Operational risks in banks

Through empirical analysis of operational risk in a bank we derive a model to represent the distribution of losses, which we compare with more traditional models of operational risk. Our findings suggest that the generalised extreme value distribution provides a good fit to the annual loss distribution and that some conventional methods to model severity are inadequate because they neglect the extreme percentiles which are important in the type of analysis required under Basel II.[1]




JOHN EVANS SF Fin is Associate Professor, Australian School of Business, University of NSW. Email: john.evans @unsw.edu.au

ROBERT WOMERSLEY is Associate Professor, Faculty of Science, University of NSW.

DANNY WONG is an Honours Student, Faculty of Science, University of NSW.

GREG A. WOODBURY Ernst & Young, Australia.

The nature of operational risk

Operational risk is usually defined as the risk resulting from inadequate or failed internal processes, people and systems, or from external events, including legal errors.

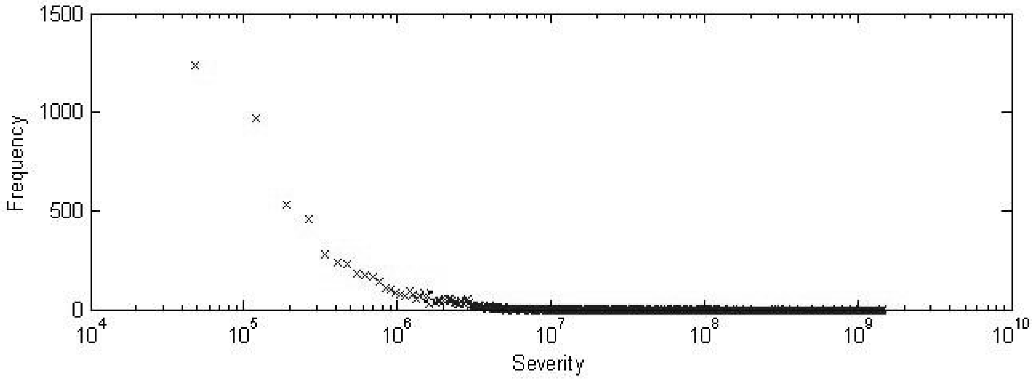
Regulators in Australia require banks to model their operational risks and to satisfy them not only of the suitability of the modelling process but also of the appropriate level of capital required to manage the residual risk such that there is a 99.9% overall probability of the survival of the bank. This implies a very high level of accuracy of modelling. Naturally, banks want the smallest estimate of expected losses that satisfies this confidence limit to minimise capital required.

One of the impediments to meeting this requirement is that most banks have not collected much operational risk data in the past as it was generally not needed and the cost was deemed to be unjustifiable. Even if the data had been collected, accuracy would have been an issue as indirect losses such as system errors, which cause delays in transactions, may produce losses which are not readily quantifiable and the duration of operational loss events can vary significantly.

There is also the problem of 'truncation', which refers to the minimum loss for reporting purposes, and this usually changes over time, and varies between banks, making interbank comparisons difficult.

There are two separate issues that need to be considered when evaluating operational risk: the severity of losses, i.e. the amount, and the frequency of losses. In this paper we will concentrate more on the severity issue.

FIGURE 1 **Frequency vs severity of operational risk losses**



Data

Data was obtained from a bank operating predominantly in the retail market for the period 1988 through 1996. The data reported here have been scaled so as not to identify the bank concerned.

The following table shows the summary statistics of the operational losses reported:

TABLE 1: **Summary statistics**

Statistic	Value
Mean	$2 million
Median	$360,000

Observations on the loss distribution

Some of the main observations are:

- There are a large number of small losses combined with a small number of large losses as indicated in the frequency versus severity plot in Figure 1.
- The time series plot in Figure 2 reveals clear evidence of extreme values. Figure 3, showing the empirical density of the losses and the empirical distribution on a logarithmic scale, also supports this view.
- The occurrences of the losses are irregularly spaced in time, suggesting non-stationarity.
- The severity and frequency of losses tend to decrease with time. This does not support the hypothesis that a reporting bias exists as suggested in some previous studies (e.g. Chavez-Demoulin and Embrechts [1] and Embrechts et al. [3]) where the severity and frequency tend to increase in time, suggesting an increased awareness and reporting of operational losses. Our results may reflect improved risk management practices and/or a modified data collection process.
- The data is very skewed and kurtotic. The kurtosis stems from the concentration of data points in the lower losses and the skewness is due to the extreme data points with the largest loss being approximately 64 standard deviations away from the mean.

