

ESTIMATING LOSS SEVERITY DISTRIBUTION: CONVOLUTION APPROACH

Ro J. Pak

Department of Applied Statistics, Dankook University, Korea

Received 2014-04-21; Revised 2014-04-24; Accepted 2014-05-02

ABSTRACT

Financial loss can be classified into two types such as expected loss and unexpected loss. A current definition seeks to separate two losses from a total loss. In this article, however, we redefine a total loss as the sum of expected and unexpended losses; then the distribution of loss can be considered as the convolution of the distributions of both expected and unexpended losses. We propose to use a convolution of normal and exponential distribution for modelling a loss distribution. Subsequently, we compare its performance with other commonly used loss distributions. The examples of property insurance claim data are analyzed to show the applicability of this normal-exponential convolution model. Overall, we claim that the proposed model provides further useful information with regard to losses compared to existing models. We are able to provide new statistical quantities which are very critical and useful.

Keywords: Convolution, Heavy-Tailed Distribution, Loss, Value at Risk (VaR)

1. INTRODUCTION

Operational risk is one of the most important risks for commercial banks; hence, it is essential to estimate the economic capital for operational risk. Operational risk is defined as 'the risk arising from inadequate or failed internal processes, people and systems or from external event' (BCBS, 2001; 2004a). The Basel report (BCBS, 2001; 2004b) suggests three methods for calculating operational risk capital charges: (i) The basic indicator approach, (ii) the standardized approach and (iii) the Advanced Measurement Approach (AMA). In particular, the Loss Distribution Approach (LDA) is the major part of the AMA. Here, estimating a loss severity distribution is one of the key processes. The parametric models in particular, such as the lognormal, the Weibull and generalized Pareto distributions, are used to estimate a particular severity distribution. Sometimes, mixture models combining the parametric models are used (Carvalho and Marinho, 2007).

In this article, a new type of loss severity distribution is suggested and the suggested distribution is based on the convolution of an expected loss severity distribution

and an unexpected loss severity distribution. Unlike the existing models, the new model provides the expected values and the VaR's for both expected and unexpected losses. Even when a future loss is anticipated, the new model provides the conditional expected values given a loss as well as the conditional VaR have given a loss for both expected and unexpected losses. In section 2 and 3, we briefly review the definition of loss and introduce the normal-exponential convolution model (norm-exp). Further, in section 4, two examples are investigated for verifying the usefulness of the normal-exponential convolution model in estimating loss severity.

2. REDEFINITION OF LOSSES

The Basel report (BCBS, 2001a; 2004b) classifies financial losses due to operational factors into the following two types:

- Expected losses: These are considered as the 'normal' losses that occur frequently, as part of everyday business, with low severity. Examples include losses due to accidentally miscalculated foreign exchange transactions

- Unexpected losses: These are the unusual losses that occur rarely and have a high severity. Examples include losses resulting from a major fraud activity

The current Basel II regulatory text does not provide a description/definition of the Expected Loss (EL). Consequently, discussion has often fluctuated between three different interpretations of EL (FSA, 2005):

- A business/management style definition in which EL is related to a future amount of expense/loss that is predicted on the basis of past experience, e.g., for credit card fraud; past experience of losses allows a projection of future losses, which is budgeted/priced for. The events giving rise to these future losses have not yet occurred
- A mathematical style definition in which EL is equated to the mean (50th percentile) of a loss distribution
- A financial accounting style definition in which EL, which describes losses expected from identified events, for which a reserve has been established. A common example of this is where a large legal cost is anticipated, but the exact amount of the legal settlement is not yet known

Figure 1 shows the distinction between expected and unexpected losses from the loss distribution point of view. The demarcation is purely arbitrary; however, the expected value of loss ($E[L]$) is often used as a point separating the expected loss and unexpected loss. In order to quantify both expected and unexpected losses, it is essential to find a loss distribution whose tail represents the unexpected losses (Neil *et al.*, 2005).

