We're Skewed—The Bias in Small Samples from Skewed Distributions

By Kirk G. Fleming, FCAS, MAAA

Abstract People in insurance work all the time with financial processes that are best modeled with skewed distributions. Despite our constant exposure to skewed distributions, I believe when we study sample averages from these skewed distributions we think and work with them as if they were samples from normal symmetrical distributions. In this paper I want to discuss the idea that a sample average is biased lower than the actual mean of a skewed distribution – an amount that depends on the sample size and how skewed the distribution is. I will talk about the implications that this small sample bias has for credibility procedures. Why do people ignore outliers? I will offer up some possible reason for why we ignore outliers and why deals get done despite what the data indicates. I will talk about the winner's curse or why we lose even as we win. Finally, I will offer a small sample of skewed random thoughts on why these ideas explain everything from people engaging in risky behaviors to the property/casualty insurance cycle.

INTRODUCTION

People in insurance work all the time with financial processes that are best modeled with skewed distributions. Despite our constant exposure to skewed distributions, I believe when we study sample averages from these skewed distributions we think and work with them as if they were samples from normal symmetrical distributions.

In this paper, I will show through computer simulations that the expected value of a sample average from a skewed distribution varies between the mode of the distribution and the true mean of the underlying distribution. Where the expected value of the sample average falls between those two values will depend upon how skewed the distribution is and the sample size. For small samples, the expected value of the sample average will be near the mode of the distribution and for some skewed distributions, "small" samples can be unexpectedly big. The implication is that while we are searching for information on a population's mean by examining the averages of small data samples from skewed distributions, we will most likely be getting indications that could be significantly lower than the population mean. This is in contrast to the situation when we are sampling from a symmetrical distribution where the expected value of a sample average is equal to the mean of the distribution regardless of the sample size.

I would also like to talk about some of the implications of people not realizing or ignoring that the expected value of the sample average from a skewed distribution is biased lower than the mean of a positively skewed distribution. I will talk about small sample bias and credibility procedures. I will talk about why people tend to ignore outliers and why deals get done in spite of what the data indicates. I will offer an explanation on why we can't win for losing or why making money in insurance is no easy matter. Finally, I will offer up a small sample of skewed random thoughts on how these ideas help to explain everything from people engaging in risky pursuits to the property/casualty insurance cycle.

In the paper I talk a great deal about the mode and the mean because I think those are concepts that are common ground for all of us in insurance. I hope to reach a bigger audience of insurance professionals than just actuaries. To that end, I relegated all formulas to the appendices. However, I must share a word of caution to actuaries who want to discuss these ideas with others outside our field. I have tried it and I have seen strange reactions from professionals of all kinds. People have played dead so that I would just go away and leave them alone. Others have fought back violently. I have seen our outside audit partner a hardened insurance veteran who has "seen it all" practically break his leg as he tried to escape from my office when I even hinted at these ideas in answer to his question. You have been warned.

A PRACTICAL PROBLEM

Consider the following scenario – we have a customer who has written some business in a particular state and it turns out to be profitable business. The customer would like to expand into the state to write more of this good business. Our job is to produce forecasted financial statements for this customer so that they can present their business plan to management.

Because our customer does not have a great deal of existing business in the state, we use an industry average loss ratio – a ratio that happens to be higher than our customer's actual experience when we produce a first draft of the forecasted financials. Our customer objects to the higher loss ratio since he knows that his past business has been better than the industry result. In order to acknowledge his concerns we credibility weight his past experience with the industry average to give some credence to the actual experience. By using the credibility procedure, we are recognizing that our customer's experience might actually represent a profitable niche as opposed to being just be a random fluctuation from the industry average.

However there is another explanation for why the small sample average based on our customer's past experience is different from the long term industry average as opposed to it being a profitable niche or a random fluctuation. If we are sampling from a typical positively skewed distribution, the most likely value of that small sample average will be less than the true average of the distribution simply because it is a small sample from a skewed distribution. For very small samples from highly skewed distributions, the sample average will more likely be closer to the mode rather than to the mean.

When we do a single sample from any discrete distribution, the most likely value that we will see is the mode of the distribution. That's the definition of the mode – the observation that appears most often, or in other words, has the greatest probability of occurring. The mode is one of those statistics that we learn about when we first do statistics but then we never hear much about it again unless we are trying to avoid distortions associated with extreme values. That is an injustice to the mode; it actually deserves more attention.

For a symmetrical distribution with one mode like a bell curve, the mode is equal to the mean. But for a typical distribution that we might encounter in insurance that is skewed to the right and which has only one mode, the mode is less than the median which is less than the mean. (For an example of an atypical skewed distribution where the mode is greater than the mean, see Appendix A). When we do small samples from typical skewed distributions, the most likely value for the sample average will be somewhere between the mode and the mean of the distribution. How close to the mode or how close to the mean will depend on how skewed the distribution is and the sample size. Moreover, for some skewed distributions, "small" samples can be surprisingly big.

For some insurance examples, this relationship should be in the back of our minds. Take for example the annual sample from a highly skewed distribution like the annual hurricane losses in the city of Miami. For any particular year, the most likely loss we will observe is zero -- the mode of the distribution. Every so often there will be a hurricane loss that will bring the long-term average above the zero mark but most of the losses we see will be zero.

