Operational Risk Quantification
Mathematical Solutions for Analyzing Loss Data

Gene Álvarez, Ph.D.
May 2001

Table of Contents

Introduction 1
1 Severity Distribution for Historical Loss Data 1
  1.1 Definition of a Tail-Adjusted Lognormal Distribution 1
  1.2 Tail-Adjusted Lognormal Distribution Determination from Truncated Data 2
  1.3 Which to Use: Lognormal or Tail-Adjusted Lognormal? 5
2 Capital Risk Exposure Calculation 6
  2.1 TALN and VaR 7
  2.2 Which to Use: Lognormal or Tail-Adjusted Lognormal? Part II 8
  2.3 Observed Mean versus Theoretical Mean 8
3 TALN versus Weibull 9
4 Frequency Distribution: How to Quantify the Occurrences of Loss Events 11
5 Event Classification 12
  5.1 When Data is Poorly Organized 12
  5.2 Loss Data Classification Scheme 15
Conclusion 16
References 18

"When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind." William Thompson, Lord Kelvin

Introduction

The daily operations of running and supporting a business are not without risk. Every organization is exposed to human errors, external failures, and technology breakdowns, to name just a few potential problems. The classification of these issues is called *operational risk*. The Basel Committee on Banking Supervision has defined operational risk to be *the risk of direct or indirect loss resulting from inadequate or failed internal processes, people, and systems, or from external events*. Every institution is exposed to risks that fit this definition and as such it is logical that every institution must focus on operational risk management. The ideal firm is proactive in operational risk management: it monitors its performance by periodically assessing its personnel and departments; keeps a record of loss events that have occurred; and analyzes these losses to determine what lessons may be learned and to develop strategies for minimizing or preventing future problems. As with all management responsibilities, there are several challenges that must be overcome before any institution will conform to this ideal. This article will address the issue of analyzing and quantifying loss data.

The primary objectives of quantifying operational risk are—first—to calculate the exposure of a business line, department, or firm and—second—to facilitate the process of deciding how to minimize, control, or mitigate risk. Before any calculation can be performed, though, one has to address a major obstacle: understanding the fundamental nature of the collected loss data. In other words, how can one mathematically describe the shape of a given set of loss data? Knowing the functional form of a data set allows for great accuracy in all calculations that derive from the data. In the coming sections, we will present a probability density function that accurately describe the monetary loss data distribution—or, a "severity distribution", and then apply the function to calculate potential risk exposures. We will also discuss the choice of the frequency probability function to describe the occurrence of losses, and the impact of this choice on operational risk quantification. And finally, a proposed operational risk loss event classification scheme is presented in order to promote a logical structure to categorize loss events that removes most of the uncertainty in the analysis of loss data.

1 Severity Distribution for Historical Loss Data

1.1 Definition of a Tail-Adjusted Lognormal Distribution

Generally, analysts assume that the severity distribution of operational risk monetary losses is best expressed by the Lognormal function, a Normal distribution whose variable x is logarithmically transformed ($x' = \text{Log}(x)$). Even though the body of the monetary loss distribution may be accurately described by a Lognormal curve, the tails of the empirical data are not precisely expressed by this function. Therefore, other statistical functions, such as Pareto or Extreme Value Distributions, are used to describe the tail, and the intrinsic limitation that they will not accurately represent the body is accepted. Analysts do this because catastrophic loss calculations are sensitive to the tail of the distribution. Figure 1 illustrates the distinction between the body and tail regions of monetary loss data. Rather than accept the aforementioned limitation, we will describe the entire empirical data set by employing a mixture of two well known functions: the Lognormal and Gamma functions. In mathematical literature, such mixtures are called *compound normal distributions* [1].

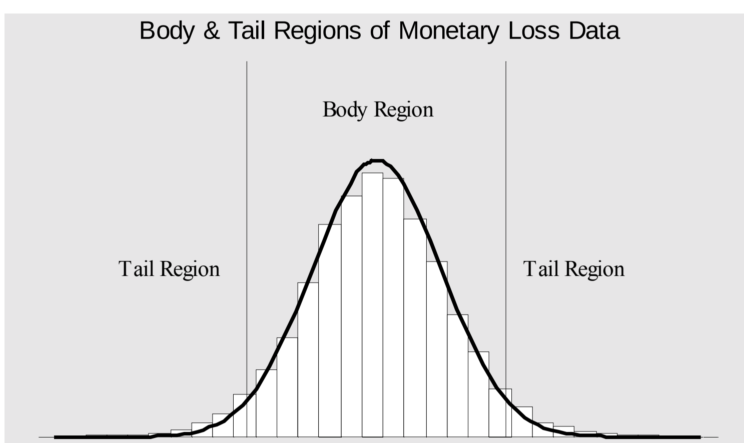


Figure 1: Distinction between the body and tail regions of monetary loss data when represented by a "normal" distribution.

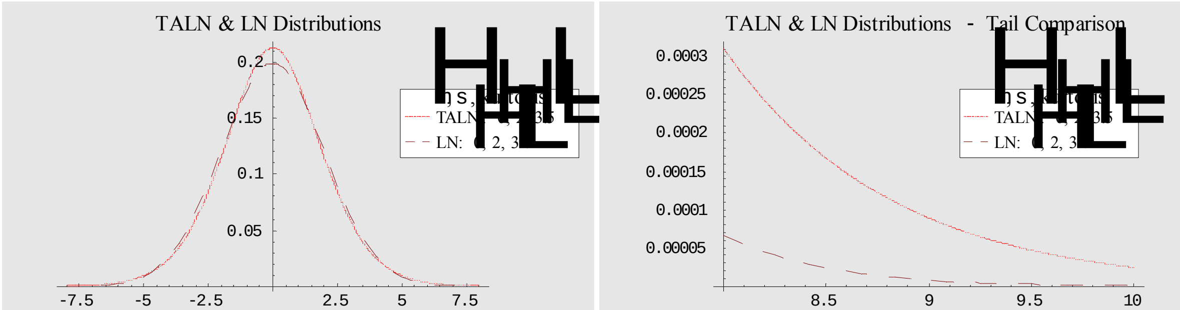


Figure 2: Contrast between Tail-Adjusted Lognormal and Lognormal distributions. The mean and standard deviation for each are 0 and 2, while their respective kurtosis is 3.5 and 3 (definition of a normal distribution).

The motivation to use the Lognormal-Gamma mixture is simple. Because the two Lognormal parameters (mean μ and standard deviation σ) reasonably depict the body of the loss distribution, a third parameter—kurtosis, or "the fourth moment"—must be introduced to satisfactorily depict the tail. Parameterizing kurtosis via a Gamma distribution will result in the desired functional form. By definition, the kurtosis of a normal distribution is three; using the Lognormal-Gamma mixture, however, will allow the kurtosis parameter to be greater than or less than three. From this point forward, we will refer to the mixed distribution as a Tail-Adjusted Lognormal (TALN) density function. A Lognormal-Gamma mixture allows an operational risk loss severity distribution to be parametrically represented by a function with three parameters: mean, standard deviation, and kurtosis.

1.2 Tail-Adjusted Lognormal Distribution Determination from Truncated Data

The determination of the three parameters that depict the TALN distribution for a given set of monetary loss data is critical. One must also keep in mind that operational risk loss data is almost always collected above a truncation threshold because it is impractical to gather data from zero: to collect losses as small as a few cents, for example, is an impossible task. Therefore, for any given data set, the measurable quantities—μ, σ, and kurtosis—represent a truncated (conditional) distribution and not a full (unconditional) distribution. Despite this apparent limitation, all is not hopeless. One can calculate the parameters of the full distribution from the truncated parameters. The procedure is as follows:

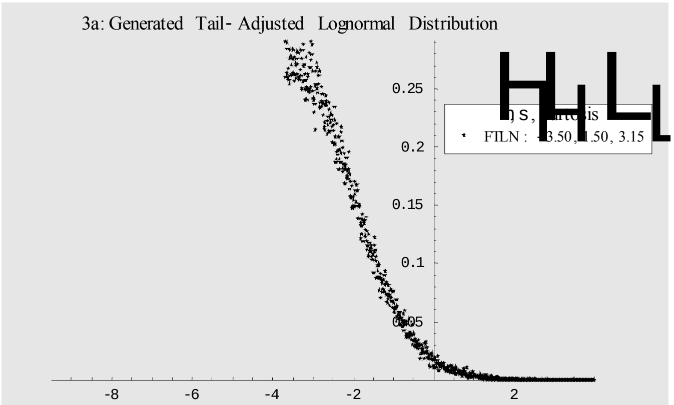
a) Derive the first, second, and fourth moments of the truncated TALN probability distribution function[*]
b) Calculate the empirical data truncated mean, standard deviation, and kurtosis
c) Equate each truncated moment expression to its respective truncated measured quantity, *i.e.* the truncated mean is equated to the truncated first moment equation
d) Solve the three simultaneous equations.

