Parameter Uncertainty in Loss Ratio Distributions and its Implications

Michael G. Wacek, FCAS, MAAA

Abstract

This paper addresses the issue of parameter uncertainty in loss ratio distributions and its implications for primary and reinsurance ratemaking, underwriting downside risk assessment and analysis of sliding scale commission arrangements. It is in some respects a prequel to Van Kampen's 2003 CAS Forum paper [1], which described a Monte Carlo method for quantifying the effect of parameter uncertainty on expected loss ratios. He showed the effect was especially significant in pricing applications involving the right tail of the loss ratio distribution. While Van Kampen focused purely on the objective of quantification, this paper develops the functional form of the loss ratio distribution incorporating parameter uncertainty that is implicit in his approach. This paper thus both underpins Van Kampen's work and allows us to apply it more efficiently, because it is easier to work with the loss ratio distribution directly than to perform Van Kampen's simulation.

Suppose we have a set of on-level loss ratios from a stable portfolio of business of substantial enough size that it is plausible that the loss ratios can be viewed as a sample arising from an approximately normal or lognormal distribution, the parameters of which are unknown. What is the distribution of the prospective loss ratio? This paper discusses the drawbacks of using the "best fit" normal or lognormal distribution to model the loss ratio, particularly for pricing or risk assessment applications that depend on the tails of the distribution. While one fit is "best", frequently a number of parameter sets provide nearly as good a fit. Choosing only the "best fit" distribution means ignoring the information contained in the sample about the other possible distributions. That information can be reflected in the loss ratio distribution by weighting together *all* the plausible normal or lognormal distributions, given the sample, by their relative likelihoods. In the continuous case, where the weighting function is the density function of the parameters, the resulting distribution is the Student's *t* or log *t* distribution, respectively. This distribution, which incorporates the uncertainty about the parameters, is preferable to the "best fit" distribution for modeling the prospective loss ratio.

The paper illustrates applications ranging from aggregate excess reinsurance pricing to measurement of underwriting downside risk to estimation of the expected cost or benefit of sliding scale commissions, in each case comparing the results arising from underlying normal and lognormal assumptions and both parameter "certainty" and parameter uncertainty.

Keywords: Parameter uncertainty, aggregate loss, aggregate excess, lognormal, Student's t, downside risk

1. INTRODUCTION

This paper addresses the issue of parameter uncertainty[1] in loss ratio distributions and its implications for actuarial applications. Very few CAS papers have dealt with the subject of parameter uncertainty, notably Van Kampen [1], Meyers [2], [6], Kreps [3], Hayne [4] and Major [5]. The number is small compared to the dozens of papers that have discussed

[1] Sometimes also referred to as "parameter risk"

methods of addressing process risk. In fact, there may be more papers containing caveats saying they do *not* deal with parameter risk than there are papers that address it! In the view of this author the subject deserves more attention. As actuaries develop increasingly sophisticated models of risk processes, it is critical that we take account of our lack of knowledge of the true parameters of these models. Failure to do so can lead to systematic overconfidence and wrong conclusions.

This paper was inspired by Van Kampen's 2003 CAS Forum paper, "Estimating the Parameter Risk of a Loss Ratio Distribution,"[1] in which he presented a Monte Carlo simulation based approach for quantifying the impact of parameter risk in certain applications. Both his presentation of the problem and his solution were refreshingly clear. Unfortunately, in practice his simulation approach is a cumbersome one. This paper develops the functional form of the loss ratio distribution incorporating parameter uncertainty that is implicit in Van Kampen's approach. It thus both underpins his work and allows us to apply it more efficiently, because is it easier to work with the loss ratio distribution directly than to perform the simulations.

1.1 Organization of Paper

The paper is organized into six sections. The first section is the Introduction, where we describe the general framework. In the context of a given set of loss ratio experience that has been adjusted to the prospective claim cost and rate levels, we define the prospective loss ratio density $f_x(x)$ as the integral of the product of the conditional loss ratio density $f_x(x\,|\,\theta)$ and the joint density function of the parameters $f_\theta(\theta)$.

Section 2 introduces the assumption that the conditional loss ratio distribution is normal, which allows us to use results from normal sampling theory to describe the densities of the parameters. We discuss the drawbacks of choosing the "best fit" normal distribution $f_x^F(x)$ as the model of the loss ratio distribution in light of the uncertainty in the "best fit" parameters, especially in the case of small sample sizes.

In Section 3 we show how to incorporate parameter uncertainty by applying the general framework described in Section 1 to the normal scenario introduced in Section 2. We show that the result is a Student's t density. We also show how that Student's t density can be approximated as a weighted average of normal densities, where the weights are discrete probabilities associated with the parameters of the plausible normal densities, which we can estimate from the information contained in the loss ratio experience.

In Section 4 we change the assumption about the form of the conditional density to lognormal. Because the lognormal density can be derived from the normal by a simple change of variable, we can easily determine the formulas for incorporation of parameter uncertainty in the lognormal case from the formulas developed in Section 3. The resulting distribution is a "log *t*", which is the Student's *t* analogue to the lognormal. We compare the "best fit" lognormal and the log *t*.

In Section 5 we illustrate the four models (normal and lognormal under conditions of parameter uncertainty and parameter "certainty") in the context of three applications: 1) aggregate excess pricing, 2) downside risk measures, and 3) sliding scale commissions.

Section 6 contains the Summary and Conclusions, where we recap the main objectives of the paper, which are described as: 1) demonstrating how to derive and use the density function of the prospective loss ratio $f_x(x)$ in pricing and risk assessment applications, given on-level loss ratio experience and a normal or lognormal loss ratio process, and 2) showing, mainly by means of examples, that $f_x(x)$ has fatter tails than the "best fit" alternative $f_x^F(x)$, which implies greater loss exposure in high aggregate excess layers and greater exposure to frequency and severity of underwriting loss than that indicated by $f_x^F(x)$.

1.2 Framing the Problem

Suppose we have *n* accident years of loss ratio experience from a stable portfolio of business, where the loss ratios have been adjusted to the projected future claim cost and rate levels. Assuming the "on level" adjustments have been made perfectly and the accident years are independent, we can treat the *n* loss ratio observations as a random sample arising from the stochastic process governing the generation of loss ratios from this portfolio. Let x represent the random variable for the prospective loss ratio and let $x_1, x_2, x_3, ..., x_n$ denote the observed loss ratios. Then the sample mean is $\bar{x} = \sum_{i=1}^{n} x_i$ and the unbiased sample variance is $s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$.

In the basic actuarial ratemaking application, we need to determine the mean of the prospective loss ratio distribution $E(x)$. If x is symmetrically distributed about the mean, then we know $E(x)=\bar{x}$. If all we need is $E(x)$, then we don't need to know any more about x. On the other hand, if x is not symmetrically distributed about the mean, then not only is $E(x)\neq\bar{x}$, but to determine its value it is necessary to evaluate $\int_{-\infty}^{\infty} x\cdot f_x(x)dx$, which requires knowledge of $f_x(x)$. Likewise, in more advanced ratemaking applications, e.g., pricing aggregate excess coverage or structuring a loss-sensitive rating plan, and in cases where x is not symmetrically distributed, we need to know the distribution of x.

In this paper we will discuss how to use on-level loss ratio experience to determine the distribution of x, given varying degrees of certainty about the parameters of the underlying stochastic process, for the cases where that process is (a) normal, and (b) lognormal[2]. Because parameter uncertainty can have a significant impact on the nature of the loss ratio distribution, it is critical to the soundness of the pricing (and reserving) process that such uncertainty is taken into account.

Let θ refer to the set of parameters of the stochastic process that gives rise to the prospective loss ratio. If $f_x(x\,|\,\theta)$ is the density function of the loss ratio, given the parameter set θ, then the marginal density function of x is:

$$f_x(x)=\int_{\theta} f_x(x\,|\,\theta)\cdot f_{\theta}(\theta)d\theta \tag{1.1}$$

Formula (1.1) shows that $f_x(x)$ can be seen as a weighted average of a set of distributions of the form $f_x(x\,|\,\theta)$ where $f_{\theta}(\theta)$ is the weighting function. If there is no uncertainty about the value of the parameter set, $f_{\theta}(\theta)$ collapses to a discrete probability function with $\Pr ob(\theta)=1$ for $\theta=\theta_0$ and 0 for all other values of θ. In that case $f_x(x)=f_x(x\,|\,\theta_0)$ and for notational convenience the θ_0 is usually omitted. However, in cases where the values of the parameters are uncertain, care must be taken to maintain the distinction between $f_x(x)$ and $f_x(x\,|\,\theta)$.

2. $x\,|\,\theta$ NORMALLY DISTRIBUTED

Assume $x\,|\,\theta$ is normally distributed with parameters $\theta=\{\mu,\sigma^2\}$, these parameters representing the population mean and variance, respectively. The values of the parameters

[2] The parameter uncertainty regarding the correct distribution family is beyond the scope of this paper.

are unknown. Treating these unknown parameters in Bayesian fashion as random variables, in this context formula (1.1) can be rewritten as:

$$f_x(x) = \int_\mu \int_{\sigma^2} f_x(x \mid \mu, \sigma^2) \cdot f(\mu, \sigma^2) d\sigma^2 d\mu$$

$$= \int_\mu \int_{\sigma^2} f_x(x \mid \mu, \sigma^2) \cdot f_\mu(\mu \mid \sigma^2) \cdot f_{\sigma^2}(\sigma^2) d\sigma^2 d\mu \tag{2.1}$$

where
$$f_x(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2.2}$$

is a normal density that depends on μ and σ^2.

