Estimating the Variability of Loss Reserves

By Richard E. Sherman, FCAS, MAAA

This paper presents two major refinements to common practices in estimating the variability of loss reserves. The first is specification of a closed form density function that very closely approximates the aggregate loss distribution resulting from the convolution of the typical frequency and severity distributions—even at the extreme tails. This new distribution apparently has the following useful properties: (1) the sum of two of these distributions is also a distribution of the same form, regardless of whether the original distributions are independent or correlated; and (2) the values of the coefficients of the combined distribution is a function of the coefficients of the original distributions. These two properties should significantly streamline the process of determining the aggregate distribution for loss reserves.

This paper openly questions the validity of the practice of approximating the confidence level factors for the aggregate distribution of loss reserves by using those factors from the aggregate distribution of all claims incurred for one (or more) accident years. It presents an alternative approach that avoids the likely inaccuracies of this common practice by using the new density function to represent: 1) the variability of each accident year piece of the loss reserve (at successive, advancing ages) and 2) the variability of the sum of these pieces.

Estimating the Variability of Loss Reserves

This paper presents two major refinements to common practices in estimating the variability of loss reserves. The first is specification of a closed form density function that very closely approximates the aggregate loss distribution resulting from the convolution of the typical frequency and severity distributions—even at the extreme tails. This new distribution apparently has the following useful properties: (1) the sum of two of these distributions is also a distribution of the same form, regardless of whether the original distributions are independent or correlated; and (2) the values of the coefficients of the combined distribution is a function of the coefficients of the original distributions. These two properties should significantly streamline the process of determining the aggregate distribution for loss reserves.

This paper openly questions the validity of the practice of approximating the confidence level factors for the aggregate distribution of loss reserves by using those factors from the aggregate distribution of all claims incurred for one (or more) accident years. It presents an alternative approach that avoids the likely inaccuracies of this common practice by using the new density function to represent: 1) the variability of each accident year piece of the loss reserve (at successive, advancing ages) and 2) the variability of the sum of these pieces.

Estimating the Variability of Loss Reserves

This paper challenges two notions widely held by casualty actuaries:

- One can obtain accurate estimates of the confidence level factors for the aggregate distribution of loss reserves at any given evaluation date by using the confidence level factors for the aggregate loss distribution of all claims incurred for one (or more) accident years.

- There is no closed form probability distribution that can closely approximate the aggregate loss distribution which is the result of the convolution of a poisson frequency distribution with a severity distribution such as the lognormal (or weibull or pareto).

The research described in this paper indicates that the confidence level factors for the aggregate ***reserve*** distribution (as contrasted to the aggregate ***loss*** distribution) differ significantly from those that have been commonly derived for the aggregate ***loss*** distribution. Furthermore, the confidence level factors (for all open and IBNR claims) for a given accident year evolve considerably as the accident year matures. This is an expected result because:

- The typical severity distribution for ***all*** claims incurred for one or more accident years includes a large number of small quickly settling claims, whereas the severity distribution for all open and IBNR claims at some age such as 12 or 24 or 36 months naturally excludes most of these smaller claims.

- The number of open and IBNR claims declines substantially as an accident year ages, causing the confidence level factors to spread out.

While it is possible that under some circumstances the compacting effect of the first phenomena will exactly offset the dispersing effect of the second, it is reasonable to expect that this will be the exception rather than the rule.

This paper also presents a new probability distribution which approximates simulations of aggregate loss (and reserve) distributions much better than any of the standard distributions currently in use. The closeness of this approximation holds up very well even at the extreme tails, including confidence levels of 98%, 99%, 99.5% and 99.9%. It is in this region that the Heckman-Meyers algorithm tends to produce approximations that are not as good as would normally be desired.

While this paper presents a new probability distribution that provides very close approximations when the underlying severity distribution is lognormal, the form of this new density function can be extended by analogy to forms that produce close approximations of aggregate loss (or reserve) distributions which result when the underlying severity distribution is either a weibull or a pareto distribution. We have used the lognormal distribution as the basis for most of the development work in deriving the formula for the density function of the aggregate loss (or reserve) distribution because the form of this new density function is more obviously related to that of the lognormal density function.

We will first present this new distribution and then conclude the paper by using it to present and support our first contention.

Defining A Density Function That Closely Approximates Aggregate Distributions

Several CAS papers in the past two decades have focused on different methodologies for approximating aggregate loss distributions which are the result of the convolution of a frequency distribution with a severity distribution. Often, either a Poisson or a negative binomial distribution is used to represent the variability inherent in the number of claims which may occur during a given year. Then, either a lognormal, Pareto or Weibull distribution is used to represent the variability inherent in the size of a claim, given that a claim has occurred. While these papers have provided very useful methods for approximating aggregate loss distributions, to this author's knowledge, none have presented a family of probability distributions that simultaneously fit aggregate loss distributions at both the tails and the mid-sections. This paper presents such a family and example fits to simulated aggregate loss distributions. It also examines various properties of the parameters and also provides specific applications to the problem of estimating the variability of loss reserves.