The solution set to the three truncated simultaneous equations will yield the three unconditional parameters.

To illustrate this point, consider two TALN distributions with parameters (μ, σ, kurtosis) = (-3.50, 1.50, 3.15) and (-4.0, 1.8, 3.8). For each distribution, one hundred thousand points were randomly generated above a truncation threshold of twenty-five thousand dollars (see Figure 3). The truncated μ, σ, and kurtosis parameters were calculated and the simultaneous truncated moment equations were solved after each expression was set equal to its respective measured truncated value. As one can readily see from Table 1 and Figure 4, the values obtained from solving the TALN simultaneous equations are in good agreement with the starting parameters. The difference is attributable to numerical rounding in solving the equations.

Figure 4 also plots a third distribution: the Lognormal distribution if one solved the truncated Lognormal first and second moment equations that were equated to their respective measured quantity. Figures 4b and 4d illustrate that, even if the kurtosis turns out to be relatively small (*e.g.*, 3.15), ignoring kurtosis will grossly misrepresent the parameterization of the tail and will, therefore, underestimate any calculation that is tail-dependent, such as VaR.

[*] One needs to go through the derivation only once. Retain the expressions for all future analyses.



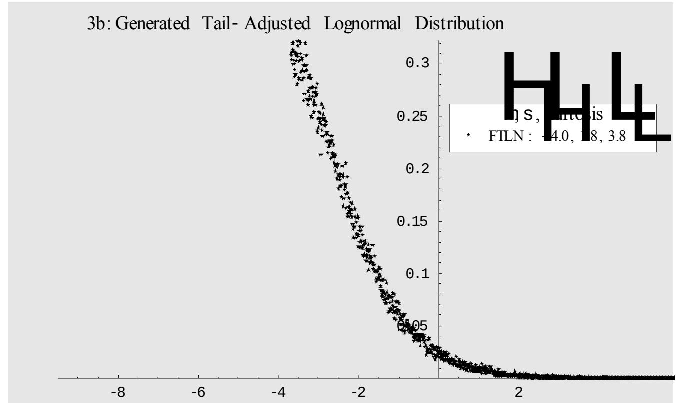


Figure 3: a) Histogram of the randomly generated losses for the (μ, σ, kurtosis) = (-3.50, 1.50, 3.15) TALN density function. Losses were binned in monetary buckets in units of $10,000.
b) Histogram of the randomly generated losses for the (μ, σ, kurtosis) = (-4.0, 1.8, 3.8) TALN density function. Losses were binned in monetary buckets in units of $10,000.

TALN Distribution Parameter Values

	Given		Truncated		Solved	
μ	-3.50	-4.00	-2.43	-2.39	-3.49	-4.03
σ	1.50	1.80	0.95	1.10	1.49	1.81
kurtosis	3.15	3.80	4.06	5.86	3.16	3.69

Table 1: Comparison of known conditional parameters with parameters obtained from solving simultaneous truncated moment equations.

Another way to determine the unconditional parameters from the truncated data set is to employ *the principle of least squares* [2]. In general, one must minimize a set of *m*×*n* simultaneous equations, where *m* is the number of variables involved in the minimization procedure and *n* is the number of data points used for the minimization procedure. Please note, the minimization of *m*×*n* equations can be numerically challenging. To demonstrate the implementation of the principle of least squares to arrive the full parameters for a given TALN distribution, we used the same randomly generated data set for the (-4.0, 1.8, 3.8) TALN density as in our previous example.

We chose to minimize the TALN cumulative distribution function (cdf) and used eleven data points corresponding to 10 percentiles, ranging from 5 to 95% in units of 10%, and the constraint that the function must include the origin. Naturally, the objective function that was minimized accounted for the truncation threshold of the randomly generated data set. In fairness, we ran the optimization program three times: once without weight and twice with different weights. The weights used, to remove subjectivity, were $(1-y_i)^{-1}$ and $n_i\left(\sum_{i=1}^{k} n_i\right)^{-1}$, where (x_i, y_i) is the i^{th} cdf point for a given momentary amount and its corresponding percentile, and n_i is the number of losses for the i^{th} point. The results from the three fits are listed in Table 2 and the parameters of the two weighted results are illustrated in Figure 5. Even with a casual glance at either Table 2 or Figure 5, one can surmise that the three 3-parameter sets obtained via the principle of least squares are not in good agreement with the original parameters. One may argue that 11 points are not a sufficient number of points to use in an optimization method; however, one must keep in mind that the number of equations to be optimized is $3n^{\dagger}$. Hence, the computational complexity increases geometrically. In contrast, our first discussed method required solving three simultaneous equations to determine the three parameters of the full TALN function to represent the empirical data—far more simple, efficient, and elegant.

† For every data point, one must optimize three equations for each unknown parameter: μ, σ, and kurtosis.

