# Modeling Loss Ratios for Aggregate Features

## *Stephen Mildenhall, May 1997*

I have never been happy with using a lognormal to model loss ratios. This note explains why and offers an alternative. The alternative is something we could easily implement in a spreadsheet and involves parameters with a clear "real world" interpretation. I would be interested in your comments.

If you compute the mean and CV of a loss ratio, fit a lognormal, and then simulate loss ratios from the lognormal the resulting series will look nothing like the experience. It will be too extreme and have too many high, and particularly, too many very low observations. On a generic primary book with an expected 65% loss ratio, you are never going to observe a 25% loss ratio, but lognormal models of the book—with reasonable CV's—will give you a 25% loss ratio. This means that the lognormal is not skewed enough. I found this puzzling, since in my investigations of aggregate loss distributions (for modeling excess treaties) I have found the lognormal is *too* skewed. Can an assumption of extreme skewness for the loss ratio distribution be justified on theoretical grounds?

To try to find such a justification, I considered a model of the loss ratio $L$ as $L=B+S$, where $B$ is represents the base loss ratio and $S$ represents shock losses. Shock losses could be cat losses, very large losses or some other similar component. Base losses are the normal, expected losses. For example in auto, $B$ may be all non-cat physical damage and property damage, and all liability losses up to some threshold. $S$ would be cat physical damage and property damage losses plus shock liability losses greater than the threshold (the whole loss, not the excess portion). Assume that both $B$ and $S$ are modeled by lognormal distributions, $B$ with a low CV and $S$ with a high CV. Assuming that $B$ and $S$ are independent[1] it is possible to compute the mean and CV of the sum, that is of $L$. At this stage we have simply have arrived at the same parameters we would have gotten looking at $L$ and not breaking it into parts. However, we can use information derived from our assumptions about $B$ and $S$ to get more insight into the form of $S$.

In particular, the assumptions that $B$ and $S$ are lognormal also allows us to compute their skewness. Assuming independence we can then compute the skewness of their sum[2]. This is new information that was not available before. Note that it is typically hard to compute the skewness from a sample—the higher the moment the more observations you need for a reasonable estimate. If the shock component includes catastrophes then CATMAP will give some idea of its skewness as a test of the lognormal assumption[3].

We now have the mean, CV and skewness of the loss ratio. Using our usual techniques we could fit a lognormal to the mean and CV. The skewness of the resulting lognormal[4] will be less than the skewness computed by splitting $L$ into $B$ and $S$. It is possible to increase the skewness of the lognormal by shifting it, that is, by modeling $L=L'+t$ where $t$ is some shifting constant and $L'$ is a lognormal with mean $E(L)-t$ and

---

[1] $B$ and $S$ would be independent by construction. Notice that it is important that $S$ contains all of a shock loss, not just the excess portion, since otherwise it would be correlated with $B$, (c.f. Meyer's CME formula for ILF's).

[2] To compute the skewness of the sum first compute the third moments of the sum, which is easy given independence: $E(X+Y)^3 = EX^3 + 3EX^2EY + 3EX\,EY^2 + EY^3$; all the required moments are available since we know the mean, variance and third moments of the components.

[3] There is no particular magic about the lognormal; we could assume a normal for $B$ and a Weibull for $S$. The point is that the distributional assumption gives the skewness—which is not readily available from a small sample. The skewnesses of $B$ and $S$ then gives the skewness of $L$.

[4] The skewness of a lognormal with coefficient of variation $CV$ is $CV(CV^2+3)$.

the same variance as $L$.[5] This process, with three variables (lognormal mean, lognormal variance and $t$) will match the mean, variance and skewness of the loss ratio distribution.

In order to see if this all has a material impact on pricing, I constructed the following example. The assumptions are:

1. $B$ has a mean of 55% and a CV of 15%,

2. $S$ has a mean of 10% and a CV of 177%, and therefore,

3. $L=S+B$ has a mean of 65% and a CV of 30%—the 177% was backed-into to get a 30% overall CV.