Because $\bar{x}$ is the unbiased and maximum likelihood estimator of μ and s^2 is the unbiased estimator of σ^2, it is tempting simply to treat μ and s^2 as parameter constants instead of as random variables[3], and set $\mu = \bar{x}$ and $\sigma^2 = s^2$ in formula (2.2), deem $Prob(\mu = \bar{x})$ and $Prob(\sigma^2 = s^2)$ to be close to 1, and conclude that, for practical purposes, the density $f_x(x)$ can be approximated by the normal density:

$$f_x^F(x) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{s}\right)^2} \tag{2.3}$$

Figure A is a graph of $f_x^F(x)$ with $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$.

[3] The reader might find it confusing that we sometimes treat μ and s^2 as parameter constants and sometimes as parameter random variables. However, to avoid overly cumbersome notation and discussion that would detract from the conceptual development, we will assume the reader can discern from context which form we are discussing.

FIGURE A

Density Function $f_x^F(x)$

Given $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$

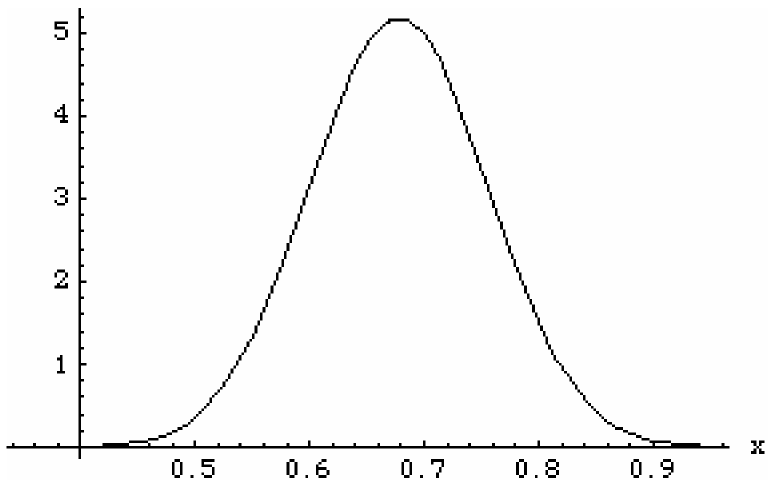


Figure A and $f_x^F(x)$ represent what is frequently called the "best fit" distribution given the sample data. However, we should be cautious about adopting this distribution as $f_x(x)$ without first examining the error structure of the sample-based parameters, which we will now do.

Given a random sample of *n* loss ratio observations, a Bayesian interpretation of results from normal sampling theory allows us to specify the densities $f_{\sigma^2}(\sigma^2)$, $f_\mu(\mu\,|\,\sigma^2)$ and $f_\mu(\mu)$.[4] We will use those results to examine the risk in the sample-based parameters, beginning with $f_{\sigma^2}(\sigma^2)$:

$$f_{\sigma^2}(\sigma^2) = \frac{1}{\sigma^2 \cdot 2^{\frac{n-1}{2}}\,\Gamma(\frac{n-1}{2})}\left(\frac{(n-1)s^2}{\sigma^2}\right)^{\frac{n-1}{2}} \cdot e^{-\frac{1}{2}\left(\frac{(n-1)s^2}{\sigma^2}\right)} \tag{2.4}$$

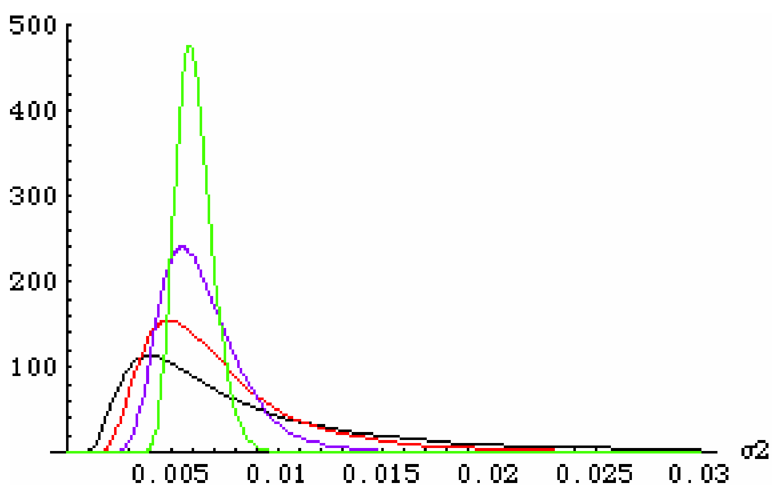
[4] Strictly speaking, we should refer to $f_{\sigma^2}(\sigma^2\,|\{x_i\})$, $f_\mu(\mu\,|(\sigma^2,\{x_i\})$ and $f_\mu(\mu\,|\{x_i\})$. However, because that notation is cumbersome and the conditionality should be clear from context, we will drop the reference to the sample $\{x_i\}$.

Because $y_{n-1} = \frac{(n-1)}{\sigma^2} \cdot s^2$ is a chi-square square random variable with *n*-1 degrees of freedom, the density represented by (2.4) is sometimes called the inverse chi-square[5]. Figure B shows $f_{\sigma^2}(\sigma^2)$ graphically for values of *n* equal to 5, 10, 25, and 100, respectively, given $s^2 = 0.0771^2$. The graph for *n*=5 is the most skewed. As *n* increases, both skewness and dispersion decreases. The graph for *n*=100 appears nearly symmetrical.

FIGURE B

Density Function $f_{\sigma^2}(\sigma^2)$

Given $s^2 = 0.0771^2$, *n* = 5, 10, 25, 100



The mean of σ^2 is a function of *n* whose value approaches s^2 as *n* approaches infinity:

$$E(\sigma^2) = s^2 \cdot \tfrac{n-1}{n-3} \tag{2.5}$$

A measure of the confidence we should feel about ascribing to σ^2 a value of s^2 is the probability that σ^2 falls within a certain tolerance of s^2. Because we want to be highly confident that $\sigma^2 = s^2$, let's set the tolerance at ±1% of s. Because $\sigma^2 = \frac{(n-1)}{y_{n-1}} \cdot s^2$, the bounds of this interval are $\frac{(n-1)}{y_{n-1}} = (.99)^2$ and $\frac{(n-1)}{y_{n-1}} = (1.01)^2$ and thus associated with chi-square values, y_{n-1}, of $\frac{(n-1)}{(.99)^2}$ and $\frac{(n-1)}{(1.01)^2}$, respectively. The probability associated with this interval is $F_{n-1}(\frac{n-1}{.99^2}) - F_{n-1}(\frac{n-1}{1.01^2})$, where F_{n-1} denotes the chi square cdf with n-1 degrees of freedom. The results are tabulated in Table 1, which shows that $Prob(.99^2 s^2 \le \sigma^2 \le 1.01^2 s^2)$ $= Prob(.0763^2 \le \sigma^2 \le .0779^2)$ is only 2% for n=5, rising to 11% for n=100. There is very little basis for having much confidence in $\sigma^2 = s^2 = 0.0771^2$ and no basis for claiming total confidence!

TABLE 1
Probability of σ within +/- 1% of s = 7.71%
Given Sample Size n

n	Degrees of Freedom	Probability σ < 7.63%	Probability σ < 7.79%	Probability 7.63% < σ < 7 .79%
5	4	39.51%	41.68%	2.17%
10	9	42.06%	45.38%	3.32%
25	24	43.40%	48.89%	5.49%
100	99	42.50%	53.67%	11.17%

Let's now turn to the distribution of μ. From sampling theory we know that the density of $\mu \mid \sigma^2$, given a sample of size n, is:

$$f_{\mu|\sigma^2}(\mu \mid \sigma^2) = \frac{1}{\sigma / \sqrt{n}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu - \bar{x}}{\sigma / \sqrt{n}}\right)^2} \tag{2.6}$$

[5] See Appendix A for derivation from the chi square with a change of variable.

which is recognizable as a normal density. The marginal distribution $f_\mu(\mu)$ is given by:

$$f_\mu(\mu) = \frac{\Gamma(\frac{n}{2})}{s/\sqrt{n}\sqrt{(n-1)\pi}\cdot\Gamma(\frac{n-1}{2})}\cdot\left(1+\frac{1}{n-1}\left(\frac{\mu-\bar{x}}{s/\sqrt{n}}\right)^2\right)^{-\frac{n}{2}} \tag{2.7}$$

which is a Student's *t* density with *n*-1 degrees of freedom. The mean and variance of μ are given below as formulas (2.8) and (2.9):

$$E(\mu) = \bar{x} \tag{2.8}$$

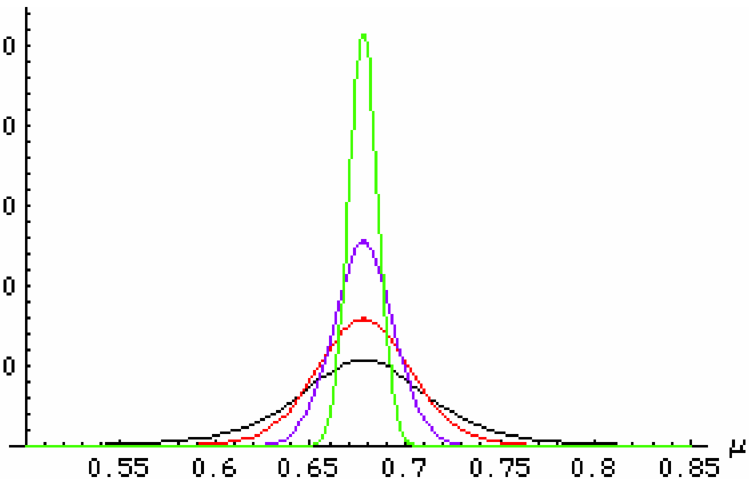
$$Var(\mu) = \frac{s^2}{n}\cdot\frac{n-1}{n-3} \tag{2.9}$$

Figure C shows $f_\mu(\mu)$ graphically for values of *n* equal to 5, 10, 25, and 100, given $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$. All the graphs are symmetrical about $\bar{x}$. The graph for *n*=5 shows the greatest variance and that of *n*=100 the least.

