, \tilde{H}(\hat{q}_{20}) = \tilde{p}_{20} \) and \( \tilde{H}(\hat{q}_{100}) = \tilde{p}_{100} \), i.e. the equivalents of \( T(q_c) = p_c \) 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. Thus this new estimated severity distribution estimate \( \tilde{H} \) ‘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 \( \tilde{R}(7), \tilde{R}(7,20), \tilde{R}(20,100) \) and \( \tilde{R}(100) \) in (9) 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 estimate VaR using this method are as follows:

**Venter’s VaR estimation algorithm**

1) Generate \( N \sim Pois(\hat{\lambda}) \);

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

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

The SLA again sheds some light on this method. As noted above the SLA implies that we need to estimate
\[ q_{1000} = T^{-1} \left( 1 - \frac{1}{1000 \lambda} \right) \] and its estimate would be 
\[ \hat{q}_{1000} = H^{-1} \left( 1 - \frac{1}{1000 \lambda} \right) \] (see Figure 1). Some algebra shows that the equation 
\[ F(\hat{q}_{1000}; \theta) = \frac{H^{-1}(\hat{p}_{1000})}{H(100)} \] needs to be solved for \( \hat{q}_{1000} \). Depending on the choice of the family of distributions \( F(x; \theta) \), 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 Venter’s approach.

In order to illustrate the properties of this approach we assume that the true underlying severity distribution is the Burr(1.0, 0.6, 2) as before. We then construct a ‘false’ severity distribution as the fitted distribution to the distorted sample depicted in Figure 2(c), i.e. the Burr(1.00,0.52,2.26). We refer to the true severity distribution as Burr_1 and the false one as Burr_2. In Figure 4(a) the box plots of the VaR approximations of the two distributions are given (using the same input for the MC simulations). We then illustrate the performance of the GPD and Venter approach in two cases. The first case assumes that the correct (oracle) quantiles of Burr_1 are supplied, but that the loss data are distributed according to the false distribution Burr_2. In the second case the quantiles of the false severity distribution is supplied, but the loss data follows the true severity distribution. The box plots of the VaR estimates are given in Figure 4(b) for case 1 and Figure 4(c) for case 2.

![Box plots of VaR approximations](image)

**Figure 4:** Comparison of VaR results for the GPD and Venter approaches.

The behaviour of the GPD approach is as expected and the box plots corresponds to the quantiles supplied. Clearly the quantiles and not the loss data dictates the results. On the other hand the Venter approach is affected by both the loss data and quantiles supplied. In the example studied here it seems as if the method is more affected by the quantiles than by the data. However, the more loss data are supplied, the more \( H(x) \) will be influenced by \( F(x; \hat{\theta}) \), the distributed fitted to the historical data.

**GPD AND VENTER MODEL COMPARISON**

In this section we conduct a simulation study to investigate the effect on the two approaches by perturbing the quantiles of the true underlying severity distributions. We assume the six parameters sets of Table 1 as the true underlying severity distributions and then perturb the quantiles in the following way. For each simulation run, choose three perturbation factors \( u_7, u_20, u_{100} \) independently and uniformly distributed over the interval \([1 - \epsilon, 1 + \epsilon]\) and then tentatively take 
\[ \tilde{q}_7 = u_7 q_7, \tilde{q}_{20} = u_{20} q_{20} \text{ and } \tilde{q}_{100} = u_{100} q_{100} \] 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 \( \epsilon \) 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 experts 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 \( \epsilon \) 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
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 \([1 - \epsilon, 1]\).

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 \(\epsilon\) the following steps are followed:

1) Use the VaR approximation algorithm in the second section to determine the 99.9\% VaR for the Burr Type XII with the current choice of parameters. Note that the value obtained here approximately equals the true 99.9\% VaR. We refer to this value as the approximately true (AT) VaR.

2) Generate a data set of historical losses, i.e. generate \(K \sim Poi(7\lambda)\) and then generate \(x_1, x_2, ..., x_K \sim iid\) Burr Type XII with the current choice of parameters. 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.

3) Add to the historical losses three scenarios \(\tilde{q}_7, \tilde{q}_{20}, \tilde{q}_{100}\) generated by the quantile perturbation scheme explained above. Estimate the 99.9\% VaR using the GPD approach.

4) Using the historical losses and the three scenarios of item c), calculate the severity distribution estimate \(\hat{H}\) and apply Venter’s approach to estimate the 99.9\% VaR.

5) Repeat items 1)-4) 1000 times and then summarise and compare the resulting VaR estimates.

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 we construct 90\% VaR bands as before for the 1000 repeated GPD and Venter VaR estimates, i.e. \(\frac{\text{VaR}_{(51)}}{\text{MedAT}} - 1, \frac{\text{VaR}_{(951)}}{\text{MedAT}} - 1\). The results are given in Figure 5 below. Note that light grey represents the GPD band and dark grey the Venter band, while the overlap between the two bands are even darker.
From Figure 5 we make the following observations:

For small frequencies ($\lambda \leq 10$) the GPD approach outperforms the Venter approach, except for short tailed severity distributions and higher quantile perturbations. When the annual frequency is high ($\lambda \geq 50$) and for moderate to high quantile perturbations ($\epsilon \geq 0.2$) the Venter approach is superior, and more so for higher $\lambda$ and $\epsilon$. Even for small quantile perturbations ($\epsilon = 0.1$) and high annual frequencies ($\lambda \geq 50$) the Venter approach performs reasonable when compared to the GPD.