Normality analysis

The distorted nature of the normality plot in Figure 4 clearly supports the hypothesis that operational risk data is not normally distributed, with both the QQ-plots and PP-plots deviating from the 45 degree linear reference line significantly.

FIGURE 2 **Time series plot of operational risk losses**

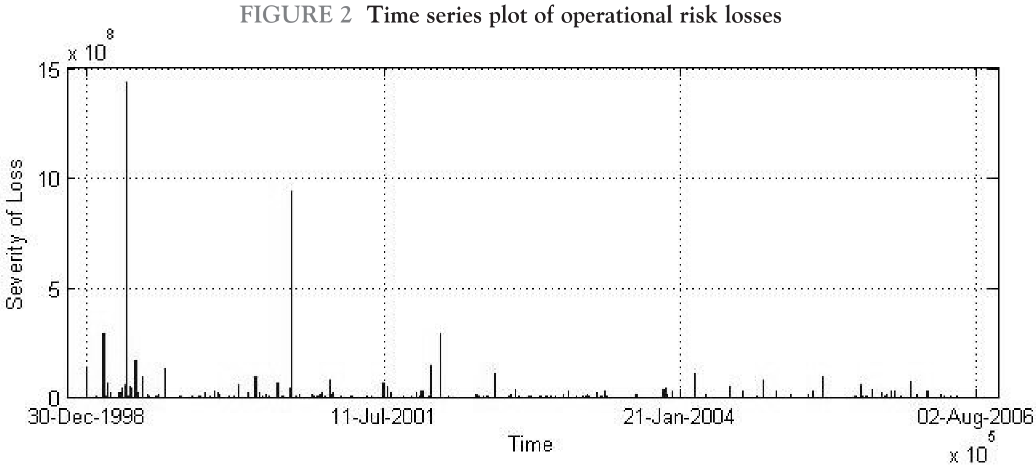


TABLE 3: **Parameter estimation for the severity of losses using the GDP distribution**

	Parameter	Estimate	95% CI	
MLE	σ	1966911.1422	1680562.7830	2302049.9327
($\mathbf{G}^{MLE(S)}$)	ξ	1.0397	0.8810	1.1986
PWM	σ	2438920.8784	2116459.8983	2903196.9420
($\mathbf{G}^{PWM(S)}$)	ξ	0.8099	0.6928	0.8704

TABLE 4: **Number of violations calculated using the fitted distributions**

	Number of Violations			
α	Theoretical	$G^{MLE(S)}$	$G^{PWM(S)}$	Lognormal
0.950	347	346	321	286
0.990	69	72	83	90
0.995	35	32	43	61
0.997	21	21	28	45
0.999	7	5	9	25

Lognormal analysis

Significant improvements are made when we model the data with a lognormal distribution as seen in Figure 5. The PP-plot almost coincides with the reference line and the majority of the QQ-plot is linear. It would seem the QQ-plot is much better at depicting the nature of operational risk tail events. However, due to the curvature at the tails, it is evident that even the lognormal distribution is unable to properly account for the extreme nature of the data.

Single distribution modelling

Both the Poisson and negative binomial distributions were applied to the data to model the frequency of losses. The Poisson distribution is inappropriate for modelling the frequency of losses as the ratio between the sample variance and sample mean should be approximately equal to 1 for the Poisson distribution to be suitable and, in our case, this ratio is 10.8. The estimated parameters and the corresponding 95% confidence levels are shown in Table 2. Comparisons of the expected results using the Poisson and negative binomial distributions and the actual results are shown in Figure 6.