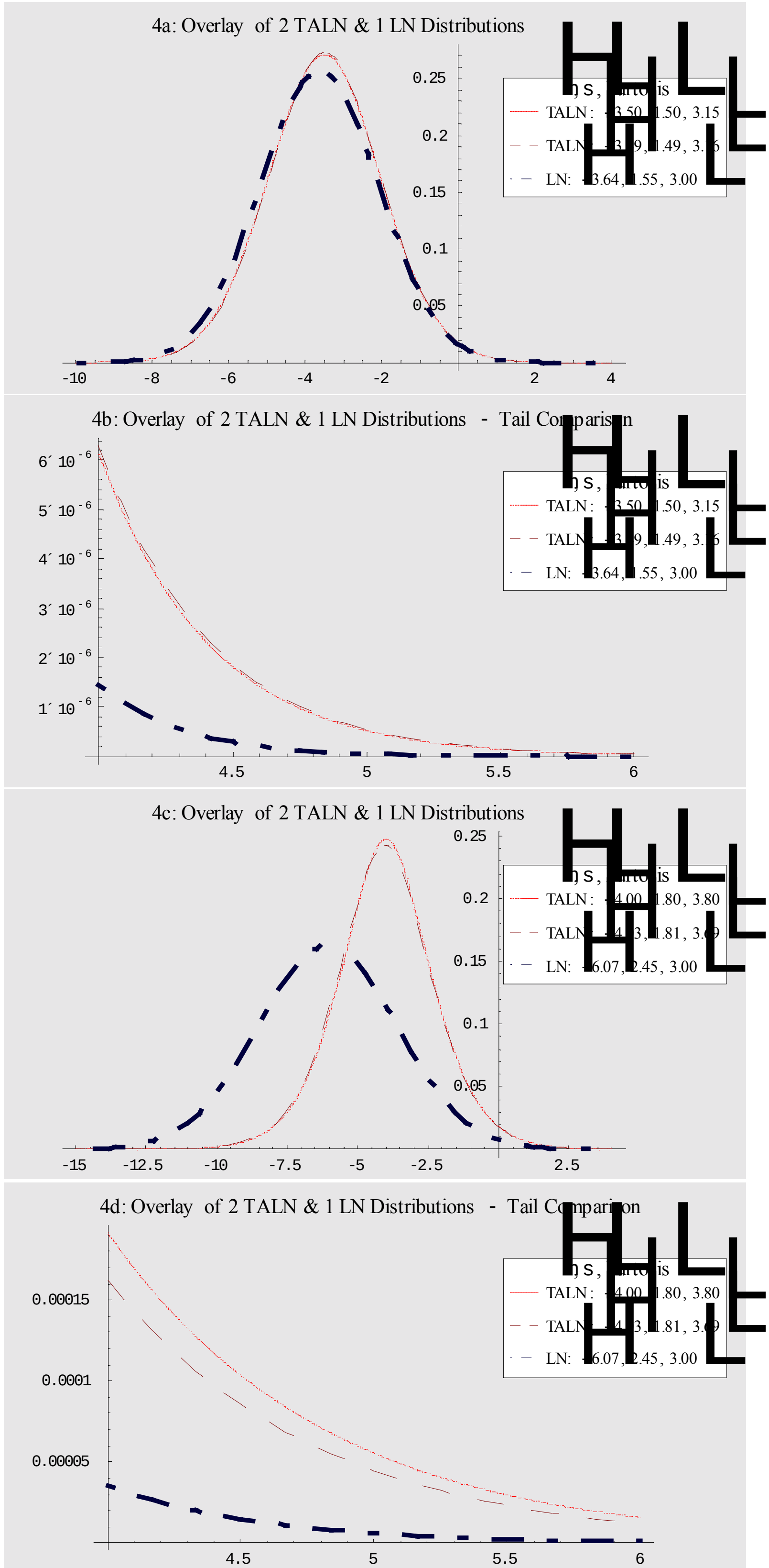


Figure 4: Comparison between TALN and LN distributions whose parameters were obtained from solving simultaneous truncated moment equations, and the original TALN distribution (solid curve).

TALN Distribution Parameter Values

	Given	Fitted Parameter Values		
		No Weight	Weight: $(1 - y_i)^{-1}$	Weight: $n_i\left(\sum_{i=1}^{k} n_i\right)^{-1}$
μ	-4.00	-2.45	-2.43	-2.09
σ	1.80	0.72	0.42	1.45
kurtosis	3.80	3.03	3.00	3.33

Table 2: Comparison of known conditional parameters with parameters obtained via the Principle of Least Squares.

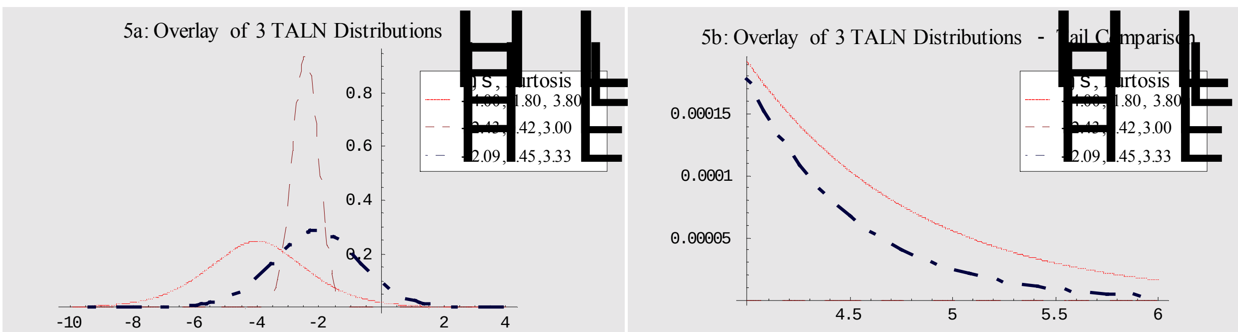


Figure 5: Comparison between the two TALN distributions whose parameters were obtained via the Principle of Least Squares and the original TALN distribution (solid curve).

One may also use the Adaptive Grid Refinement (AGR) algorithm [3] to determine the unconditional parameters. AGR is an *n*-dimensional adaptation of Newton's method related to adaptive mesh generation. Even though AGR will find all possible solutions over the defined region of possible solutions, the limitations of the algorithm are: i) it does not allow constraints; and ii) it does not define the optimal solution region. The second limitation requires a user to define a sufficiently large grid region to capture all possible solutions, if such a region can be defined. For example, the solution space for the μ must be large enough to anticipate any potential solution. Otherwise, the results found in the smaller solution space for the mean may not be the true optimal answers. One will have similar concerns for the other two parameters: σ and kurtosis. Hence, one must define a three dimensional grid, with granular lattice spacing, sufficiently large enough to accommodate any possible solution combination for the three parameters. Clearly, it is more desirable to solve the TALN truncated moment equations to obtain the exact solutions for the full distribution parameters than to deal with the limitations and complications of the AGR algorithm.

1.3 Which to Use: Lognormal or Tail-Adjusted Lognormal?

Given a truncated data set, one is able to calculate the truncated mean, standard deviation, and kurtosis parameters. Then what? How does one choose between the Lognormal and TALN density functions to obtain the full distribution that represents the data? If the truncated kurtosis is extremely large—*e.g.* greater than six—then one should choose the TALN density. But in other cases, the measured truncated kurtosis will not be such an obvious indicator. The solution to this dilemma is straightforward: solve the truncated moment equations for both the LN and TALN distributions, then perform an Anderson-Darling goodness-of-fit test to resolve the indecisiveness.

Let us consider an example to demonstrate this point. We begin with a known LN distribution with μ = -3.5 and σ = 1.5 and generate two-hundred-fifty thousand random losses above the truncation point of $25,000 (see Figure 6a). The truncated parameters calculated for the generated data set are (μ, σ, kurtosis) = (-2.42, 0.94, 3.71). For the LN and TALN densities, we solve their respective simultaneous set of truncated moment equations. The solution for the full TALN distribution is (μ, σ, kurtosis) = (-3.37, 1.46, 3.04), while for the full LN distribution we obtain (μ, σ) = (-3.51, 1.51). Figures 6b and 6c depict the resultant curves.

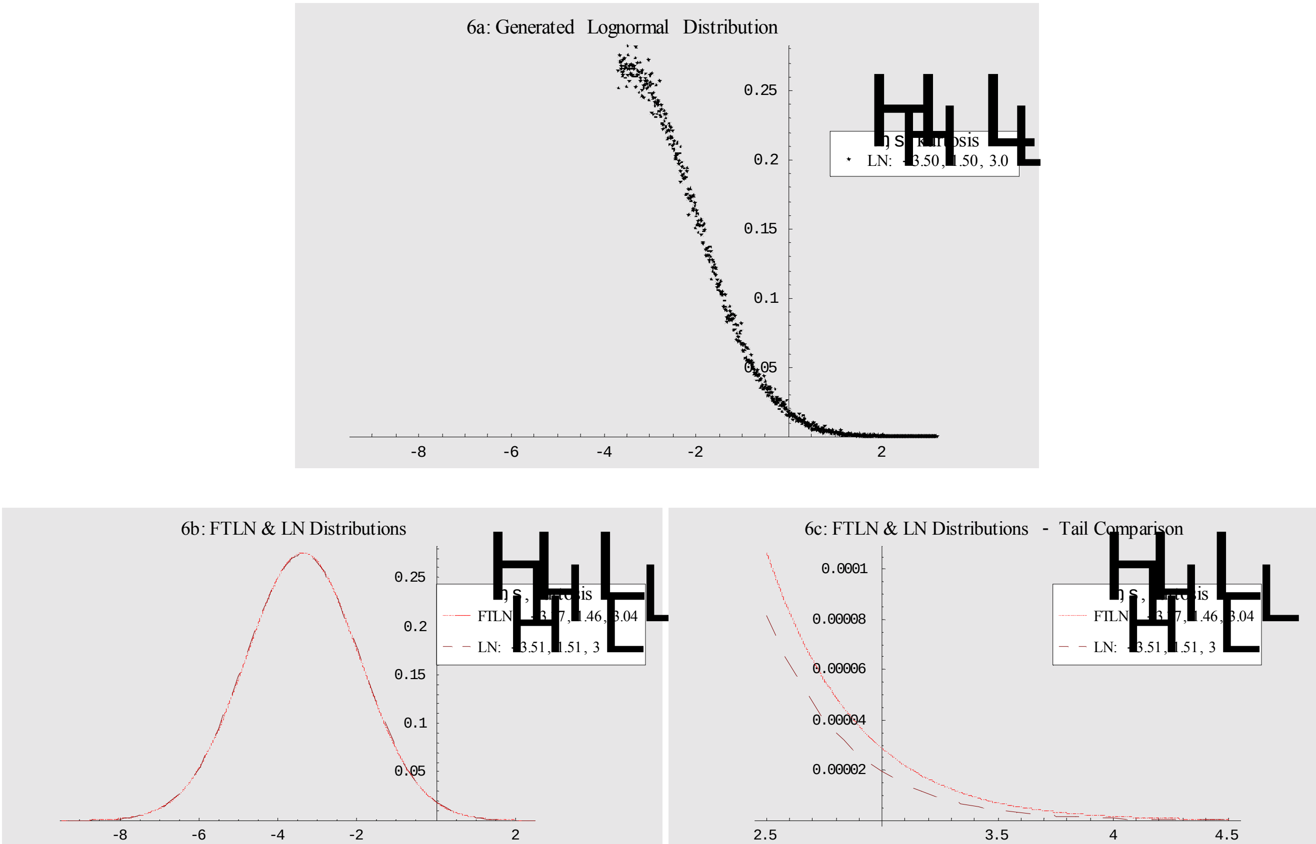


Figure 6: a) Histogram of the randomly generated losses for the (μ, σ) = (-3.50, 1.50) LN density function. Losses were grouped in monetary buckets in units of $10,000.
b,c) Comparison between TALN and LN distributions that describe the generated data as shown in Fig. 6a.

