A time scaling rule for operational VAR approximation in particular to the Pareto severity model, for which they also obtain a simple closed-form approximation for operational VAR can be obtained. They apply this approximation in particular to the Pareto severity model, for which they also obtain a simple time scaling rule for operational VAR.

According to the final proposals of the Basel Committee on Banking Supervision (2004), operational risk will be a determinant of the new capital requirements as of 2007, along with market and credit risk. Every bank will have to calculate explicit capital charges to cover operational risk by means of one of three approaches: the basic indicator approach, the standardised approach or the advanced measurement approach (AMA). This ‘continuum of approaches’ reflects different levels of sophistication and risk sensitivity. AMA, as the most flexible approach for operational risk quantification, allows a bank to build its own internal operational risk model and measurement system, comparable with market risk standards. Instead of prescribing a particular type of value-at-risk model, the committee requires a set of quantitative and qualitative standards to be fulfilled. The following two are most relevant for the issues discussed in this paper:

- The operational risk measure is VAR at a confidence level of 99.9% with a holding period of one year.
- The measurement approach must capture potentially severe tail loss events.

The most popular method in the industry to satisfy the AMA standards is the loss distribution approach (LDA), which is based on modelling the probability distribution of operational losses using banks’ internal and external data.

Despite the current lively debate, operational risk is not a new phenomenon. We all recall the operational risk event that happened on February 26, 1995, when the prestigious UK merchant bank Barings had to declare bankruptcy. The reason was an accumulated loss of £625 million in its Singapore division caused by a trader, who was hiding loss-making positions in financial derivatives.

Obviously, an important objective of any financial institution is to manage risk. The only feasible way to manage risk is by identifying and minimising it. This can only be done successfully by the development of adequate quantification techniques. It seems to be worthwhile to consider actuaries’ approach to similar problems. Dealing with randomly occurring insurance claims, they have been at the very core of issues similar to operational risk for more than 100 years. Hence, it is not surprising that LDA models have their roots in insurance risk theory, which goes back to the early work by Filip Lundberg in 1903. In this respect, modelling and quantifying operational risk can be viewed as a 100-year-old science!

In this article, we suggest and investigate a model that indeed originates in insurance. We first introduce a standard LDA, which contains the compound Poisson and the negative binomial models as special cases. Exploiting the common wisdom that severity distributions for operational risk are typically very heavy-tailed, we derive a closed-form approximation for operational VAR. Extending the model from a static model to a dynamic one, we further show that such models have an in-built α-root-of-time rule for some $\alpha > 0$, which usually differs from the well-known square-root law. Finally, we introduce a new simple operational VAR estimate, which can be applied for scenario generation or for expert-based risk assessment.

The loss distribution approach

In the context of LDA models, a widely accepted procedure consists of splitting up the total loss amount over a certain period into a frequency component, that is, the number of losses, and a severity component, that is, the individual loss amounts. The total loss is then obtained by compounding the frequency and the severity information. A prototypical model of this kind, which is currently best practice and implemented in various commercial software packages, is the following.

**Definition (standard LDA).**
- The severity process. The severities $X_k, k \in \mathbb{N}$, are positive independent and identically distributed (IID) random variables describing the magnitude of each loss event.
- The frequency process. The number $N(t)$ of loss events in the time interval $[0, t]$ for $t \geq 0$ is random. The resulting counting process $(N(t))_{t \geq 0}$ is generated by a sequence of points $(T_k)_{k \geq 1}$ of non-negative random variables satisfying:
  \[ 0 \leq T_1 \leq T_2 \leq \ldots \text{ a.s.} \]
  and:
  \[ N(t) = \sup \{ n \geq 1 : T_n \leq t \}, \quad t \geq 0 \]
- The severity process and the frequency process are assumed to be independent.
- The aggregate loss process. The aggregate loss $S(t)$ up to time $t$ constitutes a process:
  \[ S(t) = \sum_{i=1}^{N(t)} X_i, \quad t \geq 0 \]

Note that we do not require $X_i$ to have finite mean and/or variance. This is in accordance with empirical research: Moscadelli (2004) showed convincingly that typical severity distributions for operational risk are very heavy-tailed such that even moments of low order may not exist.