TABLE 2: **Parameter estimation for the frequency of losses**

Distribution	Parameters	95% Confidence Level	
Poisson	$\lambda = 75.4176$	73.6333	77.2019
Negative	$r = 7.2923$	4.8808	9.7038
Binomial	$\frac{1}{1+\beta} = 0.0882$	0.0608	0.1155

The peaks over thresholds methodology in Extreme Value Theory was used to model the severity of the losses. Two different parameter estimation methods were used, namely maximum likelihood estimation (MLE) and probability weighted moments (PWM). The results can be seen in Table 3 where $\mathbf{G}^{MLE(S)}$ and $\mathbf{G}^{PWM(S)}$ denote the severity models corresponding to MLE and PWM estimated parameters, respectively.

From the QQ and PP plots in Figure 7 it is evident that using GDP does improve the fit for the loss data as the plots are fairly linear and coincide well with the 45 degree line. It is also interesting that again the QQ-plot fits the tail events better, whereas the PP-plot conceals the tail event information.

We considered the effect on VaR of $\mathbf{G}^{MLE(S)}$ and $\mathbf{G}^{PWM(S)}$ along with the lognormal distribution for comparison, and compared the number of violations from the model with the expected number of violations. A violation occurs when the data value exceeds the calculated VaR(α). For a confidence level α with n observations, the expected number of violations will be $n(1\text{-}\alpha)$. If the number of violations is higher than the expected number of violations, then the model underestimates the extreme risk. From Table 4 it is clear that $\mathbf{G}^{MLE(S)}$ provides the best fit in terms of least violations while the number of violations for the lognormal distribution significantly increases as the confidence level increases. Both the GPD models pass the Kupiec test in that they coincide with the null hypothesis that the data conforms with the selected model, whereas the lognormal rejects the null hypothesis at all α levels.

An aggregate loss distribution was formed by simulating annual aggregate losses and then fitting those losses to an appropriate distribution. One hundred thousand simulations were used. The frequency distribution used was the negative binomial with parameters as in Table 2. The severity was simulated using both sets of parameters $\mathbf{G}^{MLE(S)}$ and $\mathbf{G}^{PWM(S)}$. The simulations generated from $\mathbf{G}^{MLE(S)}$ and $\mathbf{G}^{PWM(S)}$ are denoted as $S(\mathbf{G}^{MLE(S)})$ and $S(\mathbf{G}^{PWM(S)})$, respectively. The statistical characteristic of the resulting simulation is shown in Figure 8. The

FIGURE 3 **(Top) Empirical distribution of losses; (Bottom) Empirical distribution of losses on a logarithmic scale (the dotted line represents the mean)**