To choose between the two densities, one may apply either the Chi-Square or Kolmogorov-Smirnov goodness-of-fit test. However, the Anderson-Darling test should be employed instead, because it gives more weight to the tail of the distribution. This is an important criterion since the slope of the tail reflects the sensitivity of the distribution. When we calculated the AD-test value for each of the resultant distributions, the TALN value was approximately two percent smaller than the LN value. In this example, given that we started out with a known LN distribution, the TALN turned out to be the better fit. The minute discrepancy in the quality of fit is due to numerical round-off errors in solving the respective sets of truncated equations. Thus, in general, the TALN density function will adequately describe the severity distribution of the monetary loss data, even when a LN distribution may be used. When we discuss VaR calculations in section 2.2, we will return to this example.

2 Capital Risk Exposure Calculation

In the analysis of data, a closed, analytical function is advantageous because it produces a continuous depiction of the underlying data set. The analytical function can be evaluated to obtain values of the function or of its inverse (if one exists), for various numerical calculations. In this section, we will demonstrate the application of the TALN probability density function in capital risk exposure calculations, and discuss its advantages.

Before we proceed, we must note that, although empirical data can be grouped in discrete monetary buckets to calculate values such as probabilities, this practice must be used with caution. If we limit our discussion to probability calculations—unless the data is smooth throughout the entire data capture range, and is of statistical significance to estimate probabilities with small errors—we will underestimate or overestimate the probabilities and produce inaccurate values. And, if one needs a probability value beyond the captured monetary range, interpolation of the data definitely will result in erroneous numbers. One of the virtues of a parameterized functional form is that these shortcomings can be avoided.

2.1 TALN and VaR

There are several methods that can be used to calculate an institution's capital risk exposure. For the upcoming discussion, we will limit ourselves to Monte Carlo simulation. A Monte Carlo model is a statistical simulation that utilizes sequences of random numbers to perform statistical sampling experiments on a computer. The power of the Monte Carlo technique is that it can be applied to problems with no probabilistic content, as well as to those with inherent probabilistic structure [4]. For our discussion on capital risk exposure, or VaR (Value-at-Risk), we will compare and discuss VaR charges at two confidence levels for the two TALN distributions we have used thus far, as listed in Table 3. Our Monte Carlo model simulated 100,000 years of losses, assuming a Poisson mean frequency of 15 loss events per year above $100,000[‡].

In Table 3, we see that the three methods to determine the unconditional parameters—TALN truncated equations, LN truncated equations, and the principle of least squares—produce similar values at the 95.5% confidence level. This consistency does not hold at the 99.9% confidence level: there, we see a significant disparity between the values calculated from the known distributions, and those evaluated using the parameters obtained from the truncated LN moment equations and the principle of least squares. The difference is especially apparent when one examines the values for the known TALN density: $\mu = -4.0$, $\sigma = 1.8$, and kurtosis = 3.8.

Only the VaR values calculated from the parameters of the TALN truncated equations are in solid agreement with the known distributions. Because VaR calculations are sensitive to the tail of the distribution, it is extremely

Monte Carlo Simulated VaR Values [†]
(Values are in Millions of Dollars)

Distribution Parameters
(μ, σ, kurtosis)

Given Distribution		**TALN Truncated Eqns.**		**LN Truncated Eqns.**		**Principle of Least Squares** *	
(-3.50, 1.50, 3.15)		(-3.49, 1.49, 3.16)		(-3.64, 1.55, 3.00)		(-2.21, 1.17, 3.17)	
95.5% CL	99.9% CL	95.5% CL	99.9% CL	95.5% CL	99.9% CL	95.5% CL	99.9% CL
9.96	31.83	10.26	30.13	9.66	22.24	10.18	22.47
Mean: 5.18 MM		Mean: 5.25 MM		Mean: 5.09 MM		Mean: 5.57 MM	
(-4.00, 1.80, 3.80)		(-4.03, 1.81, 3.69)		(-6.07, 2.45, 3.00)		(-2.09, 1.45, 3.33)	
95.5% CL	99.9% CL	95.5% CL	99.9% CL	95.5% CL	99.9% CL	95.5% CL	99.9% CL
29.52	425.07	27.78	435.19	21.90	113.40	20.87	93.87
Mean: 11.77 MM		Mean: 11.48 MM		Mean: 9.07 MM		Mean: 9.52 MM	

Table 3: VaR comparisons for two known TALN distributions (first column) with respect to calculated values obtained from three different approaches to arrive at the known conditional TALN density functions. The method of solving the three simultaneous truncated moment equations gives the best agreement with the known distributions (second column).

[†] Simulated 100,000 years using a Poisson mean frequency of 15 loss events per year above 0.1MM.

[‡] Throughout the article, whenever we discuss results from a Monte Carlo simulation we will always use these simulation conditions, unless otherwise stated.

[*] Corresponding to the parameters obtained using $n_i \left(\sum_{i=1}^{k} n_i \right)^{-1}$ as weights.

important to correctly parameterize the tail to avoid under- or over-estimated VaR values. In practice, with an accurate functional representation of the data set in question, capital risk exposure calculations can be carried out to three standard deviations (~ 99.73% confidence level[§]) with mathematical certainty.

2.2 Which to Use: Lognormal or Tail-Adjusted Lognormal? Part II

In section 1.3 we discussed, given a truncated data set, which of the two density distributions—TALN or LN—should be used. We demonstrated, at least via the Anderson-Darling test, that the two functional forms were adequate mathematical representations for the example we used. For thoroughness, we will compare the VaR values for the known distribution, to the values for the solved TALN and LN parameters. The results of this comparison are given in Table 4. As in our previous VaR calculation exercise, we simulated 100,000 years of losses, assuming a Poisson mean frequency of 15 loss events per year above $100,000. As one can see from Table 4, the choice of probability density function does not significantly affect the VaR results in comparison with the VaR values of the known LN density function, so the choice of a TALN density did not skew any VaR calculation.

VaR Comparisons
(Values are in Millions of Dollars)

Distribution Parameters	Mean	95.5% CL	99.9% CL
μ = -3.50, σ = 1.50	4.971	9.293	21.508
μ = -3.51, σ = 1.51	4.982	9.212	20.694
μ = -3.37, σ = 1.46, kurtosis = 3.04	4.974	9.372	20.346

Table 4: Comparison of VaR values of a known LN distribution (first row) against those calculated from solving the LN and TALN truncated moment equations. Two hundred fifty thousand losses above $25,000 were randomly generated for the known distribution. The truncated parameters were (μ = -2.42, σ = 0.94, kurtosis = 3.71). The VaR values are a result of a simulation of 100,000 years, with a Poisson mean frequency of 15 loss events per year above 0.1MM.

2.3 Observed Mean versus Theoretical Mean

Let us now focus our attention on the importance of distinguishing between the *theoretical mean* and the *observed mean*. The theoretical mean is simply the mean value of a density distribution (the first moment), while the observed mean is the mean of a truncated data set. As we have noted, operational risk loss data is invariably truncated. Hence, the observed mean does not equal the theoretical mean: it may be below or above the theoretical mean depending on the truncation threshold. As a result, any calculation that relies on the observed mean, as proxy for the theoretical mean, will give misleading results.