Typical examples for LDA models are obtained by specifying the frequency process in the following way. The Poisson LDA is a standard LDA, where $(N(t))_{t \geq 0}$ is a homogeneous Poisson process with intensity $\lambda > 0$, in particular:

\[ P(N(t) = n) = p_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n \in \mathbb{N}_0 \]

The negative binomial LDA is a standard LDA, where $(N(t))_{t \geq 0}$ is given by a negative binomial process satisfying for $\beta, \gamma > 0$.
\[ P(N(t) = n) = p_n(t) = \binom{\gamma + n - 1}{n} \left( \frac{\beta}{\beta + t} \right)^n \left( \frac{t}{\beta + t} \right)^{\gamma}, \quad n \in \mathbb{N}_0 \]

The negative binomial distribution is a gamma mixture of a Poisson distribution, that is, it can be viewed as a Poisson distribution whose parameter \( \lambda \) is a gamma distributed random variable. This allows for modelling over-dispersion, which means that for all \( t > 0 \) the variance of \( N(t) \) is greater than its mean, whereas for the Poisson LDA \( \text{var}(N(t)) = EN(t) \) holds. However, we will see later that as far as the operational VAR approximation is concerned, over-dispersion is of minor importance.

- **Subexponential severity distributions.** Concerning the severity, we have to take – in accordance with Basel II – the heavy-tail property of operational losses into account. Some popular distributions are given in table A. All of them are heavy-tailed or, more precisely, they belong to the class of so-called subexponential distributions, meaning that their tails decay slower than any exponential tail.

> The defining property of subexponential distributions is that the tail of the sum of \( n \) IID severities is most likely to be large because of one of the terms being large, or, with emphasis on operational risk, severe overall losses are mainly due to a single big loss rather than the consequence of accumulated small independent losses. Of course, this insight should have implications for operational risk management.

The goal of every LDA model is to determine the aggregate loss distribution, which for the standard LDA can be written as:

\[ G_t(x) = P(S(t) \leq x) = \sum_{n=0}^{\infty} p_n(x)P(S(t) \leq x|N(t) = n) = \sum_{n=0}^{\infty} p_n(x)F^*t(x), \quad x \geq 0, \quad t \geq 0 \]

where \( F(t) = P(X_i \leq t) \) is the distribution function of \( X_i \), and \( F^*t(x) = P[\sum_{i=1}^{k} X_i \leq x] \) is the \( n \)-fold convolution of \( F \) with \( F^* = F \) and \( F^k = I_{[0,\infty)} \).

> For most choices of severity and frequency distributions, \( G_t \) cannot be calculated analytically. Approximation methods to overcome this problem include Panjer recursion, Monte Carlo simulation and fast Fourier transform methods (see Klugman, Panjer & Willmot, 2004, for an overview). The drawback of these methods is, however, that their results remain a 'black box', and the interaction between different model parameters and their impact on the final result, that is, the operational VAR, is only interpretable through extensive sensitivity analyses.

As both regulatory capital and economic capital are based on a very high quantile of the aggregate loss distribution \( G_o \), a natural estimation method for operational VAR is via asymptotic tail and quantile estimation. Instead of considering the entire distribution, it is sufficient to concentrate on the right tail \( P(S(t) > x) \) for very large \( x \). Now, in actuarial science, the tail behaviour of \( G_o \) has been extensively studied both for small claims and large claims models. For the latter, a key result states that for a standard LDA with subexponential severities one has under weak regularity conditions (see the theorem in the appendix, equation (7)) for every fixed \( t > 0 \):

\[ G_o(x) = \text{EN}(t)\bar{F}(x), \quad x \rightarrow \infty \tag{2} \]

where \( \text{EN}(t) \) is the expected frequency and \( \bar{F}(x) = 1 - F(x) \) and \( G_o(x) = 1 - G_{oo}(x) \) are the tail distributions of severity and aggregate loss, respectively. The symbol \( \sim \) means that the quotient of the right-hand and left-hand side tends to one, that is, \( \lim_{x \rightarrow \infty} G_o(x) = EN(t) \) for every fixed \( t > 0 \).