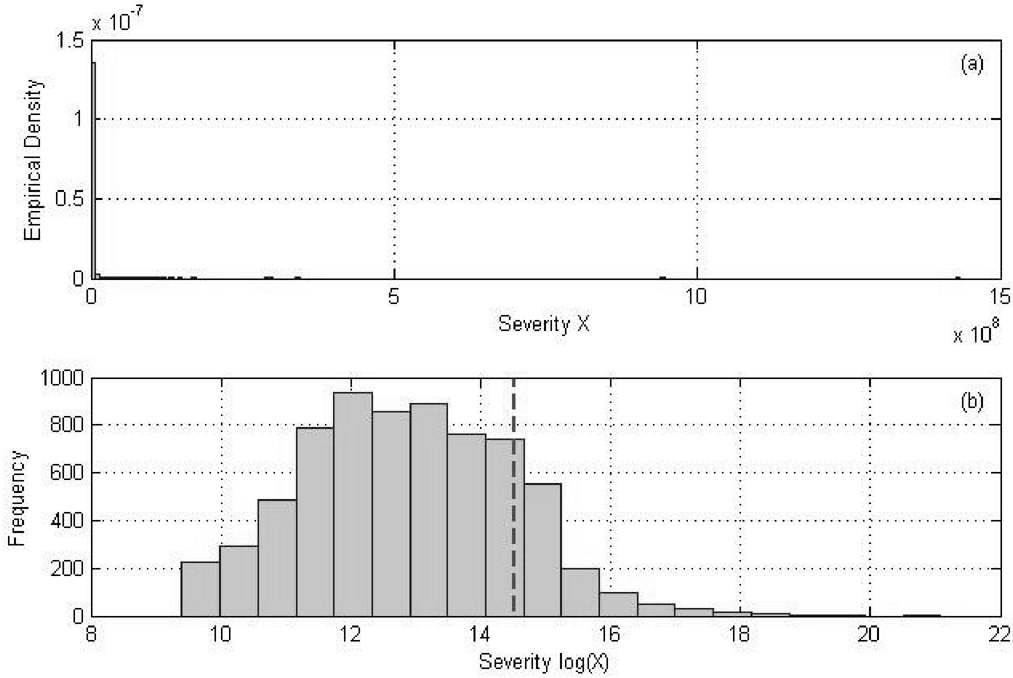


FIGURE 4: **Normality plot of the severity of losses (the diagonal line represents the 45 degree reference line)**

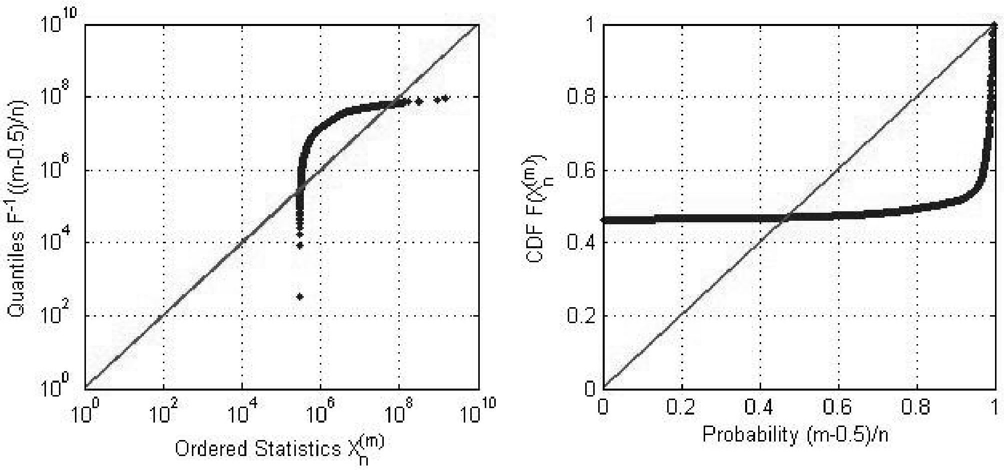
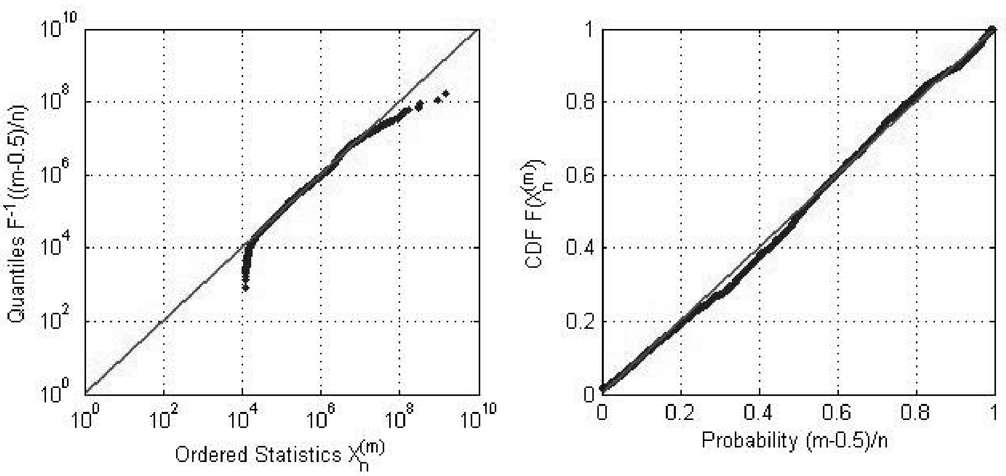


FIGURE 5: **QQ-plot using the lognormal distribution of the severity of losses (the diagonal line represents the 45 degree reference line)**



simulated data continues to show clear evidence of skewness and kurtosis even on a logarithmic scale.