Unlike the existing definition of losses identified by a separation (**Fig. 1**), we would like to redefine the (total) loss as a sum of the expected loss and unexpected loss, such that the distribution of loss is a convolution of the distributions of the expected loss and unexpected loss. The existing definition says that loss is really a numerical summation of expected and unexpected losses, while we propose that loss is a sum of random variables in probabilistic sense.

The general consensus is that heavy-tailed distributions fit historically observed losses better than light-tailed distributions; the typical probabilistic models that have proposed to describe the severity of losses are the Lognormal, Gamma and Weibull distribution. We propose to let the expected loss distribution and the unexpected loss distribution be the truncated normal distribution and the exponential distribution, respectively (**Fig. 2**).

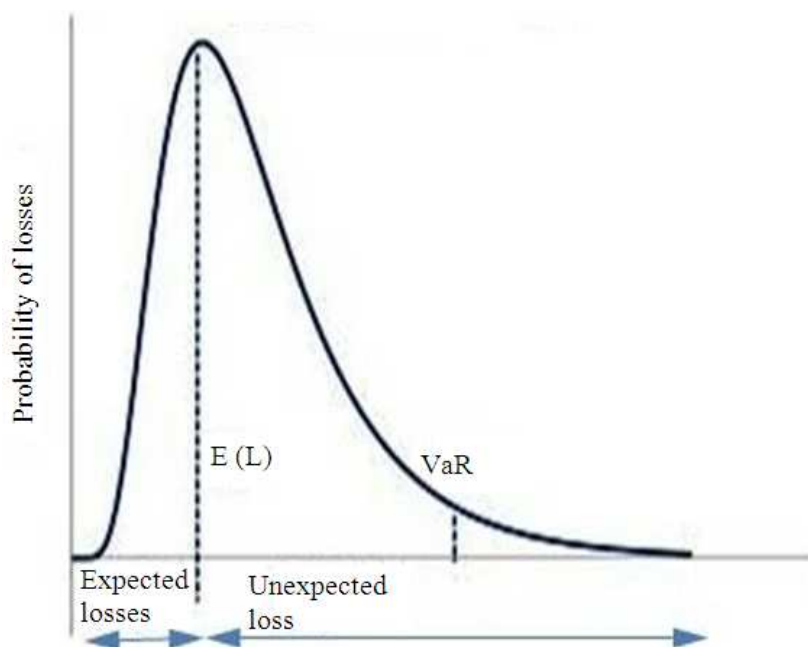


Fig. 1. Loss distribution

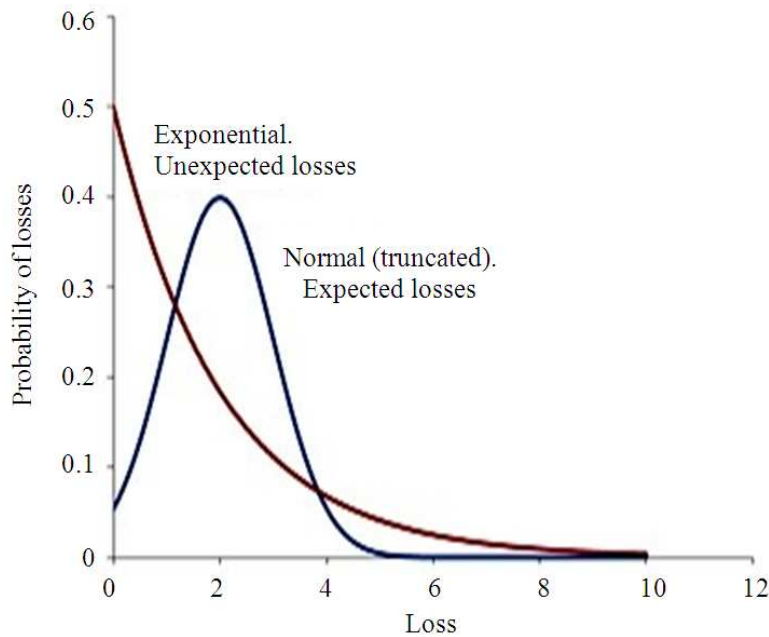


Fig. 2. Expected and unexpected losses

3. NORMAL-EXPONENTIAL CONVOLUTION

If we let X and Y represent the expected loss and unexpected loss, respectively, then S represents the loss which is $X+Y$. Let X be normally distributed with mean μ and variance σ^2 , but truncated at 0 and Y be exponentially distributed with mean $1/\alpha$. Then, $S = X+Y$ has a density:

$$f_s(s; \alpha, \mu, \sigma) = \alpha \exp\left(\frac{1}{2}\alpha^2\sigma^2 - \alpha(s - \mu)\right) \left[\Phi\left(\frac{\mu}{\sigma} + \alpha\sigma\right) - \Phi\left(\frac{\mu - s}{\sigma} + \alpha\sigma\right) \right],$$

where $\Phi(\cdot)$ is a cumulative normal distribution. The $f_s(\cdot)$ is called as the norm-exp density (Bolstad, 2004). Parameter μ controls the horizontal position, parameter σ controls the width near peak and parameter α controls the length of the tail (Fig. 3). The conditional density of X given S and Y given S are:

$$f_{X|S}(x | s) = \frac{\frac{1}{b}\phi\left(\frac{x + a - s}{b}\right)}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s - a}{b}\right) - 1}, f_{Y|S}(y | s) = \frac{\frac{1}{b}\phi\left(\frac{y - a}{b}\right)}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s - a}{b}\right) - 1}$$

where, $a = s - \mu - \alpha\sigma^2$ and $b = \sigma$, respectively.

The (conditional) expected value of the expected loss given the loss becomes:

$$E(X | S = s) = \mu + \alpha\sigma^2 - b \left[\phi\left(\frac{a}{b}\right) - \phi\left(\frac{s - a}{b}\right) \right] / \left[\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s - a}{b}\right) - 1 \right]$$

And the (conditional) expected value of the unexpected loss given the loss then becomes:

$$E(Y | S = s) = s - E(X | S = s).$$

$VaR(X|S)$, $VaR(Y|S)$, which can be obtained by taking the percentile of the corresponding conditional loss distributions at the desired confidence.

All of the mathematical expressions in this section can be obtained by closely following Bolstad (2004), while Bolstad (2004) provided the expressions only for Y . We add some mathematical derivations about the norm-exp at the end of this article.

4. DATA ANALYSIS

Example 1

The Danish data on large fire insurance losses, which were collected by a Danish reinsurance company from Jan. 3, 1980 to Dec. 31, 1990, are considered as an example.

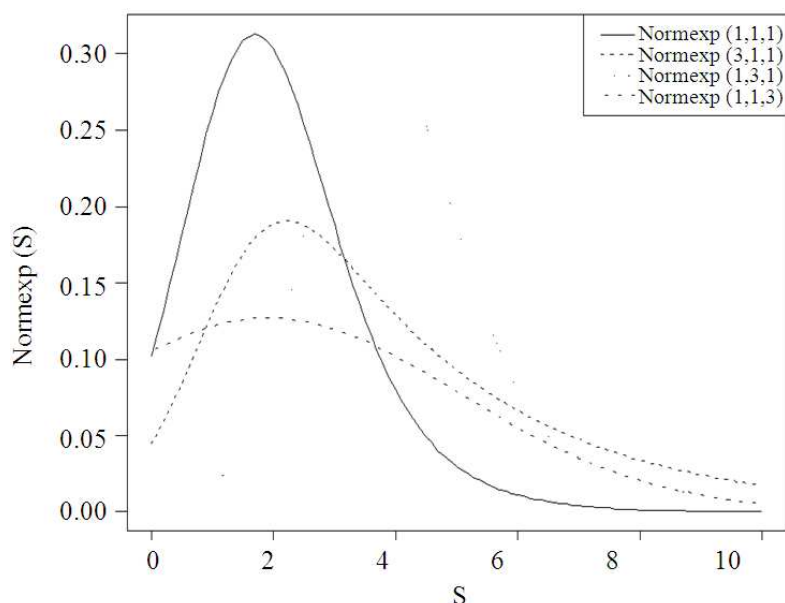


Fig. 3. Various norm-exp densities

This dataset contains 2,167 individual losses in million Danish Kroner. The dataset is accompanied with the R package *fitdistr* and it analyzed with R packages of both *fitdistr* and *limma*.