The starting point of our search was the typical lognormal severity distribution. Its density function for the natural logarithms of the fitted amounts is of the form

$$A * EXP(-B*Z^{2.0}),$$

where

$$Z = (X - \mu)/\sigma,\ A = 1/((2*\pi)^{0.5})*\sigma \text{ and } B = 1/(2*\sigma^2).$$

We will refer to this as the "two parameter" aggregate distribution. As has been well known for many years, the lognormal distribution does not produce good fits to the aggregate distribution resulting from the convolution of a poisson frequency distribution and a lognormal distribution. Although the shape of the resulting aggregate distribution roughly resembles that of the underlying lognormal severity distribution, it systematically differs from the best fitting lognormal distribution in specific ways. By noting the nature of those differences, we postulated ways of modifying the standard lognormal distribution by adding a third parameter to approximate the warping effects of the process of convolution on the shape of the aggregate loss distribution. After testing numerous different reasonable modifications of the lognormal distribution, we concluded that a three parameter distribution of the following form very closely fit simulations of the natural logs of aggregate losses:

$$A * EXP(-B*Z^{(2.0 - C*Z)}),$$

where Z is defined by

$$Z = |(X - Median)/\sigma|.$$

It was necessary to define Z so that it would always be positive so that undefined values could be averted. This is not an issue for the standard lognormal distribution, because the exponent of Z is always 2.0.

It should be noted that the median has replaced the mean. In the case of the standard "two parameter" distribution, the mean equals the median. We introduced this generalization because it tended to improve the goodness of fit in the examples we tested. In each case, the value of the parameters A, B and C were determined by deriving the best least squares fit of the "three parameter" distribution to the simulated aggregate distributions. The standard lognormal distribution becomes a special case of this family when A and B take on the values noted above ($A = 1/((2*\pi)^{0.5})*\sigma$ and $B = 1/(2*\sigma^2)$) and C is zero.

In the first example presented in this paper, we started with a poisson frequency distribution with a lambda value of 1,000 and a lognormal severity distribution with a mean of $10,000 and standard deviation of $50,000. We then performed a Monte Carlo simulation with 100,000 trials to obtain a close approximation of the resulting aggregate loss distribution.

When the three parameter distribution was fitted to these simulated aggregate loss amounts, we noted that the goodness of fit was radically improved by having C take on one constant value for all X greater than the median and a different constant value for all X less than the median. Specifically, for the upper half of the density function, C is 0.1975, while for the lower half, C is -0.2646. (We will henceforth denote the C value for the upper half as C_U and the C value for the lower half as C_L.) This means that the exponent of Z is less than 2.0 for the upper half of the density function and greater than 2.0 for the lower half. It should also be noted that the exponent of Z is a linear function of Z. This results in the exponent of Z deviating to an increasing degree from the 2.0 value of a lognormal distribution as Z increases. In other words, the degree of deviation from a lognormal distribution grows larger as X becomes farther from the median. In the case of our example, this has some interesting effects:

- The distribution is very similar to a lognormal distribution for X values near the median (or Z values near zero).

- The distribution becomes less and less like a lognormal distribution as Z increases.

- It gives the distribution a thicker tail than the lognormal distribution at its upper end. This dramatically increases the goodness of fit to the aggregate loss distribution.

- It gives the distribution a thinner tail than the lognormal distribution at its lower end. This also dramatically increases the goodness of fit to the aggregate loss distribution.

In our example, for the upper half of the fitted distribution, the density function is

$$2.9358 * \mathrm{EXP}\ (-.6845 * Z^{\,2.0+\,0.1975\,Z}) \text{ for } X > 0.0.$$

For the lower half of the fitted distribution, the density function is

$$3.0420 * \mathrm{EXP}\ (-.5617 * Z^{\,2.0\,-\,0.2646\,Z}) \text{ for } X < 0.0.$$

Exhibit 1 provides a summary comparison of the closeness of fit of the two and three parameter distributions, which are detailed in Exhibits 2 and 3 respectively. It also shows a similar comparison for the four parameter distribution and an alternate form of the four parameter distribution. Both of these distributions will be introduced later, and the results of fitting those distributions to the example aggregate loss distribution described above, are shown in Exhibits 4 and 5 respectively.

Exhibits 2U and 2L provide a comparison of the actual and fitted values of the simulated density function for the lognormal distribution. Exhibits 3U and 3L provide the same kind of comparison of the actual and fitted values for the three parameter distribution. The rightmost column of each exhibit contains the percentage differences between the simulated (and smoothed) densities and the fitted densities.

To examine the closeness of fit of each of these distributions, the average percentage difference between the simulated densities and the fitted densities within each of the indicated ranges of cumulative probabilities was calculated and displayed in Exhibit 1. For example, if the range is between 80% and 90%, then the average percentage difference is the average of the differences for the eleven integer percentage probabilities bounded by 80% and 90%.