The lognormal assumption then gives the skewness of $B$ as 0.453, the skewness of $S$ as 10.817 and the skewness of $L$ as 8.081. A lognormal distribution with mean 65% and CV 30% only has a skewness of 0.927. Modeling $L=L'+t$ with $t$=52.2% and $L'$ lognormal with mean of 12.8% and CV of 152.3% gives an overall distribution with mean 65%, CV 30% and skewness of 8.081, as required[6]. The following table summarizes the situation and provides extra information about the four distributions ($B$, $S$, $L$ modeled as a lognormal from the mean and CV and $L'$ modeled as a shifted lognormal including the skewness). The four figures in light type-face (red) are the user inputs.

| **Item** | ***B*** | ***S*** | ***L*** | ***L'*** |
|---|---|---|---|---|
| Mean | 55% | 10% | **65%** | **65%** |
| CV | 0.15 | 1.77 | **0.30** | **0.30** |
| Mode | 47% | 1% | 46% | 55% |
| Skewness | 0.4534 | 10.8166 | 8.0809 | 8.0809 |
| **Percentiles of component loss ratios** | | | | |
| 5% | 43% | 1% | 38% | 53% |
| 10% | 45% | 1% | 43% | 54% |
| 50% | 54% | 5% | 62% | 59% |
| 80% | 62% | 13% | 80% | 70% |
| 90% | 66% | 23% | 91% | 81% |
| 95% | 70% | 35% | 101% | 95% |
| 99% | 77% | 78% | 123% | 142% |

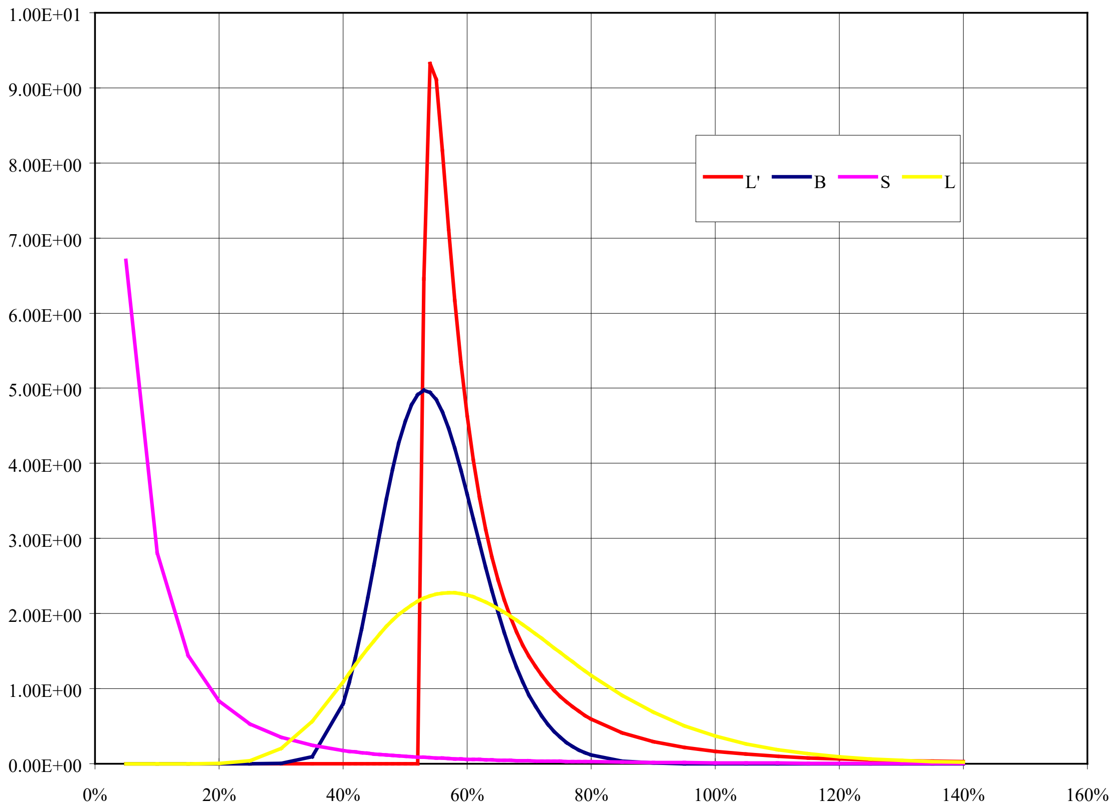
Note the percentiles look much more reasonable for $L'$ than they do for $L$. The lognormal parameters are:

| **Lognormal Parameters** | ***B*** | ***S*** | ***L*** | ***L'*** |
|---|---|---|---|---|
| mu | -0.609 | -3.011 | -0.474 | -2.653 |
| sigma | 0.149 | 1.190 | 0.294 | 1.094 |
| t | 0.000 | 0.000 | 0.000 | 0.522 |

The graph below shows the density functions of these four distributions:

[5] The CV of $L'$ is given by $CV(L') = CV(L)\dfrac{E(L)}{E(L)-t}$.