To confirm the model adequacy, some tests have been conducted and statistics are listed in **Table 1**. The critical value of K-S test at the 5% significant level can be calculated by $1.36\sqrt{1/n}$ (Barrett and Donald, 2003), which is 0.0292 for this dataset. All of the models we considered does not provide perfect goodness of fit. Among the four models, the Lognormal fits the data best; norm-exp, Gamma and Weibull followed in order in terms of the log likelihood and Akaike Information Criterion (AIC).

The VaR of loss, $VaR(S)$, based on the norm-exp is about 7.575, which is slightly larger than 6.536 by the lognormal distribution but smaller than the VaR's by Gamma and Weibull (**Table 2**). In **Fig. 4**, we plot a histogram of losses and estimated densities for the total, expected and unexpected losses. The density of expected losses is primarily located near 1.0 but the density of unexpected losses has a long tail. The sum of $VaR(X)$ and $VaR(Y)$ ($= 0.945+6.770$) is also slightly larger than $VaR(S)$ ($= 7.575$) that is, $VaR(S) < VaR(X) + VaR(Y)$, as expected by the nature of VaR. $VaR(S)$ is accounted mainly by $VaR(Y)$ (VaR by unexpected loss). We can observe that $E[X|S]$ and $VaR[X|S]$ are being bounded when s is over 1.0 and

$E[Y|S]$ and $VaR[Y|S]$ are denominating $E[X|S]$ and $VaR[X|S]$ after s is beyond 1.7, that is, if the (total) loss is larger than 1.7, the loss is taken accounted primarily by the unexpected loss.

Example 2

We deal with the Korean property insurance claim dataset as an example, which consists of the monthly total claims (losses) from March 2002 to January 2009 in Korea. The claim is monthly collected by the Insurance Statistics Information Service in Korea. The recorded data have been adjusted in 100 billion Korean won (U.S \$ 1 \cong 1000 Korean won). The Weibull density is best fitted for the given dataset from all criteria (**Table 3**). The norm-exp density is also suitable for the given data based on the K-S test.

Figure 5a and b Claims seem to follow norm-exp (0.2294, 0.4888, 0.3081). Weibull and gamma fit the data well as a whole, but lognormal and norm-exp fit the peak better than Weibull and gamma. The density of the expected claim is estimated as $N(0.4888, 0.3081)$ and that of the unexpected claim is estimated as $Exp(0.2294)$. We have $E[X] = 0.4888$, $E[Y] = 4.3592$ and $E[S] = 0.4888+4.3592 = 4.8480$.

Figure 5b We have $VaR[X] = 0.9955$, $VaR[Y] = 13.0550$, $VaR[S] = 13.5890$ ($VaR[S] < VaR[X] + VaR[Y]$). The VaR based on norm-exp is the second largest among the four VaR's (**Table 1**).

Table 1. Goodness-of-fit statistics

Distribution	Parameters	-2Log likelihood	AIC	K-S
Gamma	shape = 1.2579 Rate = 0.4107	5243.027	10488.054	0.201
Weibull	shape = 0.9475 scale = 2.9524	5270.471	10542.942	0.255
Lognormal	meanlog = 0.6718 sdlog = 0.7323	4433.891	8869.782	0.127
Norm-exp	rate = 2.2599 mean = 0.8024 sd = 0.0873	4641.394	9285.788	0.227

A-D: Anderson-Darling statistic, CVM: Cramér-Von Mises statistic, K-S: Kolmogorov-Smirnov statistic

Table 2. VaR (95%-quantile)

Distribution	VaR
Gamma	8.469
Weibull	9.398
Lognormal	6.536
Norm-exp	7.575
normal (expected loss)	0.945
Exp (unexpected loss)	6.770

Table 3. Goodness-of-fit test statistics and VaR (95%)