The distribution is again very kurtotic and right-skewed as expected. The largest loss in $S(\mathbf{G}^{MLE(S)})$ is 1.3599×10^{13} and in $S(\mathbf{G}^{PWM(S)})$ is 7.1276×10^{11}. Clearly there is great inconsistency between these two methods as one estimate is over 22 times the other. Also, given the data scaling used, the MLE estimate is many times greater than typical estimates found in other studies. This problem of overstating VaR estimates was also highlighted by King [4]. To test sensitivity of the parameters, various combinations of σ and ξ were used to perform the simulation. We found that the size of the losses was particularly sensitive to changes in ξ but even large changes in σ did not have any noticeable effect on the size of losses.

FIGURE 6: **Plot of the empirical frequency distribution fitted with the Poisson and negative binomial distribution**

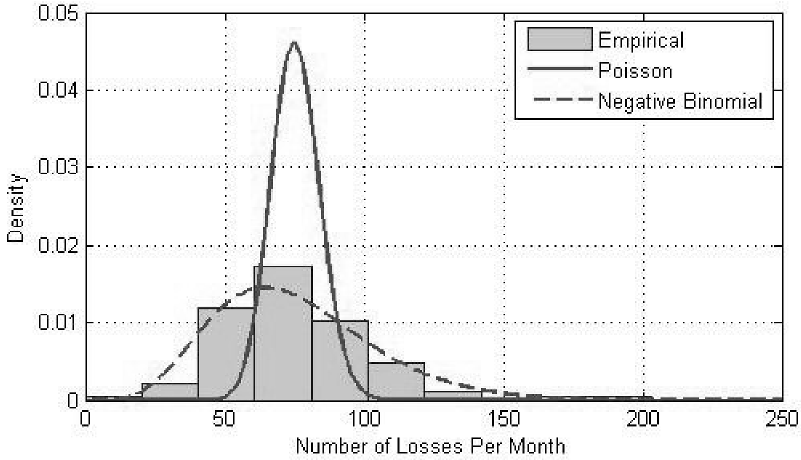


FIGURE 7: **QQ-plot (Left) and PP-plot; (Right) for the truncated severity data fitted to GDP using the MLE parameters**

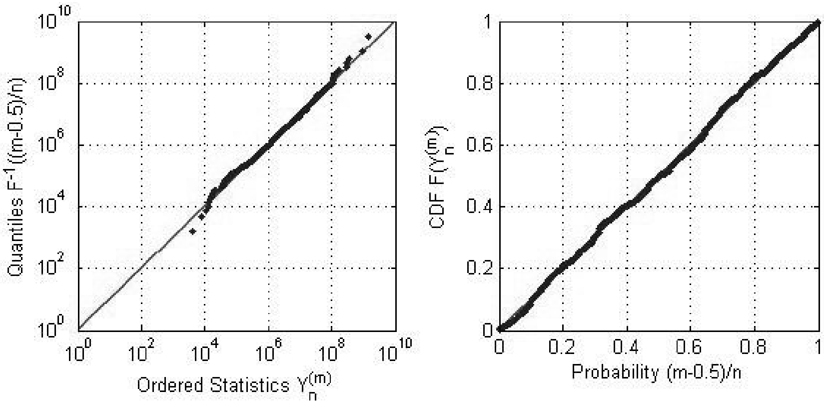


FIGURE 8: **Histogram of the aggregate losses for the simulation using the MLE parameters $G^{MLE(S)}$ on a logarithmic scale (the corresponding histogram for $\mathbf{G}^{PWM(S)}$ is similar, but slightly less extreme in nature).**