FIGURE C

Density Function $f_\mu(\mu)$

Given $\bar{x}$ = 67.79%, $s^2 = 0.0771^2$, n = 5, 10, 25, 100



By the same reasoning we described for σ^2, a measure of the confidence we should feel about ascribing to μ a value of $\bar{x}$ is the probability that μ falls within a certain tolerance of $\bar{x}$. Because we want to be highly confident that $\mu = \bar{x}$, let's set the tolerance at ±1% of $\bar{x}$. Because $t_{n-1} = \dfrac{\mu - \bar{x}}{s/\sqrt{n}}$, the bounds of this interval are $\bar{x} + t_{n-1}^L \cdot s/\sqrt{n} = .99\bar{x}$ and $\bar{x} + t_{n-1}^U \cdot s/\sqrt{n} = 1.01\bar{x}$. If $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$, this implies $t_{n-1}^U = -t_{n-1}^L = \dfrac{.01\bar{x}}{s/\sqrt{n}} = 0.0879\sqrt{n}$. The cumulative probabilities associated with the upper and lower bounds are given by $T_{n-1}(0.0879\sqrt{n})$ and $T_{n-1}(-0.0879\sqrt{n}) = 1 - T_{n-1}(0.0879\sqrt{n})$, respectively, where T_{n-1} is the Student's *t* cdf with *n*-1 degrees of freedom, which means that $Prob(.99\bar{x} \le \mu \le 1.01\bar{x}) = 2 \cdot T_{n-1}(0.0879\sqrt{n}) - 1$. The results are tabulated in Table 2, which shows that $Prob(.99\bar{x} \le \mu \le 1.01\bar{x}) = Prob(.6711 \le \mu \le .6847)$ is 15% for *n*=5, rising to 62% for *n*=100. While this is better than the case for σ^2, it still suggests that placing total confidence in $\mu = \bar{x} = 67.79\%$ is unwise, particularly for small values of *n*.

It should be clear from Figures B and C that the "best fit" parameters are far from the only reasonable choice, given the loss ratio experience. Why not incorporate information about those other reasonable parameter choices in our determination of $f_x(x)$?

TABLE 2
Probability of μ within +/- 1% of $\bar{x}$ = 67.79%
Given Sample Size *n*

n	Degrees of Freedom	Probability μ < 67.11%	Probability μ < 68.47%	Probability 67.11% < μ < 68.47%
5	4	42.69%	57.31%	14.63%
10	9	39.36%	60.64%	21.27%
25	24	33.21%	66.79%	33.59%
100	99	19.07%	80.93%	61.86%

3. INCORPORATING PARAMETER UNCERTAINTY—NORMAL CASE

3.1 Exact Density

In the previous section we showed that, especially in small sample cases, it is wrong to treat the "fitted distribution" $f_x^F(x)$ given by (2.3) as *the* distribution of x, because there is too great a probability of significant variation in the true value of the parameters from the "best fit" parameters. There are too many other good parameter choices to be sure that a single set of parameters adequately captures all the important information from that sample. In this section, we show how to use the results from sampling theory outlined in the previous section together with the information in the sample to obtain the correct characterization of $f_x(x)$.

We can express the random variables $x \mid \mu, \sigma^2$ and $\mu \mid \sigma^2$ in formulas (2.2) and (2.6) in terms of the standard normal random variable z as follows[6]:

$$x \mid \mu, \sigma^2 = \mu + z_1 \sigma \tag{3.1}$$

$$\mu \mid \sigma^2 = \bar{x} + z_2 \sigma / \sqrt{n} \tag{3.2}$$

The random variable σ^2 described in (2.4) can be expressed as:

$$\sigma^2 = \frac{(n-1)}{y_{n-1}} \cdot s^2 \tag{3.3}$$

where y_{n-1} is chi-square with n-1 degrees of freedom.

Expanding formula (3.1) by replacing the parameter μ with the random variable $\mu \mid \sigma^2$ given in formula (3.2), we see that:

$$\begin{aligned} x \mid \sigma^2 &= (\bar{x} + z_2 \sigma / \sqrt{n}) + z_1 \sigma && \text{(Because } \mu \mid \sigma^2 = \bar{x} + z_2 \sigma / \sqrt{n}\text{)} \\ &= \bar{x} + (z_1 + z_2 / \sqrt{n}) \cdot \sigma \\ &= \bar{x} + z \cdot \sigma \cdot \sqrt{\frac{n+1}{n}} && \text{(Because } (z_1 + z_2 / \sqrt{n}) = z \cdot \sqrt{\frac{n+1}{n}}\text{)} \qquad (3.4) \end{aligned}$$

Formula (3.4) implies the normal density $f_x(x \mid \sigma^2)$ given below as formula (3.5), which depends on σ^2 but not on μ:

---

[6] Subscripts are used to distinguish the separate instances of z in formulas (3.1) and (3.2).

$$f_x(x \mid \sigma^2) = \frac{1}{\sigma\sqrt{\frac{n+1}{n}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}}\right)^2} \tag{3.5}$$

We can alternatively expand (3.1) by replacing the parameter σ^2 with the random variable σ^2 given in formula (3.3) to obtain:

$$x \mid \mu = \mu + \frac{z_1}{\sqrt{\frac{y}{n-1}}} \cdot s \qquad \text{(Because } z_1 \cdot \sigma = \frac{z_1}{\sqrt{\frac{y}{n-1}}} \cdot s\text{)}$$

$$= \mu + t_{n-1} \cdot s \qquad \text{(Because } \frac{z_1}{\sqrt{\frac{y}{n-1}}} = t_{n-1}\text{)} \tag{3.6}$$

where t_{n-1} is the standard Student's *t* with ***n***-1 degrees of freedom.

Formula (3.6) implies the Student's *t* density $f_x(x \mid \mu)$ that depends on μ but not on σ^2, given below as formula (3.7):

$$f_x(x \mid \mu) = \frac{\Gamma(\frac{n}{2})}{s\sqrt{(n-1)\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot \left(1 + \frac{1}{n-1}\left(\frac{x-\mu}{s}\right)^2\right)^{-\frac{n}{2}} \tag{3.7}$$

Returning to (3.4), if we now expand that formula by replacing the parameter σ^2 with the random variable σ^2 described in (3.3), we see that:

$$x = \bar{x} + \frac{z}{\sqrt{\frac{y}{n-1}}} \cdot s \cdot \sqrt{\frac{n+1}{n}} \qquad \text{(Because } z \cdot \sigma = \frac{z}{\sqrt{\frac{y}{n-1}}} \cdot s\text{)}$$

$$= \bar{x} + t_{n-1} \cdot s\sqrt{\frac{n+1}{n}} \qquad \text{(Because } \frac{z}{\sqrt{\frac{y}{n-1}}} = t_{n-1}\text{)} \tag{3.8}$$

Formula (3.8) implies the Student's *t* density $f_x(x)$ that depends on neither μ nor σ^2:

$$f_x(x) = \frac{\Gamma(\frac{n}{2})}{s\sqrt{\frac{n+1}{n}(n-1)\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot \left(1 + \frac{1}{n-1}\left(\frac{x-\bar{x}}{s\sqrt{\frac{n+1}{n}}}\right)^2\right)^{-\frac{n}{2}} \tag{3.9}$$

This is a Student's *t* with *n*-1 degrees of freedom, mean of $\bar{x}$ and variance of:

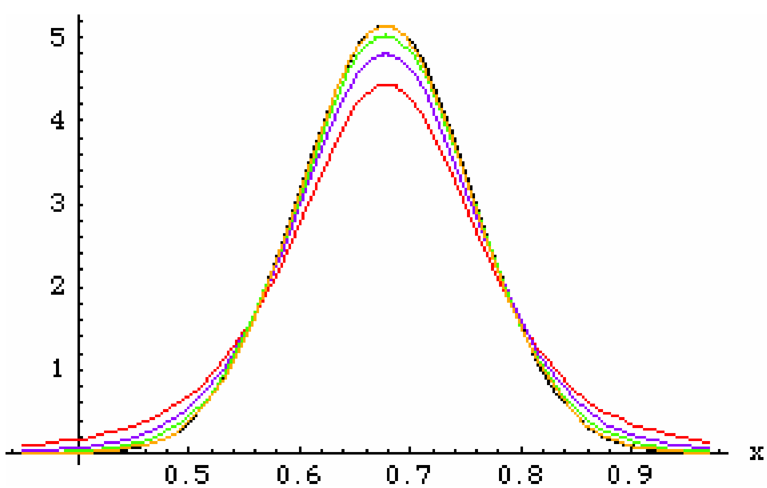
$$Var(x) = s^2 \cdot \frac{n+1}{n} \cdot \frac{n-1}{n-3} \tag{3.10}$$

Figure D shows $f_x(x)$ graphically for values of *n* equal to 5, 10, 25, and 100, respectively, given $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$. All the graphs are symmetrical about $\bar{x}$. The graph for *n*=5 shows the greatest variance and that of *n*=100 the least, with *n*=10 and *n*=25 in between. The graph corresponding to *n*=100 is visually indistinguishable from the graph of a normal density with mean 67.79% and variance 0.0771^2 (though the former has a slightly larger variance of 0.0783^2).