As Exhibit 1 plainly shows, the closeness of fit is dramatically better for each of the three and four parameter distributions—as compared with that for the standard two parameter lognormal distribution. A review of the average percentage differences for the two parameter lognormal distribution indicates that the best fitting lognormal does a poor job of fitting the aggregate loss distribution when the cumulative probability is above 95%. It also indicates that the two parameter distribution consistently overestimates the aggregate density function for cumulative probabilities between 50% and 90% and consistently underestimates it for cumulative probabilities between 10% and 50%. In simple language, the shape of the best fitting typical lognormal distribution is clearly different from that of the aggregate loss distribution.

Our goal was to define distributions that would fit the example aggregate loss distribution in such a way that: 1) the average percentage differences for each group of cumulative probabilities would be substantially smaller than those for the two parameter distribution and 2) the signs of the average percentage differences would tend to alternate. If that were to be the case, it would be reasonable to conclude that the proposed distribution either matched exactly the form of the aggregate loss distribution, or that it at least, very closely approximated it in all probability regions of interest. From a practical standpoint, finding a distribution that tightly approximated the aggregate loss distribution would be virtually as good as defining a new distribution that, from a theoretical standpoint, perfectly described it.

One advantage of focusing on percentage differences is that it highlights the goodness (or lack thereof) of fit at both ends of the aggregate loss distribution. This is not the case when the criteria for goodness of fit is a minimization of the sum of the squares of the differences.

Given the above criteria for closeness of fit, an examination of Exhibit 1 indicates that the three parameter distribution tightly fits the aggregate loss distribution in the range of cumulative probabilities of 4% through 96%. As the bottom of Exhibit 1 shows, usage of the three parameter distribution improves the closeness of fit by 80% to 90% over the two parameter distribution.

The R Squared value for the fit of the three parameter density function was 0.9981 for the upper half and 0.9968 for the lower half. This is much higher than that of the best fitting lognormal distribution—for which the R Squared value was 0.9830. This represents an 89% reduction of the difference between the R Squared value and 1.0 for the upper half and an 81% reduction for the lower half.

We also noted that the average absolute %-age difference between the fitted and actual densities was reduced from 7.07% for the lognormal distribution to 3.01% for the upper half and 2.37% for the lower half. It is also interesting to note the frequency at integer confidence level values of different size %-age errors for the two distributions:

Size of Absolute %-age Difference	**Lognormal Distribution**	**3 Parameter Distribution**
0.0% - 0.9%	10	32
1.0% - 1.9%	8	34
2.0% - 2.9%	4	22
3.0% - 3.9%	14	5
4.0% - 4.9%	8	1
5.0% - 7.4%	38	0
7.5% - 9.9%	10	0
10.0% - 24.9%	3	3
25.0% +	4	2

If we look at this on a "less than" basis, we have the following comparison:

Size of Absolute %-age Difference	**Lognormal Distribution**	**3 Parameter Distribution**
Less Than 1.0%	10	32
Less Than 2.0%	18	66
Less Than 3.0%	22	88
Less Than 4.0%	36	93
Less Than 5.0%	44	94
Less Than 7.5%	82	94
Less Than 10%	92	94
Less than 25%	95	97
25.0% +	99	99

In other words, at 88 of the 99 integer cumulative probabilities shown in Exhibits 3, the three parameter distribution estimate is less than 3% different from the actual density value. For the two parameter distribution, the fitted value is within 3% of the actual value only 22 times out of the 99. Clearly, the addition of another parameter greatly improves the goodness of fit.

The largest contiguous region where the fitted density function had errors of the same sign that were also greater than 5% was 3 successive integer cumulative percentages, and both of these regions were at the extreme tails. This may be observed in Exhibit 3.

In contrast, the best fitting lognormal distribution exhibited two sizable regions where the errors were of the same sign and were greater than 5%. The largest contiguous area covered 26 integer cumulative percentages and the second largest covered 21such percentages. This may be observed in Exhibit 2.

In addition, the best fitting lognormal distribution significantly underestimated the densities for the highest confidence levels, as indicated below:

Confidence Level	Simulated Density	Lognormal Density	%-age Underestimation of Simulated Density
99%	.077	.015	80%
98%	.185	.085	54%
97%	.300	.189	37%
96%	.404	.308	24%
95%	.510	.435	15%
94%	.613	.559	9%

Furthermore, as the above table shows, the degree of underestimation increases as the confidence level approaches 100%.

Intuitively one would expect that the density function of the aggregate loss distribution would be similar to the sum of several severity distributions. If the initial severity distribution were not skewed, one would expect that it would become less dispersed (i.e., the ratio of the standard deviation to the mean would decline) as the aggregate distribution is the result of the "sum" of more and more claims. This same phenomenon would also be expected for the lower half of a skewed distribution—because the lower half isn't very "skewed." This type of behavior is exhibited by the proposed family of density functions when the exponent of Z is greater than 2.0 for the lower half.

One might also expect (although this is less obvious) that the upper half of an aggregate loss distribution that is similar to the sum of many skewed severity distributions would become more dispersed than the initial severity distribution. This type of behavior is exhibited by the three parameter distribution when the exponent of Z is less than 2.0 for the upper half.