One VaR proposal under consideration would apply a multiplicative factor to the three-year average of the mean aggregate loss. Assuming an institution has a good procedure to capture data on loss events and their monetary impact, the truncated mean aggregate loss for a given year may be calculated fairly accurately. One the other hand, results will not be accurate for the mean of the underlying full severity distribution, which describes the collected loss data. To quantify this assertion, we simulated 1000 years for two known TALN distributions to calculate the two and three year average of the mean aggregate loss (observed mean) for two different Poisson mean frequencies above the $25,000 threshold. The frequency of 13 events per year could represent a very good control environment, while 40 events per year could represent an average control environment. Table 5 summarizes the results. All the calculations produced outcomes in which 90% of the observed means were above the theoretical mean and, in some cases, this percentage was as high as 100%. Hence, anyone using the observed mean as a substitute for the

[§] For simplicity, the author prefers the 99.9% confidence level.

theoretical mean in any monetary calculation must keep in mind that this replacement may grossly overestimate the actual amount.

Observed Mean vs. Theoretical Mean **Observed Mean Percentage Below / Above Theoretical Mean**				
	TALN Distribution Parameters			
	μ = -3.50, σ = 1.50, kurtosis = 3.15		μ = -4.00, σ = 1.80, kurtosis = 3.80	
Poisson Frequency	2 Year Average	3 Year Average	2 Year Average	3 Year Average
13	8.2% / 91.8%	2.1% / 97.9%	6.0% / 94.0%	2.4% / 97.6%
40	0.2% / 99.8%	0% / 100%	0% / 100%	0% / 100%

Table 5: Monte Carlo simulation results that calculate the percentage of time the observed mean is below or above the theoretical mean for two distinct TALN distributions. For each Monte Carlo run, 1000 years were simulated using Poisson mean frequencies of 13 and 40 to generate the *n* random loss events per year above $25K.

Furthermore, using only the mean does not take into account all that is needed to accurately describe the severity distribution—particularly the risk profile of an institution. One cannot neglect the volatility, as measured by the standard deviation, and the "heaviness" of the tail, as measured by kurtosis, because—as we have seen in various numerical examples—using the three parameters together enables a risk analyst to perform calculations with confidence and mathematical certainty.

3 TALN versus Weibull

Because of its simple functional form and ease of use, the Weibull distribution is occasionally used to describe monetary loss data. But how reliable is the Weibull distribution in depicting loss data? It is important that the function we choose to describe the monetary distribution of loss events gives an accurate representation of the underlying data set. Otherwise, all results will be questionable. In this section, we will discuss the Weibull density's accuracy in representing monetary loss data and the calculated VaR values. In all of our work we will employ a Weibull distribution with three parameters: location (ξ), scale (*c*), and shape (α).

As we have done thus far, let us consider an example. We will begin with a TALN distribution with parameters (μ, σ, kurtosis) = (-4.0, 1.8, 3.8), and a Weibull distribution with parameters (ξ, *c*, α) = (-12.0, 4.5, 8.8). For the two given densities, we will randomly generate 100,000 losses without any truncation threshold (see Figures 7a and 7b). As illustrated in Figure 7c, we will combine these two data sets to create one with 200,000 losses. Now, we will obtain the TALN and Weibull parameters that best represent the merged data set. To arrive at the respective set of parameters, we will use the principle of least squares using $n_i\left(\sum_{i=1}^{k} n_i\right)^{-1}$ as weights**. Figures 7d and 7e illustrate the results of the *fits*. As one can see, the Weibull distribution does not represent the data set as well as the TALN distribution. To quantify what the eye "knows", the Anderson-Darling goodness-of-fit test value for the TALN is approximately 60% lower than the Weibull value. Therefore, one can confidently conclude that the TALN distribution better describes the mixed losses.

** We used 100 points in the fit, corresponding to the hundred buckets used to group the data as shown in Figure 7.

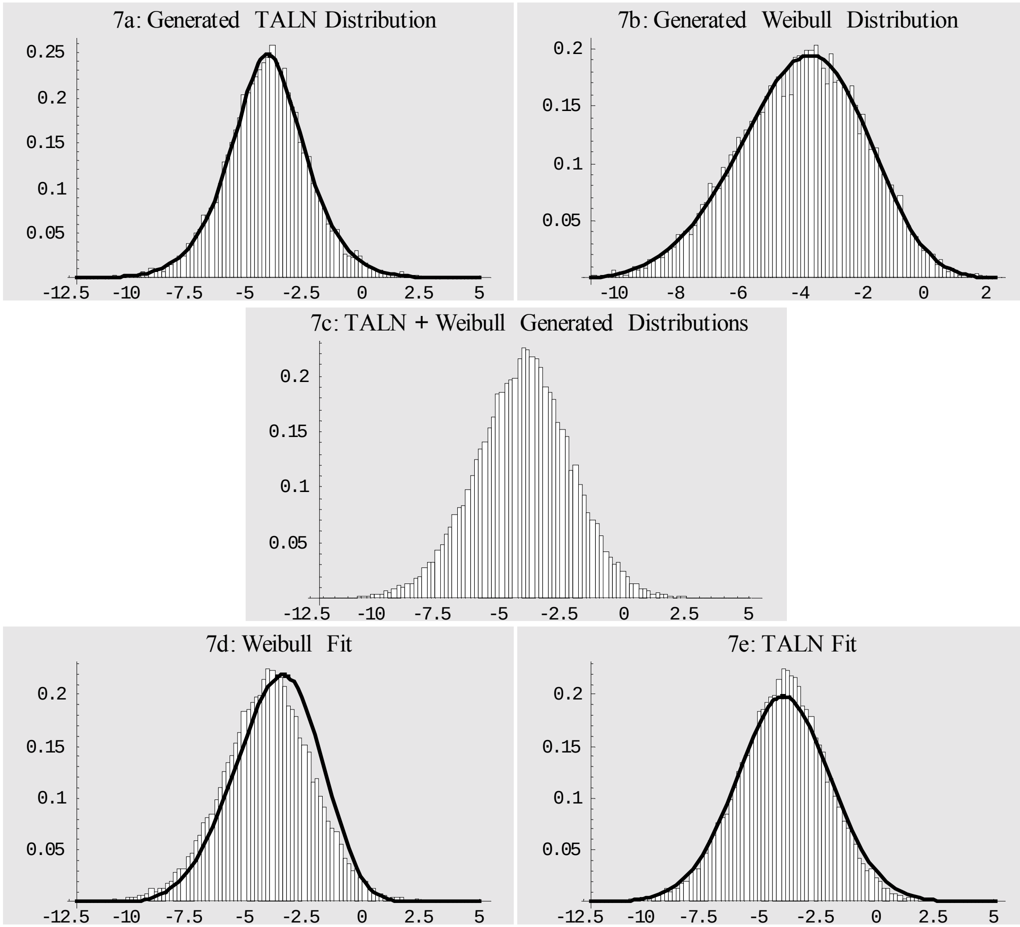


Figure 7: a) Histogram of the randomly generated losses for the (μ, σ, kurtosis) = (-4.0, 1.8, 3.8) TALN density function, as shown by the solid curve.
b) Histogram of the randomly generated losses for the (ξ, c, α) = (-12.0, 4.5, 8.8) Weibull density function, as shown by the solid curve.
c) Addition of histogram 7a plus 7b.
d) Histogram created when fitted Weibull distribution is overlaid on Fig. 7c. The Weibull distribution shown in Fig. 7d was obtained via the principle of least squares using $n_i \left(\sum_{i=1}^{k} n_i \right)^{-1}$ as weights.
e) Histogram created when fitted TALN distribution is overlaid on Fig. 7c. The TALN distribution shown in Fig. 7e was obtained via the principle of least squares using $n_i \left(\sum_{i=1}^{k} n_i \right)^{-1}$ as weights.