Distribution	Parameters	-2Loglikelihood	A-D	CVM	K-S	VaR
Lognormal	mean = 1.2863 s. d. = 0.9028	210.8466	2.3113 (0.00)	0.3499 (0.00)	0.1300 (0.12)	15.97
Gamma	shape = 1.8610 Rate = 0.3839	210.0977	1.0330 (0.00)	0.1485 (0.00)	0.0997 (0.39)	11.76
Weibull	shape = 1.5720 Scale = 5.3696	198.7650	0.7184 (0.0058)	0.0948 (0.13)	0.0850 (0.60)	10.79
Norm-exp	rate = 0.2294 mean = 0.4888 s.d. = 0.3081	411.6592	2.4926 (0.00)	0.4264 (0.00)	0.1135 (0.25)	13.58

A-D: Anderson–Darling statistic, CVM: Cramér-Von Mises statistic, K-S: Kolmogorov-Smirnov statistic.
p-values are in parentheses.

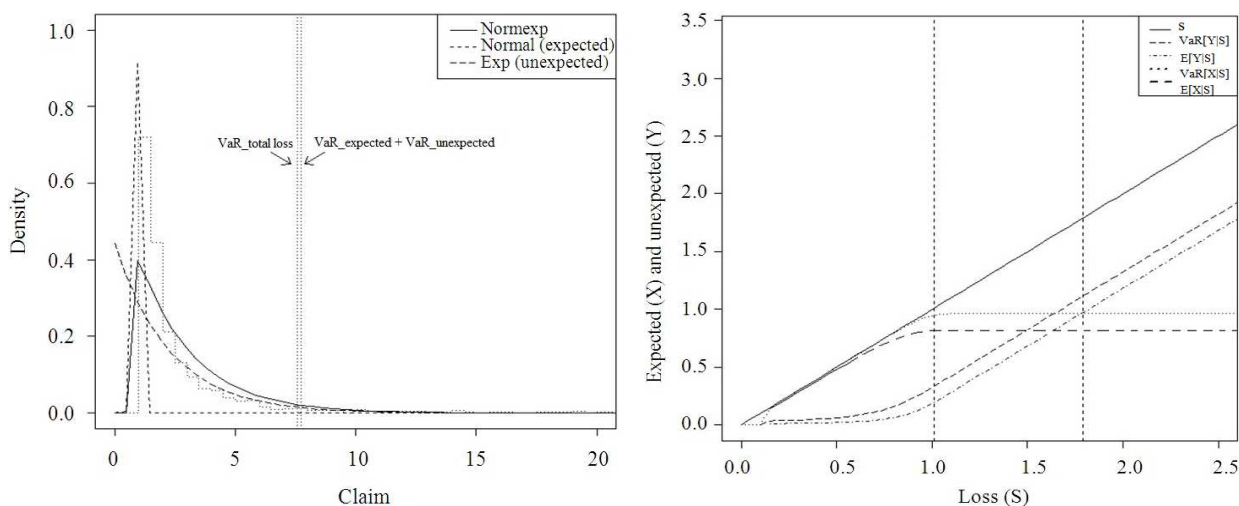


Fig. 4. Loss, expected and unexpected losses; $E[X|S]$, $VaR[E|S]$, $E[Y|S]$ and $VaR[Y|S]$