FIGURE D

Density Function $f_x(x)$

Given $\bar{x}$ = 67.79%, $s^2 = 0.0771^2$, n = 5, 10, 25, 100



3.2 Approximate Density

Note that formula (3.9) is the result of simplifying formula (2.1) by integrating over μ and σ^2. We can achieve a approximation to that integration by replacing the densities $f_\mu(\mu\,|\,\sigma^2)$ and $f_{\sigma^2}(\sigma^2)$ in (2.1) with discrete probability weights in the following summation:

$$
\begin{aligned}
f_x(x) \approx f_x^*(x) &= \sum_i \sum_j f_x(x\,|\,\mu_{ij}, \sigma_j^2) \cdot p(\mu_i\,|\,\sigma_j^2) \cdot p(\sigma_j^2) \\
&= \sum_i \sum_j \frac{1}{\sigma_j \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_{ij}}{\sigma_j}\right)^2} \cdot p(\mu_i\,|\,\sigma_j^2) \cdot p(\sigma_j^2)
\end{aligned}
\tag{3.11}
$$

where $$\sum_i p(\mu_i\,|\,\sigma_j^2) = \sum_j p(\sigma_j^2) = \sum_i \sum_j p(\mu_i\,|\,\sigma_j^2) \cdot p(\sigma_j^2) = 1$$

Assuming the analyst has access to software to do numerical or exact integration, for most applications it is both easier and more accurate to work directly with $f_x(x)$ as defined by formula (3.9) rather than with the approximation $f_x^*(x)$ given by formula (3.11)[7]. However, we believe it is instructive to use formula (3.11) to illustrate how the Student's *t* density defined by (3.9) can be constructed as a weighted sum of normal densities.

We will illustrate the case of n=5 with sample mean and variance of $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$. First, let us divide the domains of each of $f_{\sigma^2}(\sigma^2)$ and $f_\mu(\mu\,|\,\sigma^2)$ into 5 intervals associated with the following quantiles: 0, 0.04, 0.34667, 0.65333, 0.96 and 1. This results in intervals of length 0.04, 0.30667, 0.30667, 0.30667 and 0.04, which we will use as weights for the values of σ^2 and $\mu\,|\,\sigma^2$ associated with each interval. The midpoints of these intervals are 0.02, 0.1933, 0.50, 0.8067 and 0.98.

We associate a value of σ^2 with each interval such that $F_{\sigma^2}(\sigma_j^2) = midpt(j)$, which implies:

$$
\begin{aligned}
\sigma_j^2 &= F_{\sigma^2}^{-1}(midpt(j)) \\
&= \frac{(n-1)}{Y_{n-1}^{-1}(midpt(j))} \cdot s^2
\end{aligned}
\tag{3.12}
$$

[7] We have used CalculationCenter®2 by Wolfram Research to perform the integral calculations for this paper.

where $Y_{n-1}^{-1}(midpt(j))$ represents the chi-square inverse distribution function (with n-1 degrees of freedom) evaluated at the midpoint of the *j*-th interval.

Similarly, we associate a value of $\mu \mid \sigma^2$ with each interval such that $F_{\mu|\sigma^2}(\mu_i) = midpt(i)$, which implies:

$$\mu_i \mid \sigma_j^2 = F_{\mu|\sigma^2}^{-1}(midpt(i))$$

$$= \bar{x} - N^{-1}(midpt(i)) \cdot \sigma_j / \sqrt{n} \qquad (3.13)$$

where $N^{-1}(midpt(i))$ represents the standard normal inverse distribution function evaluated at the midpoint of the *i*-th interval.

Because μ is dependent on σ^2, there are five values of $\mu \mid \sigma^2$ for each μ-related interval *i*, one for each of the values of σ^2.

The results are summarized in Table 3, which show the parameters for 25 normal distributions and their associated probability weights. The interval midpoints $F_{\sigma^2}(\sigma_j^2)$ and the corresponding σ_j are shown in the first two columns.[8] The interval midpoints $F_{\mu|\sigma^2}(\mu_i)$ are displayed across the top of the table with the corresponding $\mu_i \mid \sigma_j^2$ shown in the body of the table below them. The probability weights associated with each row and column are at the right and bottom of the table respectively.

Each value of σ_j in the second column is to be paired with each of the values of $\mu_i \mid \sigma_j^2$ to its right. These parameter pairs define the normal distributions to be weighted using formula (3.11). For example, $\sigma_1^2 = 4.51\%^2$ is paired with each of 63.64%, 66.04%, 67.79%, 69.54% and 71.94% to form (μ, σ^2) parameter pairs $(4.51\%^2, 63.64\%)$, $(4.51\%^2, 66.04\%)$, $(4.51\%^2, 67.79\%)$, $(4.51\%^2, 69.54\%)$ and $(4.51\%^2, 71.94\%)$, with associated weights of $4\% \times 4\%$, $4\% \times 30.67\%$, $4\% \times 30.67\%$, $4\% \times 30.67\%$ and $4\% \times 4\%$, respectively.

[8] We display σ_j rather than σ_j^2 for presentational reasons.

TABLE 3

Parameters and Weights for Normal Densities in $f_x^*(x)$ Approximation
Example with n=5, $\bar{x}$=67.79%, s^2=0.0771^2

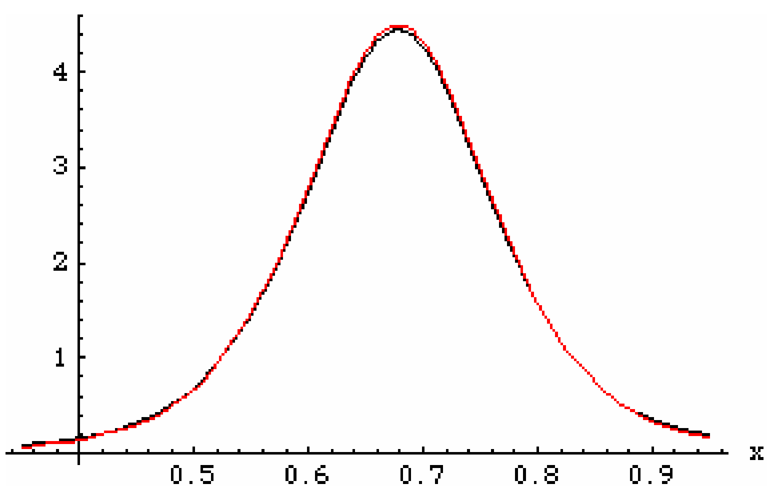
Interval Midpt $F(\sigma^2)$	σ	Interval Midpoints $F(\mu\,\vert\,\sigma^2)$ 0.0200	0.1933	0.5000	0.8067	0.9800	Row Weights
		$\mu\,\vert\,\sigma^2$					
0.0200	4.51%	63.64%	66.04%	67.79%	69.54%	71.94%	4.00%
0.1933	6.25%	62.05%	65.37%	67.79%	70.21%	73.53%	30.67%
0.5000	8.42%	60.06%	64.53%	67.79%	71.05%	75.52%	30.67%
0.8067	12.15%	56.63%	63.09%	67.79%	72.49%	78.95%	30.67%
0.9800	23.53%	46.18%	58.68%	67.79%	76.90%	89.40%	4.00%
Column Weights→		4.00%	30.67%	30.67%	30.67%	4.00%	

Figure E shows this composite density $f_x^*(x)$ based on (3.11) and represented in Table 3 to be visually identical to the Student's *t* density $f_x(x)$ defined by 3.9 for n=5.