Continuing with our example, we will calculate the VaR values[††] for the original distributions and the two fitted distributions. At the 99.9% confidence level, one might expect the VaR quantity for the combined data set to be the same order of magnitude as the average of the individual 99.9% confidence level values for the original data sets. One may propose this "guesstimate" since the losses in the tail of the original TALN distribution will influence the calculation, even though the width of the combined data set is wider than the two original distributions. As shown in Table 6, the fitted Weibull distribution grossly underestimates the VaR result, while the fitted TALN distribution agrees with our educated guess.

Qualitatively, a risk manager can conclude from this example that if a business has experienced large monetary losses, the Weibull distribution will not accurately depict the severity distribution, and that VaR calculations based

[††] Monte Carlo simulations of 100,000 years of losses assuming a Poisson mean frequency of 15 loss events per year above $100,000.

on the Weibull density will result in imprecise values. As demonstrated, the TALN distribution will describe the tail of the monetary loss distribution with better accuracy and higher precision.

VaR Comparisons
(Values are in Millions of Dollars)

		VaR Values for each Randomly Generated 10,000 Losses		VaR Values for the Combined 20,000 Losses for each Fitted Distribution	
		TALN (μ, σ, kurtosis) = (-4.0, 1.8, 3.8)	Weibull (ξ, c, α) = (-12.0, 4.5, 8.8)	TALN (μ, σ, kurtosis) = (-3.97, 2.06, 3.17)	Weibull (ξ, c, α) = (-11.12, 4.74, 8.15)
VaRs:	Mean	11.77	6.40	11.93	5.33
	95.5% CL	29.52	12.39	29.94	9.63
	99.9% CL	425.07	27.13	237.43	17.32

Table 6: VaR comparison between the original data sets as generated for the given distribution (first two columns) and the combined data set as fitted by the TALN and Weibull distributions (last two columns). The VaR values are a result of a simulation of 100,000 years, with a Poisson mean frequency of 15 loss events per year above 0.1MM.

4 Frequency Distribution: How to Quantify the Occurrences of Loss Events

Identifying the severity (monetary) distribution of loss event data with the best descriptive mathematical functional form is vital. But the analysis is not complete until one also has determined how to quantify the occurrence of loss events. The latter is equally important because the frequency distribution directly reflects the number of loss events one might expect to experience over a certain time horizon. For our investigation, we will consider three related discrete distributions to portray the frequency of loss events.

The three discrete distributions we will use are the binomial, negative binomial, and the Poisson distributions. All three satisfy the same recursive relationship and are described by one parameter or two [5]. The distinction among the three distributions is the relationship between the mean and the variance. For the binomial distribution, the variance is less than the mean; for the negative binomial distribution, the variance is greater then the mean; and, for the Poisson distribution, the variance and the mean are equal.

To illustrate the effect of the choice of frequency distribution, we calculate the VaR values implementing Panjer's recursive algorithm [6] for the two TALN distributions: (μ, σ, kurtosis) = (-3.50, 1.50, 3.15) and (-4.0, 1.8, 3.8). Panjer's algorithm was used to avoid any dependency on the random sampling of the monetary interval as it is done in Monte Carlo simulations. For the (-3.50, 1.50, 3.15) density, we assumed the mean frequency for all three frequency distributions was 72 loss events, while the variances were 10, 25, 50, and 75% of the mean value for the binomial and negative binomial distributions. For the other severity distribution in our example, we assumed the mean frequency for all three frequency distributions was 95 loss events and changed the variance as previously stated. The mean frequency represents the total number of loss events an institution may experience: the monetary threshold is zero dollars[‡‡]. A summary of the results is listed in Table 7.

As Table 7 illustrates, the choice of frequency distribution does not greatly influence the calculated VaR values, even though, as expected, the values do increase—but only slightly, as the variance increases. Therefore, any of the three discrete distributions mentioned could be used to describe the occurrence of loss events without great concern on the impact on VaR calculations.

One important observation is that, for the known TALN distribution, the VaR values calculated using Panjer's recursive algorithm and the Monte Carlo simulation given in Table 3 are consistent with each other. As long as frequencies are related via conditional probabilities, the results of VaR calculations will be similar, whether one is calculating from zero dollars or from another monetary truncation point.

[‡‡] The mean frequencies of 72 and 95 loss events per year from the threshold of zero dollars are consisted for each known TALN distribution if one assumes 15 loss events above 0.1MM for both densities, as we did in our Monte Carlo simulations.

VaR Comparisons for Three Frequency Distributions VaR Values are in Millions of Dollars										
(μ, σ, kurtosis) = (-3.50, 1.50, 3.15)		**Binomial** Mean = 72				**Poisson** Mean = 72	**Negative Binomial** Mean = 72			
Variance		7.2	18.0	36.0	54.0	72.0	79.2	90.0	108.0	126.0
VaRs:	Mean	6.51	6.51	6.52	6.52	6.52	6.52	6.52	6.52	6.52
	95.5% CL	11.96	12.00	12.06	12.12	12.18	12.20	12.24	12.30	12.36
	99.9% CL	34.15	34.17	34.21	34.25	34.28	34.30	34.32	34.36	34.40
(μ, σ, kurtosis) = (-4.00, 1.80, 3.80)		**Binomial** Mean = 95				**Poisson** Mean = 95	**Negative Binomial** Mean = 95			
Variance		9.5	23.75	47.50	71.25	95.00	104.50	118.75	142.50	166.25
VaRs:	Mean	14.17	14.18	14.19	14.19	14.19	14.19	14.19	14.19	14.19
	95.5% CL	34.27	34.38	34.56	34.73	34.90	34.96	35.07	35.24	35.42
	99.9% CL	453.72	454.00	454.52	455.05	455.45	455.72	455.98	456.51	457.04

Table 7: Two scenarios demonstrate the influence of frequency distributions on the VaR calculations. The VaR values increase, but only slightly, as the variance increases. The monetary step size of 0.005MM was used for all Panjer's recursive algorithm calculations.

5 Event Classification

5.1 When Data is Poorly Organized

We have now seen the significance of properly characterizing the severity distribution of monetary loss data. With a precise mathematical depiction of the data one is analyzing, calculations become meaningful and precise. However, regardless of the tools at one's fingertips and the mathematical rigor of the tools, if the data under analysis is suspect, then the results are meaningless. That is why good classification of data is vital. It must be done consistently, and it must follow a logical pattern. If the data is categorized haphazardly, any values derived from the data will reflect its irregularity. When we talk about categorizing data, we mean that we are grouping loss events based on their causes. Some classifications schemes advocate grouping events by their effects, but this system does not give us useful information about where losses come from. Such categorization methods lead to logical inconsistencies and yield erroneous calculated values.

Before we proceed, let us understand the distinction between *cause* and *effect* categories. The cause of an event is the action or set of circumstances that led to the event. Effect is the monetary or non-monetary loss that is incurred due to the event. In our discussion, we will not consider non-monetary losses because they are inherently difficult to quantify.

Let us now consider the following example on the importance of not misclassifying loss events. We will simply assume that we have two cause categories, fraud and sales misrepresentation, and two effect categories, regulatory fines and lawsuit settlements. If one followed an *effect* classification scheme, then one could have both fraud and sales misrepresentation loss events in the same class even though the two categories, from a management perspective, are quite distinct.

To quantify the above point, let us take into account four examples. Once again, we will employ our two favorite TALN distributions: (μ, σ, kurtosis) = (-3.50, 1.50, 3.15) and (-4.0, 1.8, 3.8). Assume the first distribution represents fraud related losses, while the second represents losses caused by sales misrepresentation. Given the one hundred thousand randomly generated losses for each distribution, we examine four scenarios. The first two scenarios examine misclassification of losses above a $50,000 threshold. The second pair examines losses with amounts greater than $100,000. For each pair of scenarios, the first example will show the results when sales misrepresentation losses are mixed into the set of fraud events. In the second example, fraud losses will be mixed into the sales misrepresentation losses. Because the four cases are analogous, we will outline the applied procedure for one and state the results for all, as shown in Table 8.