It has been shown in examples 1.3.10 and 1.3.11 of Embrechts, Klüppelberg & Mikosch (1997) that the tail estimate (2) holds for the Poisson LDA and the negative binomial LDA.

- **A closed-form approximation for operational VAR.** Given relation (2), it is straightforward to obtain an expression for the operational VAR, valid at very high confidence levels. Recall that VAR is just a quantile of a distribution function.

**Definition (value-at-risk).** Suppose \( G_t \) is the aggregate loss distribution. Then the VAR up to time \( t \) at confidence level \( \kappa \) is defined as the \( \kappa \)-quantile of the loss distribution:

\[ \text{VAR}_t(\kappa) = G_t^{-1}(\kappa), \quad \kappa \in (0,1) \]

where \( G_t^{-1}(\kappa) = \inf\{t \in \mathbb{R} : G_t(x) \geq \kappa\}, \quad 0 < \kappa < 1, \) is the generalised inverse of \( G_t \). If \( G_t \) is strictly increasing and continuous, we may write \( \text{VAR}_t(\kappa) = G_t^{-1}(\kappa) \).

Using (2) we obtain an asymptotic formula for the operational VAR:

**Theorem (analytical operational VAR).** Consider the standard LDA model for fixed \( t > 0 \) and a subexponential severity with distribution function \( F \). Assume, moreover, that the tail estimate (2) holds. Then, the \( \text{VAR}_t(\kappa) \) satisfies the approximation:

\[ \text{VAR}_t(\kappa) = F^+ \left( 1 - \frac{1 - \kappa}{\text{EN}(t)} (1 + o(1)) \right), \quad \kappa \rightarrow 1 \tag{3} \]

**Proof.** Note first that \( \kappa \rightarrow 1 \) is equivalent to \( x \rightarrow \infty \). Then recall that \( o(1) \) always stands for a function, which tends to zero, if its argument tends to a boundary, in our case if \( \kappa \rightarrow 1 \) or \( x \rightarrow \infty \). With this notation, relation (2) can be rewritten as:

\[ G_o(x) = \text{EN}(t)\bar{F}(x)(1 + o(1)), \quad x \rightarrow \infty \]

Setting the right-hand side equal to \( \kappa \) gives an asymptotic solution:

\[ F(x) = 1 - \frac{1 - \kappa}{\text{EN}(t)} (1 + o(1)), \quad x \rightarrow \infty \]

and, finally:

\[ x = G_o^{-1}(\kappa) = F^+ \left( 1 - \frac{1 - \kappa}{\text{EN}(t)} (1 + o(1)) \right), \quad \kappa \rightarrow 1 \]

This result, which holds for a quite general class of LDA models, is remarkable for two reasons. First, it says that the operational VAR at high confidence levels only depends on the tail and not on the body of the severity distribution. Therefore, if one is only interested in VAR calculations, modelling the whole distribution function \( F \) is superfluous. Second, because the frequency enters in expression (3) only with its expectation \( \text{EN}(t) \), it is also not necessary to calibrate a specific counting process; estimating the sample mean of the frequency suffices. As a consequence, over-dispersion as modelled by the negative binomial distribution has asymptotically no impact on the operational VAR.

To obtain a first-order approximation for the operational VAR for a specific LDA model, it suffices to combine (3) with the tail of the (subexponential) severity distribution \( F \). Furthermore, even closed-form solutions

---

### A. Popular severity distributions with support \((0, \infty)\)

<table>
<thead>
<tr>
<th>Name</th>
<th>Distribution function</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>( F(x) = \Phi \left( \frac{\ln x - \mu}{\sigma} \right) )</td>
<td>( \mu \in \mathbb{R}, \sigma &gt; 0 )</td>
</tr>
<tr>
<td>Weibull</td>
<td>( F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)^\alpha} )</td>
<td>( 0 &lt; 0,0 &lt; \tau &lt; 1 )</td>
</tr>
<tr>
<td>Pareto</td>
<td>( F(x) = 1 - \left( 1 + \frac{x}{\theta} \right)^{-\alpha} )</td>
<td>( \alpha, \theta &gt; 0 )</td>
</tr>
</tbody>
</table>