FIGURE E

Density Functions $f_x(x)$ and $f_x^*(x)$

Given $\bar{x}$=67.79%, $s^2 = 0.0771^2$, n = 5



A visual fit is not, of course, adequate for analytical purposes. Accordingly, if the composite density is going to be used for analysis, the number and length of the intervals should be chosen in such a way that the mean and variance of $f_x^*(x)$ and $f_x(x)$ match. Matching means is a trivial process. Matching variances is more complicated. Fortunately, there is a relationship between $Var(x)$, $Var(\mu)$ and $E(\sigma^2)$ that we can use to facilitate this process:

$$\begin{aligned} Var(x) &= s^2 \cdot \frac{n+1}{n} \cdot \frac{n-1}{n-3} \\ &= s^2 \cdot (1+\frac{1}{n}) \cdot \frac{n-1}{n-3} \\ &= s^2 \cdot \frac{n-1}{n-3} + s^2 \cdot \frac{1}{n} \cdot \frac{n-1}{n-3} \\ &= E(\sigma^2) + Var(\mu) \end{aligned} \tag{3.14}$$

This means we can test the match between $Var(x)$ and $Var(x)^*$ by separately comparing $Var(\mu)$ with $Var(\mu)^*$ and $E(\sigma^2)$ with $E(\sigma^2)^*$ (the asterisks denoting the values of the functions based on the discrete approximation).

For n=5, exact calculations give $Var(\mu) = 0.0771^2 \cdot \frac{2}{5} = 0.00238$ and $E(\sigma^2) = 0.0771^2 \cdot 2 = 0.01189$, yielding a total $Var(x)$ of 0.01427 (or 0.1195^2). This compares to $Var(\mu)^* = 0.00163$, $E(\sigma^2)^* = 0.01019$ and $Var(x)^* = .01182$ (or 0.1087^2) based on the approximation defined in Table 3. Because $Var(x)^*$ is only about 83% of $Var(x)$, this suggests the approximation could (and should) be improved by increasing the number of intervals into which the domains of each of $\mu \mid \sigma^2$ and σ^2 are divided. However, because our intent was only to illustrate a simple implementation of the approximation formula (3.11), we will not pursue the optimization of that approximation here.

3.3 Section Summary

We can summarize about how varying degrees of knowledge about the parameters are reflected in the applicable probability distribution as follows:

- If both μ and σ^2 are known, then $f_x(x \mid \mu, \sigma^2)$ is a normal density with $z = \frac{x-\mu}{\sigma}$.

- If only the value of σ^2 is known, then $f_x(x|\sigma^2)$ is a normal density with $z = \dfrac{x-\bar{x}}{\sigma\sqrt{\frac{n+1}{n}}}$.
- If only μ is known, then $f_x(x|\mu)$ is a Student's *t* density with $t_{n-1} = \dfrac{x-\mu}{s}$.
- If neither μ nor σ^2 are known, $f_x(x)$ is a Student's *t* density with $t_{n-1} = \dfrac{x-\bar{x}}{s\sqrt{\frac{n+1}{n}}}$.

Table 4 shows the 90th percentile loss ratios corresponding to these knowledge scenarios, given $\bar{x} = 67.79\%$ and $s^2 = 0.0771^2$ and sample sizes ranging from 5 to 100. Several observations can be made. First, from row 1 we see that sample size does not matter if we have certainty about both μ and σ^2. Second, because the loss ratios in row 2 are always less than those in row 3, it appears that if only one of μ or σ^2 can be known, it is more helpful to know σ^2. Third, we can see that as the sample size grows larger, $f_x^F(x) = f_x(x|\mu = \bar{x}, \sigma^2 = s^2)$ becomes an increasingly better approximation of $f_x(x)$ at the 90th percentile.

TABLE 4

90th Percentile of Loss Ratio Distribution*

Given $\bar{x} = 67.79\%$ and $s = 7.71\%$

	$n = 5$	$n = 10$	$n = 25$	$n = 100$
$f_x(x\,|\,\mu = \bar{x}, \sigma^2 = s^2)$	77.67%	77.67%	77.67%	77.67%
$f_x(x\,|\,\sigma^2 = s^2)$	78.61%	78.15%	77.87%	77.72%
$f_x(x\,|\,\mu = \bar{x})$	79.61%	78.45%	77.95%	77.74%
$f_x(x)$	80.74%	78.97%	78.15%	77.79%

The 90th percentile of the weighted normal approximation $f_x^*(x)$ illustrated in Table 3 and Figure F for n=5 is 80.30%, which is close to the true $f_x(x)$ value of 80.74%. Further accuracy could be achieved by refining the number and weights of the normal densities used in the approximation.

4. INCORPORATING PARAMETER UNCERTAINTY WHEN $x \mid \theta$ IS LOGNORMALLY DISTRIBUTED

Suppose $x \mid \theta$ is lognormally distributed with unknown parameters $\theta = \{\mu, \sigma^2\}$[9]. Then the density of $x \mid \theta$ is:

$$f_x(x \mid \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} \tag{4.1}$$

The lognormal distribution gets its name from the fact that $w \mid \theta = \ln x \mid \theta$ is normally distributed with mean μ and variance σ^2:

$$f_w(w \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{w - \mu}{\sigma}\right)^2} \tag{4.2}$$

Let $w_1, w_2, w_3, ..., w_n$ denote the natural logarithms of the respective observed loss ratios $x_1, x_2, x_3, ..., x_n$. Then the sample log mean is $\overline{w} = \sum_{i=1}^{n} w_i$ and the unbiased sample log variance is $s_w^{\ 2} = \sum_{i=1}^{n} \frac{(w_i - \overline{w})^2}{n-1}$.

We can use formula (3.9) to determine the marginal distribution of w:

$$f_w(w) = \frac{\Gamma(\frac{n}{2})}{s_w\sqrt{\frac{n+1}{n}(n-1)\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot \left(1 + \frac{1}{n-1}\left(\frac{w - \overline{w}}{s_w\sqrt{\frac{n+1}{n}}}\right)^2\right)^{-\frac{n}{2}} \tag{3.9}$$

which, with the change of variable $w = \ln x$, can be restated as a function of x:

$$f_x(x) = f_w(w) \cdot \left|\frac{dw}{dx}\right|$$

[9] Note these parameters take on different values in the lognormal case from their values in the normal case.

$$= \frac{\Gamma(\frac{n}{2})}{x\, s_w \sqrt{\frac{n+1}{n}(n-1)\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot \left(1 + \frac{1}{n-1}\left(\frac{\ln x - \overline{w}}{s_w\sqrt{\frac{n+1}{n}}}\right)^2\right)^{-\frac{n}{2}} \qquad (4.3)$$

This "log *t*" density bears the same relationship to the Student's *t* as the lognormal does to the normal.

In the same way, we can use formulas (3.5) and (3.7) together with the change of variable $w = \ln x$ to determine the densities $f_x(x \mid \sigma^2)$ and $f_x(x \mid \mu)$:

$$f_x(x \mid \sigma^2) = \frac{1}{x\,\sigma\sqrt{\frac{n+1}{n}}\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \overline{w}}{\sigma\sqrt{\frac{n+1}{n}}}\right)^2} \qquad (4.4)$$

$$f_x(x \mid \mu) = \frac{\Gamma(\frac{n}{2})}{x\, s_w\sqrt{(n-1)\pi} \cdot \Gamma(\frac{n-1}{2})} \cdot \left(1 + \frac{1}{n-1}\left(\frac{\ln x - \mu}{s_w}\right)^2\right)^{-\frac{n}{2}} \qquad (4.5)$$

Formula (4.4) is a lognormal density. Formula (4.5) is a log *t* density.

If we ignore parameter uncertainty, the "best fit" parameters of $\mu = \overline{w}$ and $\sigma^2 = s_w^2$ imply the density:

$$f_x^F(x) = \frac{1}{x\, s_w\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \overline{w}}{s_w}\right)^2} \qquad (4.6)$$

which is the lognormal analogue to formula (2.3).

As we did in the case of the normally distributed $x \mid \theta$, we again counsel caution before adopting this "best fit" lognormal $f_x^F(x)$ as the correct characterization of $f_x(x)$, because it does not account for uncertainty in the parameters.

FIGURE F

Density Functions $f_x(x)$ and $f_x^F(x)$

Given $\overline{w} = -0.3946, s_w^2 = 0.1144^2$, n = 5

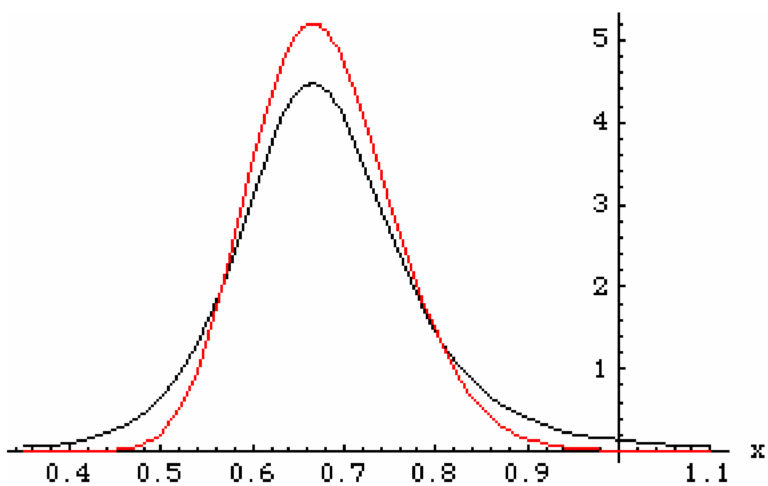


Figure F is a graph of the log *t* density $f_x(x)$ defined by formula (4.3) with n=5, plotted together with the "best fit" lognormal density $f_x^F(x)$ defined by (4.6). Values of $\overline{w} = -0.3946$ and $s_w^2 = 0.1144^2$ were determined from the same data sample that yielded $\overline{x} = 67.79\%$ and $s^2 = 0.0771^2$ used in the examples of Section 3. The log *t* distribution clearly has a larger variance and is slightly more skewed than the "best fit" lognormal. An analyst relying on the "best fit" lognormal to draw conclusions about the behavior of x, especially in the tails, will underestimate the likelihood of occurrences of x in the tails.