Results from Mixing Losses Between Two Known TALN Distribution

TALN Distribution Parameters
(μ, σ, kurtosis)

		Given		Truncated: Before Mixture	Truncated: After Mixture	Solved	
Scenario 1: A random selection of losses greater than 0.05MM from (-4.00, 1.80, 3.80) are mixed into (-3.50, 1.50, 3.15) VaR Values are in Millions of Dollars							
A		(-4.00, 1.80, 3.80)		(-2.39, 1.10, 5.86)	(-2.53, 1.08, 6.38)	(-5.01, 2.03, 3.91)	
B		(-3.50, 1.50, 3.15)		(-2.43, 0.95, 4.06)	(-2.32, 0.98, 4.78)	(-2.93, 1.33, 3.96)	
		A	B			A	B
VaRs:	Mean	12.27	5.25			23.39	7.11
	95.5% CL	30.44	10.26			45.67	15.09
	99.9% CL	461.38	30.74			1266.71	100.82
Scenario 2: A random selection of losses greater than 0.05MM from (-3.50, 1.50, 3.15) are mixed into (-4.00, 1.80, 3.80) VaR Values are in Millions of Dollars							
C		(-3.50, 1.50, 3.15)		(-2.43, 0.95, 4.06)	(-2.56, 0.94, 4.35)	(-5.05, 1.94, 2.41)	
D		(-4.00, 1.80, 3.80)		(-2.39, 1.10, 5.86)	(-2.31, 1.07, 5.56)	(-3.11, 1.50, 4.27)	
		C	D			C	D
VaRs:	Mean	5.25	12.27			9.74	11.58
	95.5% CL	10.26	30.44			23.50	26.10
	99.9% CL	30.74	461.38			334.93	378.12
Scenario 3: A random selection of losses greater than 0.1MM from (-4.00, 1.80, 3.80) are mixed into (-3.50, 1.50, 3.15) VaR Values are in Millions of Dollars							
E		(-4.00, 1.80, 3.80)		(-2.39, 1.10, 5.86)	(-2.55, 1.02, 7.19)	(-4.65, 1.85, 4.31)	
F		(-3.50, 1.50, 3.15)		(-2.43, 0.95, 4.06)	(-2.30, 1.02, 4.39)	(-3.08, 1.45, 3.56)	
		E	F			E	F
VaRs:	Mean	12.27	5.25			24.04	7.55
	95.5% CL	30.44	10.26			44.02	15.94
	99.9% CL	461.38	30.74			1121.34	84.48
Scenario 4: A random selection of losses greater than 0.1MM from (-3.50, 1.50, 3.15) are mixed into (-4.00, 1.80, 3.80) VaR Values are in Millions of Dollars							
G		(-3.50, 1.50, 3.15)		(-2.43, 0.95, 4.06)	(-2.57, 0.89, 4.88)	(-4.03, 1.54, 3.10)	
H		(-4.00, 1.80, 3.80)		(-2.39, 1.10, 5.86)	(-2.28, 1.11, 5.12)	(-3.38, 1.66, 3.76)	
		G	H			G	H
VaRs:	Mean	5.25	12.27			4.04	12.27
	95.5% CL	10.26	30.44			7.80	27.61
	99.9% CL	30.74	461.38			27.95	273.43

Table 8: Four scenarios demonstrate the effect on VaR values when one randomly mixes the losses between two known TALN severity distributions. The VaR values are a result of a simulation of 100,000 years, with Poisson mean frequencies of 13 and 17 monetary loss events per year above 0.1MM.

Let us begin with the first scenario, which is sales misrepresentation events categorized with fraud. In our example, loss events above $50,000 that belong in the (μ, σ, kurtosis) = (-4.0, 1.80, 3.80) TALN distribution are one-third of the time collected with events from the (μ, σ, kurtosis) = (-3.50, 1.50, 3.15) TALN distribution and the total number of losses remains the same (for us, that is 200,000). As a result, the two distributions are now altered—their parameters differ from what they should actually be. In scenario one, the (-4.0, 1.80, 3.80) distribution, because it lost events above 0.05MM, has been transformed into a severity density with parameters (-5.01, 2.03, 3.91). The (-3.50, 1.50, 3.15) density, because it has gained events above 0.05MM, has been transformed into a TALN density with (-2.93, 1.33, 3.96).

The impact of the transformations—the misclassification of loss events—is profound. As Table 8 illustrates, the calculated VaRs at the 99.9% confidence level for the two "new" TALN density functions are starkly different from the known TALN distributions. The VaR results increased approximately three times from the original calculated results. If one believes this problem can be resolved by taking an appropriate combination of the altered VaR numbers via a correlation analysis to arrive back at the original VaR values, such reasoning is flawed. This conclusion assumes *a priori* knowledge of the original VaR results, which we all know is not possible: if monetary losses and their occurrence could be predicted with certainty, there would be no need for a risk management department.

To exemplify the power and usefulness of the TALN density, we should note that, in scenario two, the altered distribution for the (-3.50, 1.50, 3.15) TALN density is (-5.05, 1.94, 2.41)—a density function with kurtosis less than three. As previously noted, the Lognormal-Gamma mixture allows one to have kurtosis values less than three.

A similar metamorphosis of the severity distribution will occur if one does not properly identify a loss event and its corresponding monetary impact. Let us consider, for example, a case in which a few loss events above $100,000 are not identified as one event, but as several different events, even though they are all classified in the same category. Using the (μ, σ, kurtosis) = (-4.0, 1.80, 3.80) TALN distribution, we randomly chose events with monetary amounts greater than $100,000 and divided these amounts into two or three components. The new truncated parameters are (-2.01, 1.28, 2.42). For the new truncated parameters there was no solution to the truncated TALN moment equations. Consequently, one had to solve the truncated lognormal moment equations. When one solved the simultaneous equations, the parameters for the new data set are (μ, σ) = (-3.83, 2.17). This time, the TALN distribution is transformed into a LN distribution, and the VaR at the 99.9% confidence level was greatly underestimated, as shown in Table 9. The Monte Carlo simulation did take into account the new frequency of loss events above $100,000 that resulted from splitting the losses.

Results from Splitting Losses from a Known TALN Distribution

			TALN Distribution Parameters (μ, σ, kurtosis)		
		Given	**Truncated: Before Split of Losses**	**Truncated: After Split of Losses**	**Solved**
		(-4.00, 1.80, 3.80)	(-2.39, 1.10, 5.86)	(-2.01, 1.28, 2.42)	(-3.83, 2.17, 3.00)
VaRs:	Mean	12.27			10.55
	95.5% CL	30.44			28.30
	99.9% CL	461.38			166.02
Number of Events		100,000			118,562

Table 9: A scenario that demonstrates the effect on VaR values when one randomly splits losses from a known TALN severity distribution. The VaR values are a result of a simulation of 100,000 years, with a Poisson mean frequency of 15 loss events per year above 0.1MM for the TALN density and a Poisson mean frequency of 12 loss events per year above 0.1MM for the LN density.

The two arguments just presented illustrate that, when operational risk managers collect data, they must use a logical and consistent classification scheme, with mutually exclusive categories, and coherent definitions for losses and their corresponding monetary impacts. Quantitative analysis is only as good as the data upon which it is based, and if operational risk calculations are done using an unsound classification scheme, any results will be suspect.

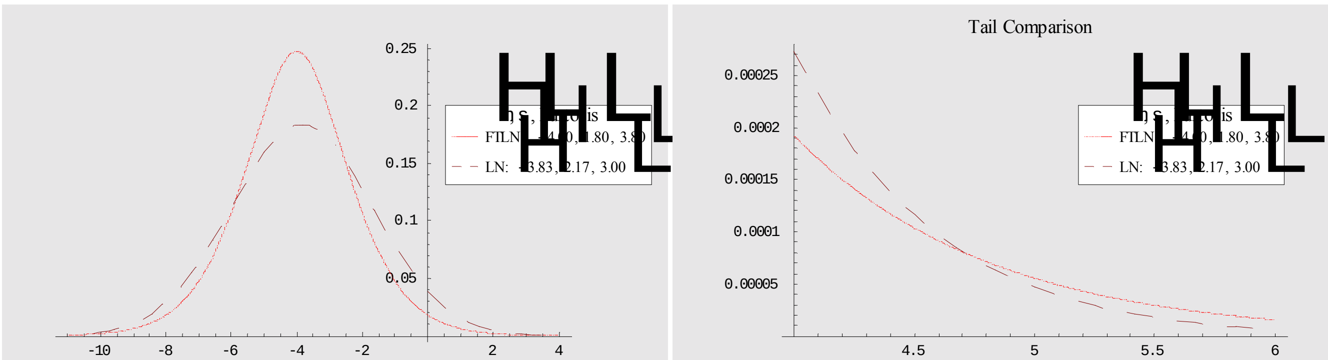


Figure 9: Histograms depict the change to a TALN distribution (solid curve) when losses are spilt (dash curve).