We have noted that aggregate loss distributions tend to be "schizophrenic" in their personality—with the shape of the lower half being less dispersed than expected and the shape of the upper half being more dispersed than expected. We have used the word "expected" here to refer to the shape anticipated if the aggregate loss distribution were the same type of distribution as the initial severity distribution. The "schizophrenic" behavior of the proposed family of density functions plays a major part in its ability to fit aggregate loss distributions much better than any of the commonly used probability distributions.

Another phenomena we noticed was that the A and B parameters of the best fitting lognormal distribution were very close to the average of those parameters for the two halves of the new distribution.

Param-eter	Upper Half of Distribution	Lower Half of Distribution	Average of Parameters	Lognormal Parameters
A	2.9358	3.0420	2.9889	2.9901
B	.6845	.5617	.6231	.6162

C	.1975	-.2646	-.0336	0.0

One weakness of the above three parameter distribution was that the goodness of fit deteriorated at the upper tail (i.e., above the 96% or 97% confidence level). This can be seen from the top of Exhibit 3U, Page 1.

Upon further research, we determined that substantially better fits could be obtained at the upper tail by adding a fourth parameter (D) and changing the form of the density function of the natural logs to

$$A * EXP(-B*Z^{(2.0 - C * Z^D)}).$$

Exhibits 4U documents the resulting curve fit for the four parameter distribution to the upper half of the distribution and thus provides a direct comparison with the best curve fit obtainable with the three parameter distribution in this region (Exhibits 3U). We also noted that there was also an additional improvement in terms of goodness of fit for regions below the 98% confidence level when this fourth parameter was added. A quick comparison of the top rows of Exhibit 4U with the corresponding rows of Exhibit 3T clearly illustrates how dramatic the goodness of fit is in the upper tail with the addition of a fourth parameter. Exhibits 4L provide similar information for the lower half of the aggregate loss distribution.

Exhibit 4T displays the goodness of fit of the four parameter distribution in the extreme tail region—where the cumulative probabilities are 98% or greater. A comparison of the fitted densities to the smoothed simulated densities shows a remarkably close fit, even up to cumulative probabilities above 99.9% and 99.95%. Considering that the Z values (which are analogous to the normalized standard deviations) are quite large in the region (e.g., 5.0 to 7.5), it is surprising that anything can be found that closely fits the aggregate loss distribution in this region.

As may be expected, adding another parameter improves the goodness of fit but also adds another level of complication in terms of determining the best fitting parameter values as well as deriving potential formulas for each of these parameters. It also makes the search for closed form formulas for various aspects of the new distribution more difficult. We invariably found that basic numerical search routines (such as EXCEL's Solver), however, quickly found minimums for such criteria as the sum of the squares of the differences or the average absolute percentage difference.

We also tried the following form of a four parameter density function of the logs of the outcomes

$$A * EXP(-B*Z^{(2.0 - C * (1 - Z)^D)}).$$

As may be seen by reviewing Exhibits 1 and 5, this form of the four parameter distribution produces fits which are quite similar in shape and in goodness of fit to those of the first type. However, the original four parameter distribution generally produced the best fits.

Comparisons with Other Common Density Functions

We would note the following observations regarding the proposed family of density functions and other common density functions:

- As noted above, the lognormal distribution belongs to the proposed family as a special case.

- The density function is similar to that of the Weibull distribution. The proposed family is more flexible and general than the Weibull density function—because the exponent of Z (or more precisely, of $((x - L)/\alpha)$), in the Weibull is a constant for the entire range of x values.

- Tails can be made quite similar, or even thicker than, those of a Pareto distribution. We have not attempted to show this via analysis of the density functions of the two distributions, but have merely noted this based on quick comparisons of corresponding values of each density function, as produced on tables in a spreadsheet.

- It seems that it is a more general family than most, if not all, of the common density functions. The linear nature of the exponent of Z as Z increases, and having two values for C (C_U and C_L), allows for much more freedom in having the example density function conform to the actual shape of the aggregate loss distribution in question.

- The proposed family could be broadened even further by making the exponent of Z a polynomial or other function (rather than a power function). Such a modification would make the proposed family of density functions even more flexible in terms of closely fitting the actual shape of the aggregate loss distribution. This may well be necessary when, for example, the aggregate loss distribution becomes more complicated—such as would be the case when that distribution is the result of the sum of several correlated aggregate loss distributions.

- Constraints might need to be applied to a polynomial or other non-linear function as the exponent of Z because the total area under the density function might become infinite.

- It is likely that one thing that might be sacrificed as the result of the more general nature and greater flexibility of the proposed family of functions is that there many not be any simple formula for the value of the constant A. This value can be approximated via numerical integration. An analysis of the results of numerous such integrations of different distributions in the proposed family might yield some clues as to possible formulae for the constant A.

Further research we have completed has indicated that both the three and four parameter distributions defined in this paper have the following properties:

- The sum of two independent distributions is another distribution of the same family. The value of the constant A appears to be $(e-1.0)^{\wedge 0.5}$ or times the constant A of the original distribution. This also appears to be true for the constant C, although the ratio between the two constants was somewhat greater than this.
- The sum of two correlated distributions is another member of the family. The values of the various constants appear to be either a linear or exponential function of the correlation coefficient.