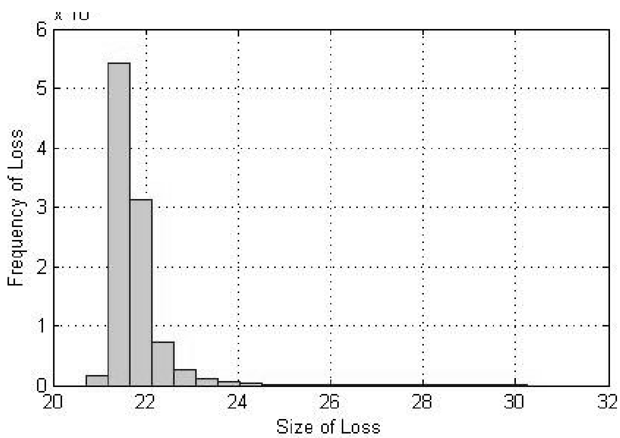


TABLE 5: Summary statistics of the aggregate losses under MLE and PWM

Statistic	Value (MLE)	Value (PWM)
Mean	4.6333×10^{9}	1.8301×10^{9}
Median	2.2529×10^{9}	1.5372×10^{9}
Variance	9.8909×10^{21}	3.1668×10^{19}
Standard Deviation	9.9453×10^{10}	5.6275×10^{9}
Semi-variance	1.0796×10^{22}	4.1584×10^{19}
Kurtosis	1.7138×10^{4}	2.0631×10^{4}
Skewness	1.2180×10^{2}	1.2985×10^{2}
Minimum	8.7975×10^{8}	6.7782×10^{8}
Maximum	1.6930×10^{13}	9.9939×10^{11}

Both the GEV and GPD were fitted to the simulated loss data. The GEV distribution provided a much better fit than the GPD so we only considered the GEV fit. We let $\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$ and $\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$ denote the GEV fit using MLE and PWM techniques to the simulated data $S(\mathbf{G}^{\text{MLE(S)}})$, respectively, and similarly, the notation for $S(\mathbf{G}^{\text{PWM(S)}})$ fits are $\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$ and $\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$. Both MLE and PWM methods gave roughly the same fit for the GEV distribution. The MLE approach seems to provide a slightly more accurate fit to the body of data when compared to the PWM approach. This can possibly be explained by the fact that we have enough simulated data points for the MLE approach to reach asymptotic convergence. The PWM, on the other hand, gives a heavier tail to the distribution and, as a result, performs better in the violations analysis shown in Table 6.

TABLE 6: Number of violations calculated using the fitted distributions

	Number of Violations				
α	Theoretical	$\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$	$\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$	$\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$	$\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$
0.950	5000	6168	3820	4833	3328
0.990	1000	2491	909	1839	843
0.995	500	1749	515	1309	512
0.997	300	1396	323	1019	346
0.999	100	817	124	660	150

The VaR(α) was calculated using the fitted GEV distributions. These results show that the VaR(α) increases with the confidence level (see Table 7). In addition, the MLE parameters $S(\mathbf{G}^{\text{MLE(S)}})$ produce significantly larger VaR(α) than the corresponding PWM parameters $S(\mathbf{G}^{\text{PWM(S)}})$. However, the inconsistencies remain with estimates varying by a factor of 22 times.

TABLE 7: Value-at-Risk values

	VaR(α)			
α	$\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$	$\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{MLE(S)}})}$	$\mathbf{G}^{\text{MLE(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$	$\mathbf{G}^{\text{PWM(A)}}_{S(\mathbf{G}^{\text{PWM(S)}})}$
0.95	6.2016×10^{9}	8.5478×10^{9}	3.0752×10^{9}	3.5329×10^{9}
0.99	1.1906×10^{10}	2.8702×10^{10}	4.6355×10^{9}	7.1614×10^{9}
0.995	1.5905×10^{10}	4.9976×10^{10}	5.5387×10^{9}	9.9851×10^{9}
0.997	1.9740×10^{10}	7.5671×10^{10}	6.3192×10^{9}	1.2854×10^{9}
0.999	3.1591×10^{10}	1.8651×10^{11}	8.4054×10^{9}	2.2479×10^{10}

Multi-distribution modelling

A possible solution for this tail estimation problem is the use of multiple distributions which will dampen the overestimating of the tails as it will place more weight on smaller losses and less weight on the tail. The use of multi-models essentially restricts the number of larger losses that can occur, thus giving a much more reliable estimate of the aggregate distribution as it takes into account the rarity of the extreme losses in the frequency.