The log *t* density representing $f_x(x)$ can be approximated as a weighted average of lognormal densities by using formula (3.11) with the normal density replaced with the analogous lognormal density. In practice, it is usually easier to numerically integrate the log *t* directly than to construct and then integrate the equivalent composite density.

One drawback to formula (4.3) is that $E(x)$ and $Var(x)$ are infinite in realistic scenarios where n is small and/or s is not small.[10] For example, if $\overline{w} = -0.3946$ and $s_w^2 = 0.1144^2$, $E(x)$ is infinite in the case of n=5. In practice, this is not as bad as it sounds. If $\int_0^{F_x^{-1}(.9999)} x f_x(x)dx$ is a plausible mean value of x, we can conclude that the non-convergence of $\int x f_x(x)dx$ is due to behavior in the extreme right tail of $f_x(x)$. For practical purposes it is safe to approximate the mean of x as $E(x) = \int_0^{F_x^{-1}(.9999)} x f_x(x)dx$. For example, in the n=5 case just cited, $F_x^{-1}(.9999) = 346\%$ and $\int_0^{3.46} x f_x(x)dx = 68.43\%$, which is a plausible value for the mean.

An implication of the assumption that $x|\theta$ is lognormally distributed that we do not fully understand is that the value of $E(x) = \int_0^{\infty} x f_x(x)dx$ calculated directly using the density function exceeds the sample mean $\overline{x}$. We find it puzzling because (a) $\overline{x}$ is the unbiased estimator of the mean of any distribution and (b) $f_x(x)$ was parameterized using the unbiased estimators $\overline{w}$ and s_w^2 for μ and σ^2, respectively. It seems both should be correct, and yet they do not match. In the example we have been following, where $\overline{x} = 67.79\%$, even using the lognormal density given in formula (4.6), which implies no parameter uncertainty, we obtain $E(x) = 67.84\%$. When we allow for parameter uncertainty (implying use of the log t density given by (4.3)), the underestimation of $E(x)$ by $\overline{x}$ increases. In particular, for n = 5, 10, 25 and 100, respectively, $E(x)$ equal to 68.43%[11], 68.02%, 67.90% and 67.85%, implying differences of 0.64, 0.23, 0.11 and 0.06 loss ratio points, respectively. The difference is particularly noteworthy for n=5.

5. APPLICATIONS

5.1 Experience Loss Ratios

In this section we illustrate the application of the foregoing to real world problems, in particular, to the pricing of aggregate excess reinsurance, the assessment of underwriting

---

[10] We draw that conclusion because our attempt to numerically integrate $\int x f_x(x)dx$ did not converge to a solution.

[11] Calculated as $\int_0^{F_x^{-1}(.9999)} x f_x(x)dx$, because $\int_0^{\infty} x f_x(x)dx$ does not converge.

downside risk and the determination of expected commissions under sliding scale arrangements.

Suppose we have been given 5 years of on-level loss ratios x_i and their logs $w_i = ln\, x_i$, which are shown in Table 5[12]. Exposure has been constant over the experience period. The sample means, variances and standard deviations based on equal weighting of the data points are shown at the bottom. We know that the historical portfolio was large enough that it is plausible that each year's loss ratio arises from an approximately normal distribution. However, it is also plausible that the loss ratio distribution has some residual skewness, which means a lognormal model might be appropriate.

TABLE 5

On-Level Loss Ratio Experience

Accident Year	Weight	x_i	$ln\, x_i$
1	20%	66.95%	-0.40125
2	20%	59.68%	-0.51623
3	20%	76.41%	-0.26911
4	20%	72.52%	-0.32126
5	20%	77.79%	-0.25118
	Mean	70.67%	-0.35181
	*Variance**	0.554%	0.01184
	*St. Dev.**	7.45%	0.10882

* Unbiased, i.e., $E(s^2) = \sigma^2$.

For the applications illustrated in this section we will use four models for $f_x(x)$ based on: (1) normal and (2) lognormal assumptions for $x \mid \theta$ under conditions of: (A) parameter uncertainty and (B) parameter certainty.

[12] The loss ratios in Table 5 were drawn from a lognormal distribution with parameters $\mu = -0.3617$ and $\sigma^2 = .0998^2$, but let us assume we do not know that.

Given the experience in Table 5, if we assume $x\,|\,\theta$ is normally distributed, then $f_x(x)$ is given by formula (3.9) with $\bar{x} = 70.67\%$ and $s = 7.45\%$. Alternatively, if we assume $x\,|\,\theta$ is lognormal, then $f_x(x)$ is given by formula (4.3) with $\bar{w} = -0.3519$ and $s_w = 0.1088$. On the other hand, if we assume $x\,|\,\theta$ is normal and we believe $\mu = \bar{x} = 70.67\%$ and $\sigma = s = 7.45\%$ with certainty, then we must use $f_x(x) = f_x^F(x)$ as given by formula (2.3). Similarly, if we believe $x\,|\,\theta$ is lognormally distributed with $\mu = \bar{w} = -0.3518$ and $\sigma = s_w = 0.1088$ with certainty we must use $f_x(x) = f_x^F(x)$ as given by formula (4.6).

These four model choices and their characteristics are summarized in Table 6. It is worth pointing out that the lognormal-based models A2 and B2 again both indicate the density-based value $E(x)$ to be greater than $\bar{x}$.

TABLE 6

Summary of Models of $f_x(x)$

Model	$f_x(x\,\|\,\theta)$	θ	$f_x(x)$	Formula	$E(x)^*$
A1	Normal	Uncertain	t	3.9	70.67%
A2	Lognormal	Uncertain	Log t	4.3	71.37%
B1	Normal	"Certain"	Normal	2.3	70.67%
B2	Lognormal	"Certain"	Lognormal	4.6	70.76%

* Given the loss ratio experience in Table 5

5.2 Aggregate Excess Reinsurance

The pure premium of an aggregate excess layer of L excess of R, where the limit L and the retention R are ratios to premiums, is given by:

$$\int_R^{L+R} (x-R)\cdot f_x(x)\,dx + L\cdot\int_{L+R}^{\infty} f_x(x)\,dx \tag{5.1}$$

Suppose we are asked to price 20 points of coverage excess of a 70% loss ratio in four layers of 5% each.

Table 7 summarizes the results of using formula (5.1) with models A1, A2, B1 and B2. The models incorporating parameter uncertainty (A1 and A2) indicate larger pure premiums in every layer than do the models that assume parameter certainty (B1 and B2). While the difference is modest in the first layer of 5% excess of 70% (on the order of 3% to 4%), it rises rapidly as the retention increases. The pure premiums for the fourth layer of 5% excess of 85% for models A1 and A2 are respectively 300% and 200% higher than from models B1 and B2! Unless the parameters really are known with certainty, it is foolhardy to use model B1 or B2 to price aggregate excess layers.

TABLE 7
Pure Premiums of Aggregate Excess Layers
Given Sample in TABLE 5

			Limit	5%	5%	5%	5%
Model	$f_x(x \mid \theta)$	θ	Retention	70%	75%	80%	85%
A1	Normal	Uncertain		2.09%	1.14%	0.56%	0.28%
A2	Lognormal	Uncertain		2.04%	1.17%	0.64%	0.36%
B1	Normal	"Certain"		2.02%	0.92%	0.30%	0.07%
B2	Lognormal	"Certain"		1.97%	0.95%	0.37%	0.12%

5.3 Downside Risk Measures

Suppose B represents the insurer's underwriting breakeven loss ratio. The expected value of the underwriting result UR is given by:

$$E(UR) = \int_0^\infty (B - x) \cdot f_x(x)\, dx \tag{5.2}$$

$E(UR)$ can be expressed as the expected contribution from underwriting profit scenarios $UP > 0$ less the expected cost of underwriting loss scenarios $UL > 0$:

$$E(UR) = E(UP > 0) - E(UL > 0) \tag{5.3}$$

$$E(UP > 0) = \int_0^B (B - x) \cdot f_x(x)\,dx \tag{5.4}$$

$$E(UL > 0) = \int_B^\infty (x - B) \cdot f_x(x)\,dx \tag{5.5}$$

As the *pure premium* cost of underwriting loss scenarios, $E(UL > 0)$ is a measure of the insurer's underwriting downside risk.

The *probability* or *frequency* of the insurer incurring an underwriting loss $UL > 0$ is given by:

$$Freq(UL > 0) = Prob(UL > 0) = \int_B^\infty f_x(x)\,dx \tag{5.6}$$

The expected *severity* of underwriting loss, given $UL > 0$, is:

$$\begin{aligned} Sev(UL > 0) &= E(UL \mid UL > 0) \\ &= \frac{\int_B^\infty (x - B) \cdot f_x(x)\,dx}{\int_B^\infty f_x(x)\,dx} \\ &= \frac{E(UL)}{Prob(UL > 0)} \end{aligned} \tag{5.7}$$

Note that $Sev(UL > 0)$ is the Tail Value at Risk (for underwriting loss) described by Meyers[2] as a coherent measure of risk and by the CAS Valuation, Finance and Investments Committee[3] for potential use in risk transfer testing of finite reinsurance contracts.