The above information suggest that provided enough loss data is available the Venter approach is the best choice to work.

IMPLEMENTATION RECOMMENDATIONS

As stated in the introduction, a major international bank approached us to implement Venter’s method. Although we are still in the process of doing so we have gained some experience and wish to share the following implementation guidelines:

i. Study the loss data carefully with respect to the procedures used to collect the data. Focus should be on the largest losses and one has to establish whether these losses were recorded and classified correctly according to the definitions used.
ii. Test the underlying assumptions, for example ask questions such as ‘Are the losses independent and identically distributed?’ and ‘Is the number of losses Poisson distributed?’ It might be necessary to correct the recorded loss data for inflation and growth of business.

iii. Experts should be presented with an estimate of \( q_1 \) (based on the loss data) and then should answer the question ‘Among those losses that are larger than \( q_1 \) what level is expected to be exceeded only once in 7 years?’ Once the expert has produced an answer \( q_\text{7} \), this answer should be taken into account when the assessment \( q_{20} \) is made. Similarly when \( q_{100} \) is assessed, \( q_{20} \) should be considered.

iv. The assessments by the expert should be checked with the condition \( \frac{q_{100} - q_7}{q_{20} - q_7} > 2.533 \). In our implementation project we seldom selected the GPD as a model for \( F(x; \theta) \). However, in the case when no historical data was available we used the GPD to model the scenario assessments. Despite the fact that the GDP was not used often, this condition bring realism for the experts as far as the ratios between the assessments are concerned. So the expert should use this as a decision making tool in step iii) above.

v. The loss data may be fitted by a wide class of severity distributions \( F(x; \theta) \). We used SAS PROC SEVERITY in order to identify the 5 best fitting distributions according to a number of statistical goodness-of-fit measures.

vi. Using (9), determine the ratios \( R(7), R(7,20), R(20,100) \) and \( R(100) \) of the 5 best fitting distributions and then select the best distribution using the ratios. Of course, one would prefer that the best fitting distribution corresponds with \( R \) ratios being close to one.

vii. Then, present the ratios of the best fitting distribution, that deviate significantly from one, to the experts for possible re-assessment. If the expert change any of the assessments, repeat steps iii to vi at least once, but preferably twice.

viii. Different data sources may be available to augment the historical data or expert assessments. For example, in operational risk external loss data is available in various pooled resources, such as the proprietary ORX database (see Operational Riskdata eXchange Association 2008) and SAS OpRisk Global Data, a repository of publicly reported operational losses. Integrating the external data with internal data is not straightforward (see e.g. Bolancè et al. 2013). However, the external data might provide added insight to the experts when they do their assessments.

CONCLUDING REMARKS

In this paper we motivated Venter’s approach as a practical and sound methodology to estimate VaR when both historical data and scenario assessments are available. 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. This methodology is currently being implemented by a major international bank and we included guidelines for its practical implementation. As far as future research is concerned, we are currently investigating the effectiveness of the ratios in assisting the experts with their assessments. Also, we are testing the effect of replacing \( q_{100} \) with a lower quantile (e.g. \( q_{50} \)) in the assessment process.

REFERENCES


ACKNOWLEDGEMENTS

We 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.

APPENDIX A

THE GENERALISED PARETO DISTRIBUTION (GPD)

The generalized Pareto distribution is defined as

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

where \( x \geq q_b \), \( q_b \) the so-called EVT threshold and \( \sigma \) and \( \xi \) respectively scale and shape parameters. The GPD is generalized in the sense that it contains a number of special cases, e.g. the ordinary Pareto and the exponential distributions. The role of the GPD in EVT is as a natural model for the excess distribution over a high threshold. The Extreme Value Index (EVI) characterizes the tail behavior of distributions. Note that heavy-tailed distributions have a positive EVI and larger EVI implies heavier tails. The EVI of the GPD distribution is \( \xi \). This follows from the fact that for positive EVI the GPD distribution belongs to the Pareto-type class of distributions, having a distribution function of the form \( 1 - F(x) = x^{-\xi} \ell_F(x) \), 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 GPD
distribution is regularly varying with index $-1/\xi$ and therefore belongs to the class of sub-exponential distributions. Note that the $\gamma$-th quantile of the GPD is $q(\gamma) = GPD^{-1}(\gamma, \sigma, \xi, q_b) = \left(q_b + \sigma \frac{\gamma - \xi - 1}{\xi}\right)$ when $\xi \neq 0$ and $GPD^{-1}(\gamma, \sigma, \xi, q_b) = q_b - \sigma \ln(1 - \gamma)$ when $\xi = 0$.

THE BURR DISTRIBUTION

The three parameter Burr type XII distribution function is

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

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 EVI of the Burr distribution is given by $EVI = \zeta = 1/\tau \alpha$. This follows 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. The $\gamma$-th quantile of the Burr distribution is $q(\gamma) = B^{-1}(\gamma; \eta, \tau, \alpha) = \eta \left((1 - \gamma)^{-1/\alpha} - 1\right)^{1/\tau}$.

CONTACT INFORMATION

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

Riaan de Jongh
Centre for Business Mathematics and Informatics (BMI), North-West University (NWU)
riaan.dejongh@nwu.ac.za
http://natural-sciences.nwu.ac.za/bmi