[6] See Daykin, Pentikainen, Pesonen, "Practical Risk Theory for Actuaries", pages 86-88 for details.




The high skewness of $L'$ is apparent. Also, I think $L'$ is looks like a more reasonable loss ratio distribution than $L$.

To assess the impact of the loss ratio model on pricing I computed the loss cost of a 10% loss ratio protection attaching at various points. The table below shows the results. For each attachment $a$ it shows, for $L$ and $L'$, the cost of 10% of loss ratio protection attaching at $a$ computed as

$$P = E(L-a)^{+} - E(L-(a+10\%))^{+},$$

where $X^{+} = \max(X,0)$, and a similar equation holds for $L'$. (You can draw an options diagram to see this is correct, the protection is the difference of two calls.)

There are several interesting things to note.

- For very low attachments $a$, $L'$ thinks there is no possibility of a loss ratio below $a$+10% and so the contract certainly costs 10%. This is a higher premium than the $L$ model. Such low attachments are not relevant to this cover, but would be relevant for a profit commission or swing commission.

- For attachments around the expected loss ratio of 65%, $L'$ is substantially cheaper than $L$. This is to be expected from looking at the density graph.

- For very high attachments, $L'$ is more expensive than $L$. This is an indication that $L'$ may capture the tail better than $L$.

| Attachment | $L$ | $L'$ | Pct Difference |
|---|---|---|---|
| 35% | 0.093 | 0.100 | 7.6% |
| 40% | 0.086 | 0.100 | 16.1% |
| 45% | 0.077 | 0.098 | 27.0% |
| 50% | 0.066 | 0.078 | 18.1% |
| 55% | 0.055 | 0.049 | -11.2% |
| 60% | 0.044 | 0.030 | -31.4% |
| 65% | 0.035 | 0.020 | -41.2% |
| 70% | 0.027 | 0.014 | -45.5% |
| 75% | 0.020 | 0.011 | -46.4% |
| 80% | 0.015 | 0.008 | -44.7% |
| 85% | 0.011 | 0.006 | -40.7% |
| 90% | 0.008 | 0.005 | -34.4% |
| 95% | 0.005 | 0.004 | -25.4% |
| 100% | 0.004 | 0.003 | -13.5% |
| 105% | 0.003 | 0.003 | 2.3% |
| 110% | 0.002 | 0.002 | 22.7% |
| 115% | 0.001 | 0.002 | 49.3% |
| 120% | 0.001 | 0.002 | 83.6% |

As a very quick test of this model I entered a series of loss ratios from homeowners, with the shock component given by cat losses. The statistics from the sample of 55 loss ratios were:

| | **Base** | **Shock** | **Total** |
|---|---|---|---|
| Mean | 46% | 33% | 79% |
| CV | 0.23 | 2.31 | 0.97 |
| Skewness | 0.23 | 4.92 | 4.89 |
| Min | 26% | 1% | 28% |
| Max | 72% | 504% | 557% |

In the table, the mean, min and max are percent loss ratios. The general form of these numbers lends support to my model. A more in-depth test would need to consider differences in base loss ratio between observations and also only combine experience where the expected shock loss ratios were the same. Combining observations with a high shock and low shock component makes the base look more variable. For example the base loss ratio in Florida homeowners will be lower than in Illinois homeowners since the cat load is higher; without adjusting for this the base loss ratio will appear too variable.

Given the theoretical justification for this model it is possible to fit the model without going through the process of splitting the loss ratio into two. The idea would be to use the known meaning of the parameters. The process would be:

- Pick $t$ to represent the "lowest conceivable" loss ratio,
- Adjust the mean and CV of the variable component,
- Fitting a lognormal as usual.

In the example presented above, one could argue that $t$=52.2% is too high, and that loss ratios in the 40's may be possible. This reflects an error in the selection of the CV's, which is certainly possible. Given three parameters, the pricing actuary should be able to select a loss ratio distribution which simulates realistic numbers.

I am going to look for a good source of loss ratios to use to test this model; if anyone has anything that might be useful, please let me know. Also, if you would like to try this method on one of your own accounts, I can e-mail you the spreadsheet.

Please let me know if you have any comments.