We can use the measures defined by formulas (5.5), (5.6) and (5.7) to describe the insurer's underwriting downside risk. Given an underwriting breakeven loss ratio of B = 75%, Table 8 shows the results of using the loss ratio experience contained in Table 5 together with the $f_x(x)$ models A1, A2, B1 and B2 discussed in our analysis of aggregate excess pure premiums. For example, given the assumption that $x|\theta$ is normally distributed with unknown parameters (model A1), there is a probability of 31.19% that the insurer will have an underwriting loss averaging 7.48 points. This equates to an expected underwriting downside cost of 2.33 points. In contrast, given the assumption that $x|\theta$ is normally distributed with "known" parameters based on the loss ratio experience (model B1), there is a probability of 28.06% that the insurer will incur an underwriting loss of average severity equal to only 4.62 points, which equates to an expected downside pure premium of 1.30 points. Similarly, the lognormal model incorporating parameter uncertainty (A2) shows much larger measures of frequency, severity and downside pure premium than the lognormal model assuming parameter certainty (B2). It should be clear that ignoring parameter uncertainty in characterizing downside underwriting risk has potentially very serious and adverse consequences for an insurer's understanding of the underwriting risk it has assumed.

TABLE 8
Measures of Downside Risk
Given Sample in TABLE 5

Model	$f_x(x\|\theta)$	θ	Freq(UL)	Sev(UL)	E(UL)
A1	Normal	Uncertain	31.19%	7.48%	2.33%
A2	Lognormal	Uncertain	30.95%	9.26%	2.87%
B1	Normal	"Certain"	28.06%	4.62%	1.30%
B2	Lognormal	"Certain"	27.78%	5.34%	1.48%

5.4 Sliding Scale Commissions

Suppose a quota share reinsurance treaty has been negotiated where the ceding commission is determined according to a sliding scale. A minimum commission of 20% is payable if the loss ratio is 70% or higher. The commission slides up at a rate of 0.5 point for every point of reduction in the loss ratio below 70%, up to a maximum commission of 25% at a loss ratio of 60% or lower. The expected value of the ceding commission *C* can be expressed by formula (5.8) below:

$$E(C) = 20\% \int_{70\%}^{\infty} f_x(x)\,dx + \int_{60\%}^{70\%} (20\% + \frac{70\% - x}{2} f_x(x))dx + 25\% \int_{0}^{60\%} f_x(x)\,dx \tag{5.8}$$

Given the on-level loss ratio experience in Table 5, what is the expected value of the ceding commission? We have calculated the expected commissions based on normal and lognormal assumptions for $x \mid \theta$ under conditions of parameter uncertainty and certainty (models A1, A2, B1 and B2) and have tabulated the results in Table 9. In all cases the modeled ceding commissions are higher than the 20% commission that would be payable at a loss ratio of 70.67%. The differences range from 1.20% to 1.42%. The commissions indicated by all the models are clustered very closely together, ranging between 21.20% and 21.42%. Because the ceding commission slides in response to loss ratios that are near $E(x)$, where the model differences are less pronounced, the effect of parameter uncertainty is immaterial (at least in this example).

TABLE 9
Expected Ceding Commissions
Given Sample in TABLE 5

Model	$f_x(x \mid \theta)$	θ	C @ 70.67%	E(C)	Diff
A1	Normal	Uncertain	20.00%	21.37%	1.37%
A2	Lognormal	Uncertain	20.00%	21.42%	1.42%
B1	Normal	"Certain"	20.00%	21.20%	1.20%
B2	Lognormal	"Certain"	20.00%	21.24%	1.24%

5.5 Unequal Loss Ratio Weights

The previous examples were based on the assumption that it is appropriate to weight each observed on-level loss ratio in the historical experience equally. While that is a convenient assumption, it is not a realistic one, because exposure tends to change from year to year. Accordingly, in the interest of providing additional examples that are also more realistic, we have tabulated another set of on-level loss ratios in Table 10. These observed loss ratios arose from the same distribution as the loss ratios in Table 5. The sample mean, variance and standard deviation statistics have been computed both on a weighted basis and on the standard unweighted basis. The formulas for weighted mean and the unbiased weighted sample variance s_c^2 are:

$$\bar{x}_c = \sum_{i=1}^{n} \frac{c_i \cdot x_i}{\bar{c} \cdot n} \tag{5.9}$$

$$s_c^2 = \sum_{i=1}^{n} \frac{c_i \cdot (x_i - \bar{x}_c)^2}{\bar{c} \cdot (n-1)}, \tag{5.10}$$

where c_i denotes the weight to be used with the i-th observation, $\bar{c}$ is the mean weight and $\bar{x}_c$ is the weighted mean.

TABLE 10
On-Level Loss Ratio Experience
2nd Sample

Accident Year	Weight	x_i	$ln\, x_i$
1	16%	53.88%	-0.44823
2	18%	53.15%	-0.63203
3	22%	70.62%	-0.34790
4	23%	73.06%	-0.31391
5	21%	56.55%	-0.56998
Unweighted			
Mean		63.45%	-0.46241
*Variance**		0.744%	0.01893
*St. Dev.**		8.62%	0.13758
Weighted			
Mean		64.00%	-0.45392
*Variance**		0.767%	0.01941
*St. Dev.**		8.76%	0.13309

* Unbiased, i.e., $E(s^2) = \sigma^2$.

Though the loss ratio experience shown in Table 10 emerged from the same underlying loss ratio distribution as that in Table 5, its mean and standard deviation are significantly different. On an unweighted basis the loss ratio mean in Table 10 is more than 7 points (more than 10%) less than the loss ratio mean in Table 5 (64.00% v. 70.67%). On the other hand, the standard deviation is more than 15% greater (8.62% vs. 7.45%). The sample variation illustrated by those differences is worth remembering when we are tempted to put great weight on the credibility of a small sample.

We have calculated the aggregate excess pure premiums for the layers defined in Table 7 using the weighted basis loss ratio experience in Table 10 and displayed the results in Table 11. As in the example based on Table 5, the pure premiums for all layers are higher when priced using the models that incorporate parameter uncertainty (A1 and A2) than the models assuming the parameters are known with certainty (B1 and B2). Again the pricing difference increases as the retentions increase. However, it is also worth noting that the differences in pure premiums based on Table 10 are far less than the difference between those pure premiums and those calculated based on the experience in Table 5. For example, in Table 11 we see the indicated model A1 pure premium for 5% excess of 70% is 0.62% compared to 2.09% in Table 7. The indicated pure premiums for all other layers and models are also much lower in Table 11 than in Table 7. Both experience samples arose from the same loss ratio distribution, but the two samples indicate dramatically different pure premiums!

TABLE 11
Pure Premiums of Aggregate Excess Layers
Given Sample in TABLE 10

			Limit	5%	5%	5%	5%
Model	$f_x(x \mid \theta)$	θ	Retention	70%	75%	80%	85%
A1	Normal	Uncertain		0.62%	0.46%	0.34%	0.25%
A2	Lognormal	Uncertain		0.59%	0.46%	0.35%	0.27%
B1	Normal	"Certain"		0.51%	0.33%	0.20%	0.12%
B2	Lognormal	"Certain"		0.49%	0.33%	0.21%	0.13%

Table 12 shows the downside risk statistics calculated on the basis of the weighted loss ratio experience in Table 10. Because the sample mean Table 10 is much lower than in Table 5, the indicated probability of underwriting loss is much reduced from that shown in Table 8. While the severity of underwriting loss is not much affected, due to the large reduction in frequency, the expected cost of underwriting losses is much lower in Table 12 than in Table 8. The difference is much greater for the parameter certainty models B1 and B2 than for models A1 and A2. Models B1 and B2 now indicate minimal downside risk as measured by $E(UL)$ values of 0.49% and 0.54%. These compare to values of 1.30% and 1.48%, respectively, in Table 8, reductions of about two-thirds. On the other hand models A1 and A2 are less sensitive to the sample variation. Model A1's $E(UL)$ of 1.40% is 40%

less than its value in Table 8. The A2 $E(UL)$ of 1.87% is about 35% less than its value in Table 8. Even at these reduced values both indicate significant downside risk and both show expected underwriting loss costs more than three times as high as B1 and B2.

TABLE 12
Measures of Downside Risk
Given Sample in TABLE 10

Model	$f_x(x \mid \theta)$	θ	Freq(UL)	Sev(UL)	E(UL)
A1	Normal	Uncertain	15.78%	8.86%	1.40%
A2	Lognormal	Uncertain	15.88%	11.75%	1.87%
B1	Normal	Certain	11.53%	4.27%	0.49%
B2	Lognormal	Certain	10.59%	5.06%	0.54%

Table 13 shows the expected ceding commissions based on the weighted loss ratio experience in Table 10. As we saw in the commissions based on the loss experience shown in Table 5 and displayed in Table 9, there is little variation in the commission estimates based on using the different models. The expected commissions in Table 9 range from 22.65% to 22.81% compared to a range of 21.20% to 21.42% in Table 13. The difference due to the variation in loss ratio experience is far more important than the difference in models. Models A1 and A2 show only about 1.3 points increase in expected ceding commission and Models B1 and B2 show only about 1.5 points increase, even though the sample loss ratio is more than 7 points lower.