5.2 Loss Data Classification Scheme

It is essential to clearly define an "operational risk loss event" and its corresponding "monetary impact," and to categorize loss events in statistically non-overlapping classification groups. As with any analysis, one should always be mindful of the acronym GIGO—garbage in, garbage out. To keep from producing "garbage," we propose that operational risk loss events be grouped into the following five major categories.

1) Employee: loss events resulting from the actions, or inactions, of a person who works for another in return for stipulated services.

2) Business Process: loss events arising from a firm's execution of business operations.

3) Relationships: loss events caused by the connection or contact that a firm has with clients, third parties, or regulators. This category focuses on the interaction between a firm and other entities; relationship risks involve both parties.

4) Technology: loss events due to a piracy, theft, failure, breakdown, or other disruption in technology, data or information; also includes technology that fails to meet the intended business needs.

5) External: loss events caused by people or entities outside a firm.

The five proposed risk categories apply to a wide spectrum of business activities, and naturally lend themselves to subcategories that are mutually exclusive to each other. For example, a few subgroup members for the Employee class are Human Resources, Errors, and Wrongdoing. Within each subcategory, further pertinent subdivisions can be defined: the Wrongdoing subgroup, for example, could include such branches as Fraud, Disclosure Issues, and Trading Misdeeds.

There are several competing classifications schemes available in the operational risk services market and the scheme that satisfies the logical structure defined above, and possesses a mutually exclusive subcategories for each risk category, is promoted and practiced by Zurich IC Squared (IC^2) [7]. Appendix A outlines IC^2 classification scheme for operational risk loss events. One should keep in mind that the proposed subgroups are not exhaustive. As we gain more experience in gathering and identifying operational risk events, the subgroups for the five major categories may be expanded to reflect our ever-growing knowledge, which is the commitment of IC^2.

When one adheres to the aforementioned categorization format the monetary impact attributable to the operational risk loss event will naturally follow. Splitting the monetary loss into multiple groups will not occur.

Conclusion

As economic globalization becomes a reality and our reliance on technically skilled people, computer technologies, and telecommunications grows, the execution of business becomes more intricate and competitive. With these changes comes an increased exposure to operational risk. Firms have always been vulnerable to problems in the realm of operational risk, but only recently have they begun to classify them as such. Managers face pressure from many sources, both inside and outside their organization, and are expected to raise profits, minimize financial losses, and promote institutional efficiency. To achieve these goals, risk managers must have at their disposal tools or methodologies that help them to comprehend their operational risk exposure. One such tool is mathematical analysis.

In this paper, we discussed the concept that operational risk loss data is inherently truncated, and proposed a continuous, analytical function (the Tail-Adjusted Lognormal) to describe the monetary severity distribution. We then demonstrated the feasibility of obtaining full severity parameters from the truncated mean, standard deviation, and kurtosis measurements. Various examples were presented to test the accuracy of the proposed function by comparing the results to those generated using known severity distributions. We then performed several VaR calculations to examine the precision of the TALN functions with respect to the VaR values of the known distributions. We also compared and discussed the shortcomings of using a Weibull distribution when one has large monetary loss events.

Finally, we presented a classification scheme that categorizes operational risk events by their cause—the action or set of circumstances that led to the event—helping us avoid some of the pitfalls of overlapping data sets, and preventing the inconsistency that arises when events are grouped by the type of loss incurred. If we revisit the Basel Committee's definition of operational risk, the proposed modification will reflect the types of operational risk that the author believes are endemic to all businesses. Operational risk should be defined as *the risk of direct or indirect loss resulting from inadequate or failed internal processes, people, relationships, and technology, or from external events*.

Appendix A: Loss Event Classification Scheme

The outline below is the operational risk classification scheme proposed by Zurich IC Squared [7]. There are five major groups with their associated subgroups listed. In parenthesis, following the subgroups, a few subsets are mentioned as examples. The subgroups and subsets of the subgroups listed here are not a complete hierarchy, but serve as examples. As institutions mature in their operational risk management practice new subgroups and subsets will emerge.

- People Risk: The risk of a loss intentionally or unintentionally caused by an employee—*i.e.* employee error, employee misdeeds—or involving employees, such as in the area of employment disputes. This risk class covers internal organizational problems and losses.
  - Employee Errors (*e.g.* general transaction errors, incorrect routing of transaction, … *etc.*)
  - Human Resource Issues (*e.g.* employee unavailability, hiring/firing, … *etc.*)
  - Personal Injury – Physical Injury (*e.g.* bodily injury, health and safety, … *etc.*)
  - Personal Injury – Non-Physical Injury (*e.g.* libel/defamation/slander, discrimination/harassment, … *etc.*)
  - Wrongful Acts (*e.g.* fraud, trading misdeeds, … *etc.*)

- Process Risk: Risks related to the execution and maintenance of transactions, and the various aspects of running a business, including products and services.
  - Business Process (*e.g.* lack of proper due diligence, inadequate/problematic account reconciliation, … *etc.*)
  - Business Risks (*e.g.* merger risk, new product risk, … *etc.*)
  - Errors and Omissions (*e.g.* inadequate/problematic security, inadequate/problematic quality control, … *etc.*)
  - Specific Liabilities (*e.g.* employee benefits, employer, directors and officers, … *etc.*)

- Relationships: Losses caused to a firm and generated through the relationship or contact that a firm has with its clients, shareholders, third parties, or regulator—*i.e.,* reimbursements to clients, penalties paid, sales practices.
  - Legal/Contractual (*e.g.* securities law violations, legal liabilities, … *etc.*)
  - Negligence (*e.g.* gross negligence, general negligence, … *etc.*)
  - Sales Related Discrimination (*e.g.* lending discrimination, client discrimination, … *etc.*)
  - Sales Related Issues (*e.g.* churning, sales misrepresentation, high pressure sales tactics, … *etc.*)
  - Specific Omissions (*e.g.* failure to pay proper fees, failure to file proper report, … *etc.*)

- Technology: The risk of loss caused by a piracy, theft, failure, breakdown or other disruption in technology, data or information; also includes technology that fails to meet business needs.
  - General Technology Problems (*e.g.* operational error – technology related, unauthorized use/misuse of technology, … *etc.*)
  - Hardware (*e.g.* equipment failure, inadequate/unavailable hardware, … *etc.*)
  - Security (*e.g.* hacking, firewall failure, external disruption, … *etc.*)
  - Software (*e.g.* computer virus, programming bug, … *etc.*)
  - Systems (*e.g.* system failures, system maintenance, … *etc.*)
  - Telecommunications (*e.g.* telephone, fax, … *etc.*)

- External: The risk of loss due to damage to physical property or assets from natural or non-natural causes. This category also includes the risk presented by actions of external parties, such as in the perpetration of fraud, or in the case of regulators, the promulgation of change that would alter the firm's ability to continue operating in certain markets.
  - Disasters (*e.g.* natural disasters, non-natural disasters, … *etc.*)
  - External Misdeeds (*e.g.* external fraud, external money laundering, … *etc.*)
  - Litigation/Regulation (*e.g.* capital control, regulatory change, legal change, … *etc.*)

References

1. Johnson, Kotz, and Balakrishnan · Continuous Univariate Distributions Volume 1, 2nd edition
2. F. B. Hildebrand · Introduction to Numerical Analysis
3. Dixon and Szego · Towards Global Optimization 2: The Global Optimization Problem
4. Fishman · Monte Carlo: Concepts, Algorithms, and Applications
5. Johnson, Kotz, and Kemp · Univariate Discrete Distributions 2nd edition
6. Panjer H.H. (1980). The aggregate claims distribution and stop-loss reinsurance. *Transactions of the Society of Actuaries* **32**, 523-545.
7. Zurich IC Squared, www.ic2.zurich.com
8. Barron's Dictionary of Banking Terms