Extensions to Approximations of Aggregate Loss Distributions Based on Underlying Weibull and Pareto Severity Distributions

In the last section of this paper we derived a three parameter density function of the logs of claim amounts of the form

$$A * EXP(-B*Z^{(2.0\ -\ C*Z)}).$$

This density function could have been presented in the following form:

$$A * EXP(-B*Z^{2.0}) * EXP(-BZ^{\ -\ C*Z)}).$$

Presented in this way, the first part is the same as the form of the density function of the logs of amounts that are lognormally distributed, while the second EXP term may be viewed as being a “convolution distortion” function. This same “convolution distortion” function, when multiplied by the density function of either the Weibull or the Pareto distribution, produces analogous results to those described above when the underlying severity distribution is lognormal.

Estimating the Variability of Loss Reserves

The problem of accurately estimating the variability of loss reserves has historically been perplexing because of the difficulty of deriving a solution purely via theoretical considerations or equations. It becomes a much simpler problem when it is approached by a combination of simulation, distribution fitting and analysis of empirical results. This process becomes illuminating as we apply it to a data base of over 5,500 claims that: 1) includes the incurred and paid values of each claim at every annual evaluation date and 2) is old enough that virtually all claims have already closed. We included accident years 1983 through 1987 in our data base. As will be seen shortly, it can be quite helpful to know, on an after the fact basis, the size of loss distribution of every reported and IBNR claims making up a loss reserve at a specific evaluation date.

By working with a data base of this kind, we have the advantage of knowing, on a hindsight basis, what the actual size of loss distribution of open and IBNR claims was for a given accident year or years as of each year end. We also know exactly how many open and IBNR claims there were as of each of these prior dates. Of course, we also know the actual size of loss distribution of open claims at each of their reserved amounts at each year end, as well as the number of open claims. By having both sets of size of loss distributions, we can note how the parameters of the best fitting distributions shift from a currently reported to an ultimate basis. By analogy, the same kinds of shifts can be inferred on the current distribution of case reserves—in order to estimate the ultimate distribution of hindsight values.

It is possible to obtain a data base similar in depth of information to that just described if we are working only with liability claims—where presumably all loss payments are made on the closure date of each claim. By making this simplifying assumption, a full history of payments and hindsight reserves can be constructed as long as the accident date, report date and closure date of each claim are known. Obviously, it would be necessary to exclude paid ALAE from such a data base.

With the above information, it was a very straightforward process to simulate the aggregate loss distribution for the outstanding reserve for any given accident year as it progressed from 0 months of development to, for example, 120 months of development, in annual increments. We proceeded as follows:

1. As of 0 months of development, we assume the number of claims is, for example, Poisson distributed with a lambda equal to the total number of claims, which we already know. We also assume that the size of loss distribution is defined by the actual distribution of all known settlement amounts on all claims. We fit various common distributions to this set of actual claim values and found that the lognormal distribution produced the best fit. It was then used in our simulation of the aggregate loss distribution as of 0 months of development. We fit the four parameter distribution described in the first section of this paper to the simulated aggregate loss distribution and obtained a set of parameters representing the aggregate loss distribution as of 0 months of development.

2. As of 12 months of development, we know exactly how much has been paid out in the first year. We also know exactly how many claims remain to be settled and the exact amounts of every claim closed after 12 months of development. This is a

very similar situation to that at 0 months of development. We proceed as we did at 0 months of development and perform a simulation to obtain the aggregate distribution of losses yet to be paid as of 12 months of development. We again fit the four parameter distribution to the simulated aggregate loss distribution and obtain another set of parameters.

3. The same process is completed as of 24 months of development, 36 months, 48 months, and so forth. The final result is a progression of 11 different aggregate distributions, from that at 0 months of development, to that at 120 months of development. Each is represented by a set of parameters. Through studying patterns in the progression of these parameters over time, we can closely describe how the aggregate loss distribution evolves as it progresses from early to late stages of development.

One objection to this above approach is that we have assumed what we are trying to estimate. While this might at first appear to be true, it is not. Our real purpose in the above exercise is to study how the parameters of the best fitting distributions change over time as the aggregate reserve distribution moves to later stages of development. By learning how aggregate reserve distributions evolve as accident years develop, we can, by analogy, use this information to derive similar estimates for the latest accident year as well as previous immature accident years.

For example, suppose we have an accident year which is at 24 months of development. We would first obtain our best estimate of the frequency and severity distributions as of 0 months of development for that year. This is the same as our best estimate of total ultimate losses for the year. This would be translated via simulation and the fitting of the four parameter distribution into a set of parameters. Having that, we would have the benefit of knowing how the aggregate reserve distributions of other previous accident years evolved from 0 to 24 months of development to help us in estimating the aggregate reserve distribution at 24 months of development. The key parameters would be estimated by analogy to the patterns derived from the full history example described above.

After performing the above sequence of steps for the data base of 1,200 liability claims for accident year 1984, we obtained a series of lognormal size of loss distributions for the hindsight reserves as well as the actual case reserves. In each case, the parameters of these best fitting lognormal distributions progressed in a smooth fashion as each accident year matured.