TABLE 13
Expected Ceding Commissions
Given Sample in TABLE 11

Model	$f_x(x \mid \theta)$	θ	C @ 64.00%	E(C)	Diff
A1	Normal	Uncertain	23.00%	22.65%	(0.35%)
A2	Lognormal	Uncertain	23.00%	22.76%	(0.24%)
B1	Normal	Certain	23.00%	22.72%	(0.28%)
B2	Lognormal	Certain	23.00%	22.81%	(0.19%)

6. SUMMARY AND CONCLUSIONS

The main objectives of this paper have been to: 1) demonstrate how to derive and use the density function $f_x(x)$ of the prospective loss ratio in pricing and risk assessment applications, given on-level loss ratio experience and a normal or lognormal loss ratio process, and 2) show, mainly by means of examples, that $f_x(x)$ has fatter tails than the "best fit" alternative $f_x^F(x)$, which implies greater loss exposure in high excess layers and greater exposure to frequency and severity of underwriting loss than that indicated by $f_x^F(x)$.

In distributional terms, we have shown that if we believe the on-level loss ratios are normally distributed, our lack of knowledge of the parameters of that normal distribution requires that $f_x(x)$ be characterized as a Student's *t* rather than a normal distribution. We may still believe the loss ratio is normally distributed, but we do not have sufficient knowledge to safely characterize it as such. The Student's *t*, which does approximate the normal for large sample sizes (see Figure D), is the best we can do.

Similarly, if we believe the on-level loss ratios are lognormally distributed, our lack of knowledge of the parameters of that lognormal distribution means that $f_x(x)$ must be characterized as a log *t* rather than a lognormal distribution, for the reasons described above.

Two other points also bear repeating. First, for right-skewed distributions, the sample mean $\bar{x}$ appears to give a lower estimate of $E(x)$ than the one determined from the density function parameterized with unbiased estimators derived from the sample. The difference is less pronounced for large sample sizes, but for small experience samples it is sizeable. We do not know what to make of this, but it adds to our discomfort about being overconfident about conclusions drawn from small samples. Second, small experience samples can exhibit significant variation from the characteristics of the population from which they arise, which *can* lead to over-pricing or under-pricing even when using the correct form of $f_x(x)$. Actuaries must resist the temptation to be overconfident about the inferences that can safely be drawn from small samples. It is wise to avoid staking too much on the conclusions of a pricing analysis based on a small sample.

Some further caveats apply. While the methods described in this paper incorporate the consequences of our uncertainty about some critical parameters into estimates of the projected loss ratio, note that they do not address other important sources of parameter uncertainty, and accordingly, are likely to underestimate the total variance of x. They address only the uncertainty arising from the sample loss ratios, given that those loss ratios are themselves certain. However, those loss ratios are estimates. Therefore, these methods do not reflect parameter uncertainty associated with loss development factors used for the projection of reported loss ratios to ultimate, nor do they reflect uncertainty in the on-level adjustment parameters. In addition, we do not know for certain that we have chosen the correct model distribution in the normal or the lognormal. Thus, while this method is an improvement over methods that do not incorporate any parameter uncertainty, a certain amount of caution remains in order.

Appendix A

Derivation of Formula (2.4)

Assume y_{n-1} is chi square with *n*-1 degrees of freedom. That implies

$$f_y(y_{n-1}) = \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \cdot y^{\frac{n-1}{2}-1} \cdot e^{-\frac{1}{2}y}$$

Perform the change of variable $y_{n-1} = \frac{(n-1)}{\sigma^2} \cdot s^2$, where σ^2 is the new random variable.

Then $\left|\frac{dy}{d\sigma^2}\right| = \frac{(n-1)}{(\sigma^2)^2} \cdot s^2$ and

$$f_{\sigma^2}(\sigma^2) = f_y(y_{n-1}) \cdot \left|\frac{dy}{d\sigma^2}\right|$$

$$= \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \cdot \left(\frac{(n-1)}{\sigma^2} \cdot s^2\right)^{\frac{n-1}{2}-1} \cdot e^{-\frac{1}{2}\left(\frac{(n-1)}{\sigma^2} \cdot s^2\right)} \cdot \frac{(n-1)}{(\sigma^2)^2} \cdot s^2$$

$$= \frac{1}{\sigma^2 \cdot 2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \cdot \left(\frac{(n-1)}{\sigma^2} \cdot s^2\right)^{\frac{n-1}{2}} \cdot e^{-\frac{1}{2}\left(\frac{(n-1)}{\sigma^2} \cdot s^2\right)} \tag{2.4}$$

7. REFERENCES

[1] Van Kampen, Charles E., "Estimating the Parameter Risk of a Loss Ratio Distribution", *Casualty Actuarial Society Forum*, Volume: Spring, **2003**, p. 177-213, http://www.casact.org/pubs/forum/03spforum/03spf177.pdf.

[2] Meyers, Glenn G., "Parameter Uncertainty in the Collective Risk Model", *PCAS* LXX, **1983**, p. 111-143, http://www.casact.org/pubs/proceed/proceed83/83111.pdf

[3] Kreps, Rodney E., "Parameter Uncertainty in (Log)Normal Distributions", *PCAS* LXXXIV, **1997**, p. 553-580, http://www.casact.org/pubs/proceed/proceed97/97553.pdf

[4] Hayne, Roger M., "Modeling Parameter Uncertainty in Cash Flow Projections", *Casualty Actuarial Society Forum*, Volume: Summer, **1999**, p. 133-152, http://www.casact.org/pubs/forum/99sforum/99sf133.pdf

[5] Major, John A., "Taking Uncertainty into Account: Bias Issues Arising from Parameter Uncertainty in Risk Models", *Casualty Actuarial Society Forum*, Volume: Summer, **1999**, p. 153-196, http://www.casact.org/pubs/forum/99sforum/99sf153.pdf

[6] Meyers, Glenn G., "Estimating Between Line Correlations Generated by Parameter Uncertainty", *Casualty Actuarial Society Forum*, Volume: Summer, **1999**, p. 197-222, http://www.casact.org/pubs/forum/99sforum/99sf197.pdf

[7] Meyers, Glenn G., "The Cost of Financing Insurance", *Casualty Actuarial Society Forum*, Volume: Spring**, 2001**, p. 221-264, http://www.casact.org/pubs/forum/01spforum/01spf221.pdf.

[8] CAS Valuation, Finance, and Investments Committee "Accounting Rule Guidance Statement of Financial Accounting Standards No. 113--Considerations in Risk Transfer Testing", *Casualty Actuarial Society Forum*, Volume: Fall, **2002**, p. 305-338, http://casact.org/pubs/forum/02fforum/02ff305.pdf .

Abbreviations and notations

CAS, Casualty Actuarial Society
C, ceding commission rate
E(*UL*), expected value cost of underwriting loss scenarios
Freq(*UL*), frequency of underwriting loss scenarios
L, aggregate excess layer limit, in loss ratio points
R, aggregate excess retention, in loss ratio points
Sev(*UL*), mean severity of underwriting loss scenarios
μ, first parameter of a normal or lognormal distribution, sometimes a random variable
σ^2, second parameter of a normal or lognormal distribution, sometimes a random variable
θ, parameter set
n, number of years in the loss ratio experience sample
c_i, weight for the *i*-th observed on-level experience loss ratio
$\bar{c}$, mean of the weights used with observed on-level experience loss ratios
s^2, variance of the on-level experience loss ratios (unbiased)
s_c^2, weighted variance of the on-level experience loss ratios (unbiased)
s_w^2, variance of logs of the on-level experience loss ratios (unbiased)
t_{n-1}, a Student's *t* distribution random variable with *n*-1 degrees of freedom
w, random var for the log of prospective loss ratio given uncertainty about underlying distribution parameters
$w \mid \theta$, random rs variable for the log of prospective loss ratio given parameters of underlying distribution
w_i, log of *i*-th observation of on-level experience loss ratios
$\bar{w}$, mean of the logs of the on-level experience loss ratios
x, random variable for the prospective loss ratio given uncertainty about parameters of underlying distribution
$x \mid \theta$, random variable for the prospective loss ratio given parameters of underlying distribution
x_i, *i*-th observation of the on-level experience loss ratios
$\bar{x}$, mean of the on-level experience loss ratios

$\bar{x}_c$, weighted mean of the on-level experience loss ratios
y_{n-1}, a chi square random variable with *n*-1 degrees of freedom
z, a standard normal random variable

Biography of the Author

Michael Wacek is President of Odyssey America Reinsurance Corporation in Stamford, Connecticut. A Fellow of the CAS and a Member of the American Academy of Actuaries, he is the author of several Proceedings and Discussion Program papers. Before joining Odyssey Re he held various actuarial and management positions at St. Paul Fire and Marine Insurance Company (a primary insurer), E.W. Blanch Company (a reinsurance broker), St Paul Reinsurance Company Limited (a U.K. reinsurer) and TIG Reinsurance Company (a U.S. reinsurer). He is a graduate of Macalester College, St. Paul, Minnesota.