Note: \( \Phi \) is the standard normal distribution function.
Cutting edge

B. First-order approximations of the $\text{VAR}(\kappa)$ as $\kappa \to 1$ for the aggregate loss distribution for popular severity distributions

<table>
<thead>
<tr>
<th>Name</th>
<th>$\text{VAR}(\kappa)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>$\exp \left[ \mu - \sigma \Phi^{-1} \left( \frac{1 - \kappa}{EN(t)} \right) \right]$</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\theta \ln \left( \frac{EN(t)}{1 - \kappa} \right)^{1/\alpha}$</td>
</tr>
<tr>
<td>Pareto</td>
<td>$\theta \left( \frac{EN(t)^{1/\alpha}}{1 - \kappa} - 1 \right)$</td>
</tr>
</tbody>
</table>

Note: set $EN(t) = \lambda t$ for a Poisson distributed and $EN(t) = \gamma t / \beta$ for a negative binomially distributed frequency

1. Comparison of the approximated VAR given by (4) (dashed line) and the simulated VAR (solid line) for the Pareto-Poisson LDA with $\theta = 1$

for the (asymptotic) operational VAR are available (see table B).

Finally, we want to emphasise that the problem of finding a severity distribution that accurately describes empirical loss data is a non-trivial task, and that the parameterisation of appropriate severity and frequency distributions is an integral part of every AMA model. A textbook treatment concerning such statistical issues as data analysis, parameter estimation and hypothesis testing in the context of general loss models can be found in Klugman, Panjer & Willmot (2004).

The Pareto severity model

Operational loss data is usually very heavy-tailed. Moscadelli (2004) investigated empirical loss data collected by the Basel Committee during the financial year 2001. Motivated by extreme value methods, for a generalised Pareto distribution (GPD) model, he estimated $1/\alpha$ in a range between approximately 0.6 and 1.5, corresponding to $\alpha$ roughly between 0.7 and 1.7. For all such $\alpha$ the severity distribution has infinite variance and for $\alpha \leq 1$ even the mean value does not exist.

Recall that the GPD model comprises distributions with compact support, exponential distributions and Pareto distributions, corresponding to $1/\alpha$ being negative, zero and positive, respectively. For such small positive values of $\alpha$ as observed by Moscadelli above, it is quite clear that it suffices to consider the GPD model family for positive finite values of $\alpha$, corresponding to a Pareto distribution.

The Pareto distribution has further properties, which we will exploit in this section.

\[ F(x) = \left( 1 + \frac{x}{\theta} \right)^{-\alpha}, \quad x > 0 \]

For simplicity, we denote $EN(t) = \lambda t$ with the obvious understanding that for a negative binomial process $\lambda$ has to be replaced by $\gamma / \beta$. As a result of the analytical operational VAR theorem, we obtain for the operational VAR:

\[ \text{VAR}(\kappa) = \theta \left( \frac{\lambda t}{1 - \kappa} \right)^{1/\alpha}, \quad \kappa \to 1 \quad (4) \]

(Actually, any severity distribution satisfying $F(x) - (x/\theta)^{-\alpha}$ as $x \to \infty$ yields approximation (4)).

Figure 1 compares the analytical VAR estimate (4) with the results of a Monte Carlo simulation for the Pareto LDA with different shape parameters $\alpha$ and $\theta = 1$. We see that the best approximation is obtained for extremely heavy-tailed data, that is, for small values of $\alpha$. Consequently, for operational loss data, our approximation should be very good.

\[ \text{Time scaling in the Pareto severity model.} \]

A well-known formula in risk management is the square-root-of-time rule for deriving multi-period VAR values from one-period values. This scaling law is based on the well-known property of the normal distribution, which says that the sum of $n$ IID centred normal random variables, when scaled by $\sqrt{n}$ is again normally distributed. As a generalisation, the central limit theorem guarantees that the sum of $n$ IID random variables with finite variance (with arbitrary distribution and centred by its mean) converges for $n \to \infty$ to a normal distribution. It can be shown that the central limit theorem holds also for Pareto LDA models, when proper adjustments have been made for the random number $N(t)$ of summands (see Embrechts, Klüppelberg & Mikosch, 1997, theorems 2.5.7 and 2.5.9). Note that for $\alpha < 2$ neither is scaling by $\sqrt{n}$ correct nor does the normal distribution appear as a limit for $n \to \infty$. Instead scaling has to follow a $1/\alpha$-root and the limit is a so-called stable distribution, which is much heavier-tailed than the normal law.