By noting the degree to which the mean and standard deviation of the best fitting lognormal distribution changed in transitioning from a current reported case reserves basis to an ultimate hindsight basis for earlier, fully developed accident years, these same relationships could be inferred on the means and standard deviations of the best fitting lognormal distribution to the size of loss distribution of the case reserves at the latest evaluation—to approximate what those statistics will progress to on an ultimate hindsight basis. Having these estimated parameters, we can approximate the variability of current loss reserves via simulation or the estimated parameters of the three or four parameter distributions set forth in this paper.

For example, the following table displays the means of the best fitting lognormal distributions for the case reserves and the hindsight case (and IBNR) reserves for accident year 1984 at each stage of maturity:

Age of Accident Year (Mos.)	**Mean of Size of Case Reserve Distribution**	**Size of Hindsight Claim Reserve Distribution**	**Ultimate of Distribution Mean (B)/(A)**
0	9,307	14,324	1.539
12	11,810	16,404	1.389
24	14,877	19,563	1.315
36	16,889	21,348	1.264
48	20,972	25,439	1.213
60	24,303	29,984	1.185
72	32,733	37,905	1.158
84	44,255	50,052	1.131
96	78,894	89,242	1.117

The above table also shows the “development factors” that were indicated for accident year 1984. These factors could then be applied, by analogy, to the mean of the size of loss distribution of case reserves for each accident year—to project the means of the size of loss distributions of the hindsight case (and IBNR) reserves. An example of such an application is shown below:

Accident Year	Age of Accident Year (Mos.)	(A) Mean of Size of Case Reserve Distribution	(B) Factor to Ultimate of Distribution Mean	(C) Projected Hindsight Mean of Size of Claim Reserve Distribution (A) x (B)
1990	96	105,586	1.117	117,940
1991	84	62,636	1.131	70,841
1992	72	47,776	1.158	55,325
1993	60	39,126	1.185	46,364
1994	48	33,884	1.213	41,102
1995	36	28,671	1.264	36,241
1996	24	26,637	1.315	35,028
1997	12	22,299	1.389	30,974

The next table displays the parameters of variation (i.e., the ratios of the standard deviations to the means) for the best fitting lognormal distributions to the size of loss distributions of both the case reserves and the hindsight reserves for accident year 1984:

Age of Accident Year (Mos.)	(A) Coefficient of Variation of Size of Case Reserve Distribution	(B) Coefficient of Variation of Size of Hindsight Claim Reserve Distribution	(C) Factor to Ultimate of Distribution Coefficient of Variation (B)/(A)
0	4.821	7.120	1.477
12	3.207	4.592	1.432
24	2.572	3.333	1.296
36	2.655	3.358	1.265
48	2.643	3.269	1.237
60	2.776	3.359	1.210
72	3.043	3.648	1.199
84	2.541	2.986	1.175
96	2.538	2.934	1.156

As with the means of the best fitting lognormals, we could also use the "development factors" of the coefficients of variations for a fully developed accident year or years to derive estimates of those factors for immature years. These estimated coefficients could then be multiplied by the projected hindsight means to obtain estimates of the standard deviations of the distributions of hindsight claim reserves for each accident year as of a current evaluation date.

In the case of accident year 1984, we performed simulations with each of the lognormal severity distributions for the hindsight reserves (with the means and standard deviations shown in the above tables) and obtained a series of resultant aggregate reserve distributions which progressed in a smooth and logical way from the typical aggregate distribution we are all familiar with (i.e., the one at 0 months of development) to ones with unique parameters and properties. The parameters for the three parameter reserve functions are shown in Exhibit 6. It should be noted that the values of these parameters shift significantly as the accident year ages. This clearly shows that the aggregate reserve distribution for a given accident year will vary noticeably in shape from that of the original aggregate loss distribution—which is the same as the aggregate reserve distribution at age 0.

One by-product of these simulations and the fitted distributions is a test of the reasonableness of the assumption that confidence level factors obtained from the aggregate loss distribution for all claims incurred in a single accident year are good approximations of those same factors for the aggregate reserve distribution for that same accident year at various stages of maturity. In the case of the test data we examined, the following table clearly shows that the confidence level factors for the loss reserve for accident year 1984 as of 12, 24 and 36 months of development are less dispersed than those at 0 months of development. However, the degree of dispersion increases beyond 24 months until it exceeds that at 0 months for the oldest stages of development (60 months and later).

Confi-dence	**Confidence Level Factor for AY 1984 Reserves at X Months of Development**								
Level	**0**	**12**	**24**	**36**	**48**	**60**	**72**	**84**	**96**
10%	0.807	0.838	0.850	0.831	0.800	0.746	0.671	0.508	0.300
30%	0.892	0.915	0.924	0.913	0.894	0.866	0.810	0.693	0.504
50%	0.964	0.978	0.984	0.978	0.970	0.963	0.927	0.862	0.717
70%	1.049	1.048	1.050	1.054	1.062	1.073	1.075	1.092	1.051
90%	1.214	1.181	1.166	1.188	1.227	1.286	1.386	1.587	1.900
95%	1.316	1.263	1.235	1.272	1.328	1.418	1.604	1.964	2.602
99%	1.714	1.484	1.400	1.498	1.637	1.804	2.327	3.141	5.174

The same process of simulation and distribution fitting was also applied to the total loss reserve for accident years 1983-1987 on a combined basis. Similar results were obtained. However, we found that the alternative four parameter distribution (Exhibit 7) fit the size of loss distribution of hindsight case reserves for all five accident years combined much better than did the three parameter distribution (Exhibit 8).