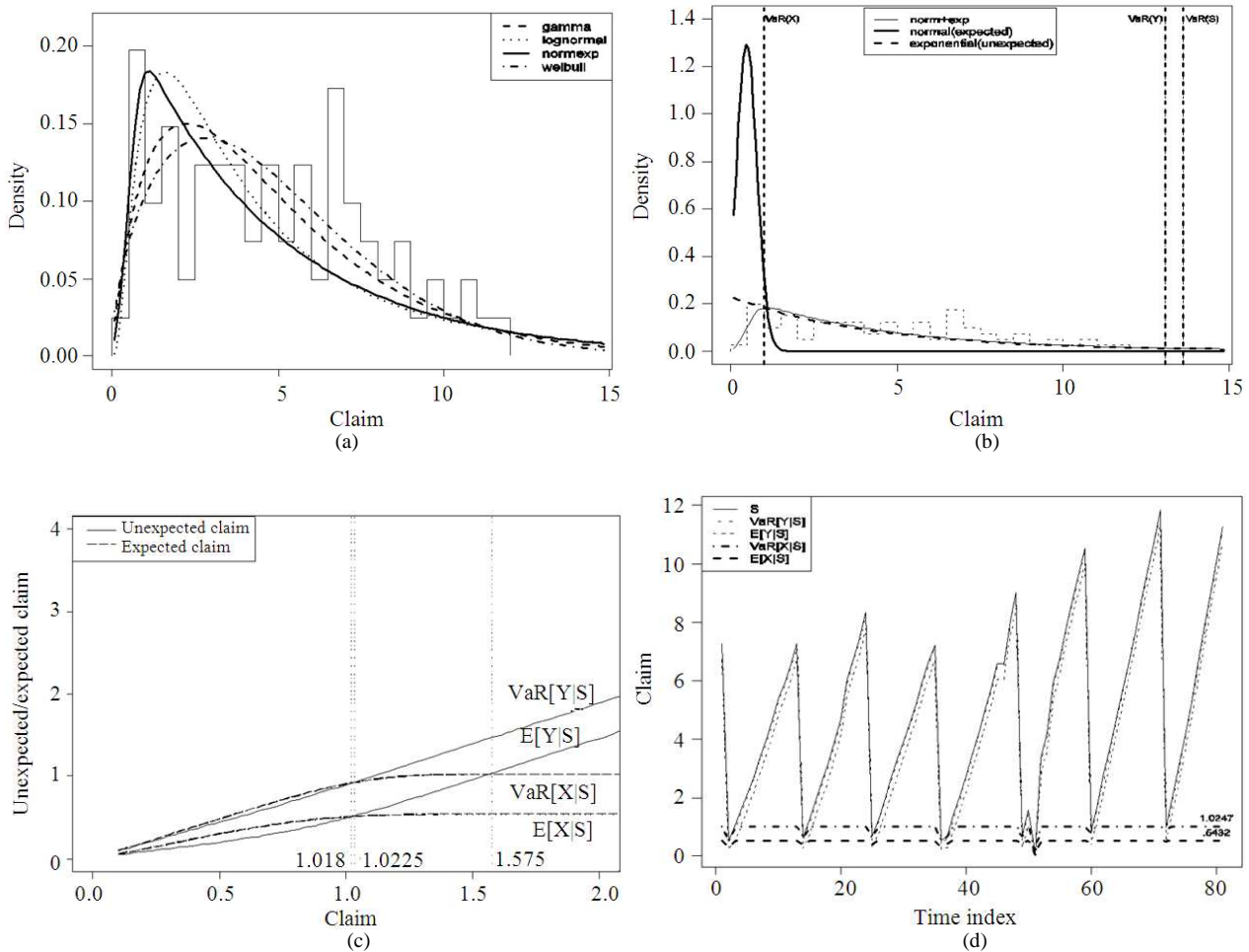


Fig. 5. Related plots for the Korean property insurance claim data (a) fitted densities (b) normal and exponential: VaR's (c) Expected and unexpected cla (d) $E[X|S]$, $E[Y|S]$, $VaR[X|S]$ and $VaR[Y|S]$

Figure 5c Claims decrease in the summer and increase in the winter. The same pattern is repeated every year. $E[X|S]$ and $VaR[X|S]$ are bounded by 0.5432 and 1.0247, respectively.

Figure 5d As claim increases, $E[Y|S]$ and $VaR[Y|S]$ are increasing while $E[X|S]$ and $VaR[X|S]$ are bounded by 0.5432 and 1.0247, respectively. If a claim is less than 1.0180, $E[Y|S] < E[X|S]$; if a claim is less than 1.0225, $VaR[Y|S] < VaR[X|S]$. Beyond a claim of 1.0180 or 1.0225, the inequalities are reversed. Furthermore, when a claim is larger than 1.5750, both $E[X|S]$ and $VaR[X|S]$ become less than $E[Y|S]$ and $VaR[Y|S]$. When a claim is lower than 1.0180 or 1.0225, the expected loss dominates the unexpected loss. More specifically, it can be said that

a loss is as usual, as expected. When a claim is between 1.0225 and 1.5750, the unexpected loss tends to be over the expected loss; however, it is still under the expected boundary. Yet, when a claim is larger than 1.5750, losses are taken accounted primarily by the unexpected losses.