We are, however, not aiming at a limit law for $n \to \infty$, respectively $N(t) \to \infty$ (which means $t \to \infty$), but for a simple multi-period VAR based on one-period values. Moreover, we consider approximations in the very far tail of a heavy-tailed distribution. Consequently, a central limit argument may be misleading, and scaling with the square-root factor is not justified, even for a finite variance model.

We may, however, infer from (4) that for all fixed $t > 0$:

\[ \text{VAR}(\kappa) \sim t^{1/\alpha} \text{VAR}(\kappa), \quad \kappa \to 1 \quad (5) \]

Consequently, in the case of a Pareto LDA model, we have an $\alpha$-root-of-time rule for the operational VAR. Inserting typical values for $\alpha$, (5) implies that the threat of losses due to operational risk increases rapidly (and much faster than the outcome of the square-root-rule) when con-
considering future time horizons. To put it simply, operational risk can be a long-term killer!

Maxima of operational losses. Consider a VAR at confidence level $\kappa$ and time horizon $t = 1$ year, that is, the potential one-year loss that is exceeded only with small probability $1 - \kappa$. From the law of large numbers we know that for large $N$ an event with probability $p$ occurs on average $Np$ times in a series of $N$ observations. Therefore, in the case of yearly data, for $\kappa = 0.1\%$, VAR can be heuristically interpreted as the once-in-a-thousand-year event. There is, however, a different interpretation of VAR that is closely related to the sample maxima among a sequence of $N$ IID loss variables $X_i$ within a given time period $[0, t]$:

$$ M(t) = \max(X_1, \ldots, X_{N(t)}), \quad t \geq 0 $$

For the standard LDA definition, setting $P(N(t) = n) = p(n)$ and defining $M_n = \max(X_1, \ldots, X_n)$ for $n \in \mathbb{N}$, we can immediately calculate the distribution function $G_M$ of $M(t)$ for any fixed $t > 0$:

$$ G_M(x) = P(M(t) \leq x) = \sum_{n=0}^{\infty} p_n M_n P(M_n \leq x) = \sum_{n=0}^{\infty} p_n F^n(x), \quad x \geq 0 $$

Example (Poisson Pareto LDA). If the frequency follows a Poisson process with intensity $\lambda > 0$, we obtain:

$$ G_M(x) = \sum_{n=0}^{\infty} e^{-\lambda t} \left( \frac{\lambda t}{n!} \right)^n F^n(x) = e^{-\lambda t f^0(x)}, \quad x \geq 0 $$

(6)

We now ask for the most probable value $x_{\text{mp}}$ of the maximum, the mode of $G_M$. If $F$ has a differentiable density $f$ with derivative $f'$, then also $G_M$ has a differentiable density $g_M$ with derivative $g'_M$. In this case, the mode of $G_M$ is determined as the solution $x_{\text{mp}}$ to:

$$ g'_M(x) = e^{-\lambda f^0(x)} \left[ \lambda f^2(x) + f'(x) \right] = 0 $$

and, thus, $x_{\text{mp}}$ is the solution to:

$$ \lambda f^2(x) + f'(x) = 0 $$

For most realistic severity distributions $x_{\text{mp}}$ will be unique. In the important example of a Pareto distribution, we have:

$$ x_{\text{mp}} = \theta \left[ \left( \frac{\alpha \lambda}{1 + \alpha} \right)^{1/\alpha} - 1 \right] = \theta \left( \frac{\alpha \lambda}{1 + \alpha} \right)^{1/\alpha} $$

(7)

Note the similarity between the VAR formula (4) and the right-hand side of (7). We finally arrive at the following approximate relationship between the operational VAR at time horizon $t$ and the most probable value of the maximum loss event during that time period for $\kappa$ near one:

$$ \text{VAR}_t(\kappa) = \left( \frac{1 + 1/\alpha}{1 - \kappa} \right)^{1/\alpha} x_{\text{mp}} $$

(8)

It is worth mentioning that this result does not depend on the frequency process, but only on the shape parameter $\alpha$ and the confidence level $\kappa$. For any given $x_{\text{mp}}$, Table C clearly shows the sensitivity of the corresponding operational VAR of the shape parameter and the confidence level.