This provided a test of the reasonableness of the assumption that confidence level factors obtained from the aggregate loss distribution for a single accident year are good approximations of those same factors for the aggregate reserve distribution for multiple accident years at a given evaluation date. In the case of the test data we examined, the following table clearly shows that the confidence level factors for the loss reserve for accident years 1983-1987 as of yearend 1987 are much less dispersed than those for the aggregate loss distribution for a single accident year (1984):

Confidence Level	Factor for Aggregate Loss Distribution for Accident Year 1984	Factor for Aggregate Reserve Distribution for AYs 1983-87 at Dec. 1987
10%	0.810	0.899
30%	0.895	0.950
50%	0.967	0.990
70%	1.051	1.033
90%	1.216	1.107
95%	1.324	1.150
99%	1.669	1.284

The smaller spread of confidence level factors for the reserve has been due in no small part to the greater number of claims comprising the reserve (3,728 claims) versus the single accident year (1,160 claims). It is also due in part to the fact that the loss reserve is comprised of a less disparate range of claim sizes than is the collection of all claims for a single accident year. The latter includes a high percentage of very small claims that close quickly, while the loss reserve tends to be populated by larger claims that more difficult to settle. Evidence of this can be found by noting that the coefficients of variation of the best fitting log normal distributions shown in the third prior table declines significantly as the age of the accident year increases.

Conclusion

The three and four parameter distributions presented in this paper can serve to facilitate the more accurate estimation of the variability of loss reserves, as well as of the aggregate loss distribution of all losses for a single accident year. Given the apparent additive properties of these distributions, further research to determine the precise relationships between the parameters of the combined distribution to those of the original distributions may well streamline the process of closely approximating the variability of loss reserves. In addition, the ability of these distributions to closely fit aggregate loss and reserve distributions at the tails should increase the accuracy of variability estimates in these regions.

The belief that confidence level factors for the variability of loss reserves can be obtained simply from using the same factors as are estimated for the aggregate loss distribution for all claim for a given accident year (or years) is suspect. As noted in this paper, the process of accurately estimating this variability is a much more complex task, but it can be achieved given an adequate data base and the performance of a series of fits of various distributions.

Appendix A

Derivation of the Example Density Function

In performing the research to derive the example density function, we used a process we will call "analysis of simulated empirical data" (ASED). We will illustrate and define this process by describing how we approached the solution of the following problem: Given a commonly used frequency distribution (e.g., Poisson or negative binomial) and a commonly used severity distribution (e.g., lognormal or Pareto), find a simply defined distribution that fits the convolution of the frequency and severity distributions.

The first step in the ASED process is to select specific distributions (including parameters) so that we may focus on only one very well defined problem. In our example, we chose a Poisson distribution with a lambda value of 1,000 for the frequency distribution and a lognormal distribution with a mean of $10,000 and a standard deviation of $50,000 as the severity distribution. The challenge was to define a simple distribution that very tightly fits the convolution.

As we all know, there is an outstanding algorithm (the Heckman-Meyers method) which can produce an approximation of the resultant convoluted distribution much more efficiently than can be done with Monte Carlo simulation. This approach, however, still leaves us with two problems: a) it does not yield a simple distribution as a solution, and b) the accuracy of the approximation weakens noticeably at the tails. Unfortunately, the values at the right tail of the convoluted distribution are often of significant importance.

The ASED approach takes a different tack. **The second step in the ASED process** is to perform a Monte Carlo simulation with, for example, 10,000 trials. This can be accomplished very efficiently using a spreadsheet add-in (e.g., @RISK or Crystal Ball) within either an EXCEL or Lotus 1-2-3 spreadsheet. The 10,000 outcomes are sorted by size.

The third step is to deploy a spreadsheet add-in (e.g., BestFit or Crystal Ball) to fit an entire battery of commonly used distributions to these outcomes. The best fitting distributions are noted, with parameters and goodness of fit statistics. Not surprisingly, in the case of our example, a lognormal distribution was the best fitting distribution.

The fourth step is to analyze the nature of the differences between the best fitting distribution and the distribution of actual outcomes. The spreadsheet add-in produces a nice graph plotting the differences between the actual and the fitted distributions. This graph is highly illuminating. In our analysis, we made the following observations: a) the distribution of outcomes differs systematically and smoothly from the best fitting lognormal distribution; and b) the curve of the cumulative differences appears to have three turning points, occurring approximately at the mean of the fitted lognormal distribution and at one normalized standard deviation above and below the mean.