The norm-exp fits the data well, especially in the peak area (**Fig. 4a**). Claims are mainly affected by unexpected claims while expected claims are somehow bounded (**Fig. 4c**). Approximately 1.0 (100 billion Korean dollars), where $E[Y|S] > E[X|S]$ and $VaR[Y|S] > VaR[X|S]$, is the turning point where we should begin to seriously worry about unexpected losses (**Fig. 4b, d**).

5. DISCUSSION

Most of quantities in the above examples are not available under the current definition of expected and unexpected losses. We are able to provide new statistical quantities which, we believe, are very critical and useful. We can provide more concrete and probabilistic relation between expected and unexpected losses on condition of (total) loss. Many representative researches (Resnick, 1997; McNeil *et al.*, 2010; Cooray and Ananada, 2005) supplied lot of important results on estimating losses, but did not give much information about what happened on expected and unexpected losses when a loss changed or when a loss was conditioned on. By the proposed method, we can visualize how expected and unexpected losses behave as a loss moves.

6. CONCLUSION

In this study, we redefine loss as a sum of expected and unexpected losses. The loss follows a distribution, which is referred to as the normal-exponential distribution (norm-exp). Once data is fitted well by the norm-exp, we can obtain useful information with regard to the expected and unexpected loss unlike the other existing distributions. In this article, we only consider the exponential distribution for unexpected loss, but there are certainly the other distributions with a long tail like Lognormal and Gamma, etc. for modeling unexpected loss. Convoluting those distributions would be a good topic for future research.

7. REFERENCES

BCBS, 2001. Consultative paper on operational risk. Consultative Paper 2, Bank of International Settlement, Basel Committee on Banking Supervision.

BCBS, 2004a. Working paper on the regulatory treatment of operational risk. Basel Committee on Banking Supervision.

BCBS, 2004b. Basel II: International convergence of capital measurement and capital standards: A revised framework. Bank for International Settlement.

Bolstad, B.M., 2004. Low-level analysis of high-density oligonucleotide array data: Background, normalization and summarization. Ph.D. Thesis, University of California, Berkeley.

Barrett, G.F. and S.G. Donald, 2003. Consistent tests for stochastic dominance. *J. Economet. Society*, 71: 71-104. DOI: 10.1111/1468-0262.00390

Carvalho, A.X. and A.S. Marinho, 2007. Mixture models in operational risk. *Development of ObjectRisk*.

Cooray, K. and M. Ananda, 2005. Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial J.*, 2005: 321-334. DOI: 10.1080/03461230510009763

McNeil, A.J., R. Frey and P. Embrechts, 2010. *Quantitative Risk Management: Concepts, Techniques and Tools*. 1st Edn., Princeton University Press, Princeton, ISBN-10: 140083757X, pp: 608.

FSA, 2005. Treatment of expected losses in capital calculations. Draft by FSA AMA Quantitative Expert Group. Financial Services Authority.

Neil, N., N. Fenton and M. Taylor, 2005. Using Bayesian networks to model expected and unexpected operational losses. *Risk Anal.*, 25: 963-972. PMID: 16268944

Resnick, S., 1997. Discussion of the Danish data on large fire insurance losses. *Astin Bulletin*, 27: 139-151. DOI: 10.2143/AST.27.1.563211

Appendix

The following proof closely follows Bolstad (2004) except that densities of X and Y have been switched and so the order of integrations also is changed. Let X be normally distributed with mean μ and variance σ^2 but truncated at 0 and let Y be exponentially distributed with mean $1/\alpha$. The joint density of X and Y is just product of densities:

$$f_{X,Y}(\alpha, x; \mu, \sigma, y) = \alpha \exp(-\alpha y) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right), \text{ where } 0 < x, y$$