The question arises of whether (8) can be used as an alternative approximation for operational VAR. Unfortunately, estimating $x_{\text{mp}}$ by a reliable empirical method would require a vast amount of loss data, which is currently not available. The underlying data should consist of annual maxima for recent years, which define a histogram, from which $x_{\text{mp}}$ can be read off. Therefore, a large amount of annual maxima would have to be collected before $x_{\text{mp}}$ could be estimated, where presumably the IID property would be violated simply by non-stationarity in a long time series.

However, the right-hand side of (8) can, for instance, be estimated by scenario analyses and expert-based risk assessment. An experienced risk manager may estimate the maximum one-year loss caused by a single event within the next year. Annual maximal losses of previous years may guide the way. Such estimates, interpreted as the most probable value $x_{\text{mp}}$, then yield an expert approximation of the operational VAR as is required by the Basel Committee.

Klaus Böcker works in the risk control department at HypoVereinsbank in Munich. Claudia Klüppelberg holds the chair of mathematical statistics at the Center for Mathematical Sciences of the Munich University of Technology. Klaus Böcker would like to thank Martin Ottmann for stimulating discussions about operational risk issues. The opinions expressed in this article are those of the authors and do not reflect the views of HypoVereinsbank. Email: klaus.boecker@hvb.de, cklu@ma.tum.de

### Appendix. Tail behaviour of the aggregate loss distribution

The following theorem covers the standard loss distribution approach (LDA) with the two frequency models given in the main text.

**Theorem** (Embrechts, Klüppelberg & Mikosch, 1997, theorem 1.3.9). Consider the standard LDA $S(t) = \sum_{i=0}^{N(t)} X_i, \; t \geq 0$ from the definition given in the main text. Assume that the severities $X_i$ are subexponential with distribution function $F$. Fix $t > 0$ and define the aggregate loss distribution by $P(N(t) = n) = p(n)$ for $n \in \mathbb{N}$. Then, the aggregate loss distribution is given by:

$$ G_n(x) = \sum_{n=0}^{\infty} p_n F^n(x), \quad x \geq 0, \quad t \geq 0 $$

Assume that for some $\varepsilon > 0$:

$$ \sum_{n=0}^{\infty} (1 + \varepsilon)^n p_n(n) < \infty $$

(7)

Then, $G_n$ is subexponential with tail behaviour given by:

$$ \overline{G}_n(x) \sim E N(t) F(x), \quad x \to \infty $$

### REFERENCES

- Basel Committee on Banking Supervision, 2004
- International convergence of capital measurement and capital standards
- Basel
- Embrechts P, C Klüppelberg and T Mikosch, 1997
- Modelling extremal events for insurance and finance
- Springer, Berlin
- Klugman S, H Panjer and G Willmot, 2004
- Loss models – from data to decisions
- Wiley, Hoboken, New Jersey
- Moscavelli M, 2004
- The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee
- Banca d’Italia, Termini di Discussione 517

C. The factor $((1 + 1/\alpha)/(1 - \kappa))^{1/\alpha}$ of equation (8) for $\alpha$ and $\kappa$ in a realistic range

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\kappa$ 99.0%</th>
<th>$\kappa$ 99.9%</th>
<th>$\kappa$ 99.95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>77</td>
<td>524</td>
<td>934</td>
</tr>
<tr>
<td>1.0</td>
<td>200</td>
<td>2,000</td>
<td>4,000</td>
</tr>
<tr>
<td>0.8</td>
<td>871</td>
<td>15,496</td>
<td>36,857</td>
</tr>
</tbody>
</table>