Difficulties were encountered in attempting to fit multi-distributions. We used MLE to simultaneously maximise the likelihood in both distributions as well as the weighting factor. The algorithm used was based on trying to maximise the log-likelihood of the mixture solution but no optimal solution was found. The process was simplified by taking a multi-step approach. The data was split into two portions – smaller than threshold u_t (body X_b) and larger than u_t (tail X_t) as represented in Figure 9.

The losses X_b were fitted with the lognormal distribution using MLE and extreme value theory techniques applied to X_t. With l_m denoting the log-likelihood function for the mixture model and f_{LN} and f_{GPD} as the densities of the fitted Lognormal and GPD, the weighting factor was varied between zero and one to maximise

$$l_m = \sum \ln\left(wf_{\text{LN}} + (1-w)f_{\text{GPD}}\right)$$

This method also failed to converge to a solution. The resulting weights were either close to zero or close to one.

An empirically based technique was then used to implement the mixture distribution. We chose u to be the truncation point used in the bank-wide analysis, and the weighting factor was the empirical estimation for the proportion of losses less than u, that is, $w = \frac{\#\text{ losses less than } u}{\text{total \# losses}}$. The Lognormal distribution was estimated using MLE yielding parameters μ=12.5693 and σ^2=1.3522. The statistical characteristics of the mixed severity distribution are shown in Table 8.

TABLE 8: Statistics for the aggregate loss simulation produced from the mixture severity distribution

Statistic	Value (MLE)	Value (PWM)
Mean	4.4979×10^{9}	1.6299×10^{9}
Median	2.3252×10^{9}	1.3954×10^{9}
Variance	6.7723×10^{18}	1.2010×10^{19}
Standard Deviation	2.6024×10^{9}	3.4656×10^{9}
Semi-variance	9.4711×10^{18}	1.6111×10^{19}
Kurtosis	2.1616×10^{4}	2.4960×10^{4}
Skewness	1.1967×10^{2}	1.2918×10^{2}
Minimum	6.1717×10^{8}	6.4101×10^{8}
Maximum	5.5123×10^{11}	7.6331×10^{11}

The use of the multi-distributions produced more consistency between the MLE and PWM estimates of maximum loss. This is evident when Table 5 and Table 8 are compared. Also, unusually large estimates of maximum loss are not generated as is the case for the MLE estimate in Table 5. Furthermore, the mean and median remained stable for both methods. The GEV distribution generated a better fit.

Comparisons with other results

Our results demonstrate major differences with other studies. Our value-at-risk amounts are smaller than those for similar-sized US banks that hold a lower proportion of residential mortgage loans in their accounts [5]. This is consistent with the notion that residential mortgages create less operational risk than other business lines and, consequently, the need to hold large capital reserves is reduced.

The analysis performed by de Fontnouvelle et al. [2] indicates that non-US operational losses are significantly larger than US losses. The percentiles for the non-US losses are approximately double the equivalent percentiles for US losses at both the aggregate and business line level. This is inconsistent with our data set, yielding capital reserves which are much smaller than would be expected from the earlier papers.

Another inconsistency is the modelling of the frequency of losses where most banks have used the Poisson distribution. This is most likely due to the greater number of favourable statistical properties inherent in the Poisson distribution rather than the ability to produce a better fit. There is also the possibility that data is reported in clusters creating an artificial over-dispersion that is then ignored when fitting a distribution.

At a recent presentation at UNSW, Mario Wuthrich of ETH Zurich [6] presented his work on the approach to determining operational risk using multiple sources of information. Essentially he suggests an approach that could be represented empirically as:

$$F(x) = w_1F_{int}(x) + w_2F_{ext}(x) + (1 - w_1 - w_2)F_{expert}(x)$$

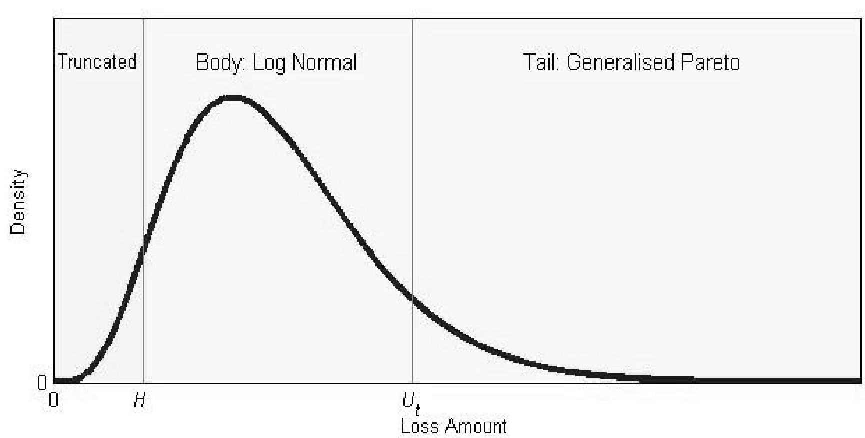
where w_1 and w_2 refer to weights given to the internally derived distribution (F_{int}) and the externally derived distribution (F_{ext}), and these weighted distributions are then combined with the distribution estimated for the current projection period by an expert (F_{expert}). While this approach is different to our multi-distribution approach, the result may be similar.

Several different approaches have been described that may improve on the VaR estimates given here. These include the incorporation of a reporting probability function to account for underreporting of smaller losses due to a reporting bias, the use of uncertainty minimisation or generalised linear modelling to combine multiple sources and kinds of data [7].

Regulatory implications

Our analysis would suggest that using the standard Poisson distribution approach to modelling operational risk may be introducing significant model errors. This would be correct if the regulatory approach were to try to fit the models as closely as possible to reality with capital levels required to be sufficient to ensure a 99.9% probability of a bank's

FIGURE 9: Illustration of the mixed distribution concept (H is the truncation point and U_t is the value where the tail is believed to begin).



survival. However, this may not be correct as it is entirely reasonable to use a simplified model that reduces costs and produces a sufficiently high capital requirement that the regulator is content with the probability of survival.

Conclusion

Extreme value theory has demonstrated significant potential to account for the heavy tail of operational losses where other conventional methods fail. We have shown statistically that the use of conventional methods to model severity is inadequate because the operational loss data exhibits kurtotic and right-skewed behaviour while conventional models place emphasis on fitting the central body of the data, and thus, neglect the extreme percentiles. It is the extreme losses that are important in the type of analysis required under Basel II.

However, a major limitation in the implementation of extreme value theory is the lack of data. This inhibits capturing the generalised Pareto nature of the excess distributions. The ability to model any sort of dependence is also limited by the availability of quality data. Even if we could overcome these limitations, the regulators may not permit the use of models based on such a small sample despite the accuracy of the dependence models. As such, it may take many years before any banks can convincingly justify the use of any sophisticated dependence structure between the various risk cells and reap the benefits of diversification.

In our view, the Poisson distribution, which is the most common distribution used to model operational risks, has proved to be inappropriate for our data set.

Of course, it is recognised that the use of internal data for the measurement of operation risk capital is not the only source; external data are also used, and scenario analysis may well overcome the shortcomings of the modelling of internal data that we have detected. We also acknowledge that it may be necessary to make adjustments for changes in business factors and changes in controls.

References

1. Chavez-Demoulin, V. and Embrechts P., 2004, *Advanced external models for operational risk*, Dept. of Mathematics ETH-Zentrum.
2. de Fontnouvelle, P., DeJesus-Rueff, V., Jordan, J. and Rosengren, E. 2003, *Using loss data to quantify operational risk*, Federal Reserve Bank of Boston.
3. Embrechts, P., Furrer, H. and Kaufmann, R. 2003, *Quantifying regulatory capital for operational risk, derivative use, trading & regulation*, Dept. of Mathematics ETH-Zentrum.
4. King, J.L. 2001, *Measurement and modelling operational risk*, Wiley Finance.
5. Moscadelli, M. 2004, The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee, discussion paper, Banca d'Italia, Temi di discussione del Servizio.
6. Wuthrich, M. 2007, Modelling operational risk: combining different sources of information, presentation at UNSW seminar, August.
7. http://oprisk.austega.com/components/com_fireboard/uploaded/files/Op_Risk_Internet_post1-156ef665ab8249f848e4586465d48e59.pdf

This paper is based on a presentation to the Institute of Actuaries of Australia Convention, 2007.