The total loss S is defined as $S=X+Y$ and then the joint distribution of X and S is with the absolute value of the Jacobian $|J| = 1$:

$$f_{X,S}(\alpha, x; \mu, \sigma, s) = \alpha \exp(-\alpha(s-x)) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right), \text{ where } 0 < x < s.$$

The marginal distribution of S is:

$$\begin{aligned}
 f_s(\alpha, \mu, \sigma, s) &= \int_0^s \alpha \exp(-\alpha(s-x)) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx \\
 &= \int_{-\mu/\sigma}^{(s-\mu)/\sigma} \alpha \exp(-\alpha(s-\sigma w - \mu)) \phi(w) dw, \\
 &\quad \text{where } w = (x - \mu) / \sigma \\
 &= \alpha \exp(-\alpha(s - \mu)) \\
 &\quad \int_{-\mu/\sigma}^{(s-\mu)/\sigma} \exp(\alpha\sigma w) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w^2}{2}\right) dw, \\
 &= \alpha \exp(-\alpha(s - \mu)) \exp\left(\frac{\alpha^2\sigma^2}{2}\right) \times \\
 &\quad \int_{-\mu/\sigma}^{(s-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(w - \alpha\sigma)^2}{2}\right) dw \\
 &= \alpha \exp(-\alpha(s - \mu)) \exp\left(\frac{\alpha^2\sigma^2}{2}\right) \times \\
 &\quad \int_{-\mu/\sigma - \alpha\sigma}^{(s-\mu)/\sigma - \alpha\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz, \\
 &\quad \text{where } z = w - \alpha\sigma \\
 &= \alpha \exp\left(\frac{\alpha^2\sigma^2}{2} - \alpha(s - \mu)\right) \times \\
 &\quad \left[\Phi\left(\frac{\mu}{\sigma} + \alpha\sigma\right) - \Phi\left(\frac{\mu - s}{\sigma} + \alpha\sigma\right) \right]
 \end{aligned}$$

The conditional density of X on S and the corresponding expectation are:

$$\begin{aligned}
 f_{X|S}(x | s) &= \frac{f_{X,S}(x, s)}{f_s(s)} \\
 &= \frac{\alpha \exp(-\alpha(s-x)) \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\alpha \exp\left(\frac{1}{2}\alpha^2\sigma^2 - \alpha(s-\mu)\right) \left[\Phi\left(\frac{\mu}{\sigma} + \alpha\sigma\right) - \Phi\left(\frac{\mu-s}{\sigma} + \alpha\sigma\right) \right]} \\
 &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu-\alpha\sigma^2)^2\right)}{\Phi\left(\frac{\mu}{\sigma} + \alpha\sigma\right) - \Phi\left(\frac{\mu-s}{\sigma} + \alpha\sigma\right)} \\
 &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu-\alpha\sigma^2)^2\right)}{\Phi\left(\frac{s-\mu}{\sigma} - \alpha\sigma\right) + \Phi\left(\frac{\mu}{\sigma} + \alpha\sigma\right) - 1} \\
 &= \frac{\frac{1}{b} \phi\left(\frac{x+a-s}{b}\right)}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s-a}{b}\right) - 1}, \text{ where } a = s - \mu - \alpha\sigma^2, b = \sigma
 \end{aligned}$$

And:

$$\begin{aligned}
 E(X | S = s) &= \frac{1}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s-a}{b}\right) - 1} \int_0^s \frac{x}{b} \phi\left(\frac{x-a}{b}\right) dx \\
 &= \frac{1}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s-a}{b}\right) - 1} \int_{\frac{a}{b}}^{\frac{s-a}{b}} (bz + a) \phi(z) dz, \\
 \text{where } z &= \frac{x-a}{b} = a + b \frac{\phi\left(\frac{a}{b}\right) - \phi\left(\frac{s-a}{b}\right)}{\Phi\left(\frac{a}{b}\right) + \Phi\left(\frac{s-a}{b}\right) - 1}
 \end{aligned}$$