The fifth step is to propose one or more hypothetical distributions which might exhibit this behavior. Since we have already fit an entire battery of commonly used distributions to the data, it makes sense to focus on modifications of the best fitting distribution as potential solutions. To simplify the problem, we decided to work with the normal distribution underlying the best fitting lognormal distribution. We produced a comparison between the values of the normal density function and those yielded by smoothing the distribution of the logs of the outcomes. The results were again illuminating. Let us define z as the standardized normal variable (i.e., x minus the mean divided by the standard deviation). What we found was that the density function of outcomes was less than the normal density function within the range of z values of -1 to +1 and was greater for z values greater than +1.

The sixth step is to determine the impact various modifications of the best fitting distribution would have in terms of closing the gap between the density functions of the modified distribution and the density function of outcomes. The normal density function may be represented by

$$A * EXP(-B*Z^{C}).$$

We varied each of the constants of this density function and noted that A had no apparent impact, B had some impact and C had a substantial impact on the shape of the modified density function. We then used the Solver add-in to EXCEL to modify A, B and C in order to minimize the differences between the density function of outcomes and the modified normal density function. We quickly found that we had to redefine Z so that it was always positive—since negative values resulted in the refusal of the computer to produce values for the density function. We noted that there was a noticeable difference of behavior for the right half of the curve versus the left half. This suggested applying the fit separately to each half. In doing so, we obtained two very good fits which virtually eliminated significant regions of the curves where there were contiguous areas of errors with the same sign. We noted that four of the five variables had nearly identical values (mean, standard deviation, A and B) while the last variable, C, had decidedly different values (2.45 for the lower half and 1.81 for the upper half). This resulted in a density function with two regions of definition:

$$.719 * EXP\ (-.681 * Z^{\,2.45}) \text{ for } Z < 0.0$$

$$.722 * EXP\ (-.687 * |Z|^{\,1.81}) \text{ for } Z > 0.0.$$

The seventh step is to attempt to make refinements to the proposed modified distribution. We tried varying C for the upper half in order to get the best fit possible for the tail, and determined that this occurred when C was approximately 1.58. Since the goodness of fit of the density function for the right tail is of great importance (because of the large values of the associated outcomes and their significant impact on the mean and standard deviation), it seemed desirable to modify the form of the density function shown above to produce a better fit in this extreme region. We noted that the C value of 1.58 for the upper half was approximately the same distance from the C value of 2.0 associated with the normal distribution as was the C value of 2.45 for the lower half. This suggested that C might possibly be replaced by a linear function which would produce something in the range of the above observed values for each half of the curve. We modified the proposed density function to the following to potentially accomplish this:

$$A * EXP(-B*|Z|^{(2.0 - C * Z)}).$$

We again used EXCEL's Solver to find the values of A, B and C that minimized the sum of the squares of the differences between the new density function and the density function of simulation outcomes. We obtained values of .721 for A, .684 for B and .143 for C, resulting in the following density function:

$$.721 * EXP(-.684*|Z|^{(2.0 - .143 * Z)}).$$

The result was an outstanding fit. The R^2 value was 0.9967. (This could be compared with the much lower R^2 value of 0.98257 for the best fitting lognormal distribution.) We reviewed a table comparing the resultant density function with the density function of outcomes and found no major areas of the curves where there were contiguous areas of errors with the same sign. This suggested that we had found a curve that actually fit the convolution of a Poisson frequency distribution and a lognormal severity distribution—subject to errors which were most likely the result of the typical random variations that result from performing a Monte Carlo simulation. (Alternatively, the entire curve consisted of areas were there large contiguous areas of errors of the same sign—that alternated with one another.)

The above results should be compared with those we obtained when we fit a lognormal distribution to the simulated aggregate loss distribution. In the first place, the R^2 value was much lower (0.98257 versus 0.9967). In the second place, the entire curve consisted of areas were there large contiguous areas of errors of the same sign—that alternated with one another. This clearly indicated that the lognormal distribution simply did not have the right shape to fit the aggregate loss distribution closely. (It should be remembered here that the lognormal distribution was the best fitting of all the 16 common distributions that were tested.)

The eighth step is to apply various reasonableness tests to the indicated solution. First, it seems appropriate that the proposed density function should have the lognormal distribution as a special case. Other reasonableness tests have been discussed above in our initial description of this family of density functions.

REFERENCES

Aitchison, J.; and Brown, J.A.C. *The Lognormal Distribution*. Cambridge University Press (1957).

Heckman, Philip E.; and Meyers, Glenn G., Ph.D. "The Calculation of Aggregate Loss Distributions From Claim Severity and Claim Count Distributions." PCAS LXX (1983).

Hogg, Robert V., and Craig, Allen T. *Introduction to Mathematical Statistics*. 4th Edition. Macmillan (1978).

Panjer, H.H.; and Wilmot, G.E. *Insurance Risk Models*. Society of Actuaries (1992).

Philbrick, Stephen W. "A Practical Guide to the Single Parameter Pareto Distribution." PCAS LXXII (1985): 44. Discussion by Kurt A. Reichle and John P. Yonkunas, p. 85.

Robertson, John P. "The Computation of Aggregate Loss Distributions." PCAS LXXIX (1992).

Venter, Gary G. "Review of the Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions." PCAS LXX (1983).