On the other hand, an industry average loss ratio is based on a sample size that we could consider for all practical purposes to be approaching infinity. If we are dealing with large samples, even from skewed distributions, we are confident that the most likely value for the sample average will be something close to the true average of the distribution. This is the law of large numbers. As the sample size increases, the probability approaches zero that the sample average differs from the mean of the distribution by any set amount as long as the samples are mutually independent and from a distribution with a finite mean and variance.

In between the two extreme cases – a sample size of one and a sample size approaching infinity – the most likely value for the average of the sample goes from the mode of the distribution up to the distribution average. How does the most likely value of the sample erage change with how skewed the distribution is and the sample size? Let us explore this question by examining results from a positively skewed distribution that is used in insurance modeling – the lognormal distribution.

RESULTS FROM A TYPICAL SKEWED DISTRIBUTION

The lognormal distribution has been used by actuaries to model losses since at least the early 1970's [1]. A lognormal has its hump to the left and a long tail to the right. Because of the shape of the curve, the lognormal implies that small losses are more likely than very big losses. How likely a small loss is as compared to a large loss depends on how skewed the particular lognormal distribution is. The chance of a large number of small losses increases with the skew of the distribution.

Rather than talk about the skew of a distribution, I am going to talk about the coefficient of variation (CV) of a distribution. When we are working with a lognormal, a higher CV is the same as a higher skew. The CV is defined as the standard deviation of the distribution divided by the mean of the distribution. Actuaries typically refer to the coefficient of variation (CV) of a distribution rather than how skewed a distribution so that they can compare the skews of two distributions with different means. Intuitively we should feel that for a family of skewed distributions, the higher the standard deviation, the higher the CV, and the more skewed the distribution. The lognormal is always positively skewed as shown in Appendix B.

Actuaries who use the lognormal for size of loss curves very often have rules of thumb for an appropriate CV depending upon the line of business. CV's of around 1 or 2 might represent low limits liability lines of business, CV's between 2 and 5 might represent mixed property and liability losses, and CV's on the order of 10 might be used for very volatile high limits excess lines of business.

Chart 1 shows three lognormal curves each with a mean of 1000 and with varying CV's. As the CV increases, the mode or highest point on the distribution is associated with lower and lower values than the mean. Other typically skewed distributions would have the same relationship between the mode and the mean – as the skew of the distribution increases, the mode gets lower and lower as compared to the mean.

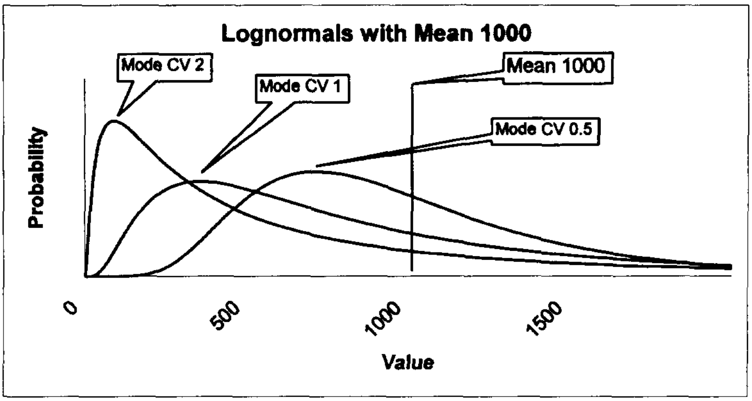

Chart 1

For the lognormal, the ratio of the mode to the mean can be written as a function of the CV. I have included that formula at the end of Appendix B for those who like formulas. Chart 2 shows the ratio of the mode to the mean for lognormal distributions with different coefficients of variation.

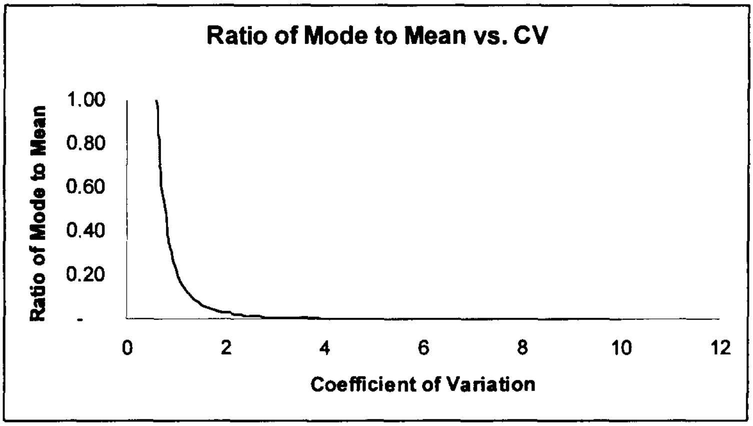

Chart 2

As the coefficient of variation increases, the ratio of the mode to the mean of a lognormal distribution drops off very quickly towards zero. What does that imply? The more skewed the distribution, the more likely a sampled mean will underestimate the true underlying mean. For a lognormal distribution with a CV over two, the most likely value for a sample of one is relatively close to zero no matter how big the mean of the distribution. For small samples, the expected value for the sample average will be close to zero.

To acknowledge that there are people who are uncomfortable with the idea of focusing on the mode rather than the mean, I offer some numbers that might help them get more comfortable. The mode is the point at the highest point on the probability density function. What I am going to show is the area under the probability density function for all points that have a value that is greater than the value associated with the mean. Chart 3 shows a lognormal distribution and we are interested in area "A" between the mean and the point to the left of the mode that has the same probably density function value as the mean.

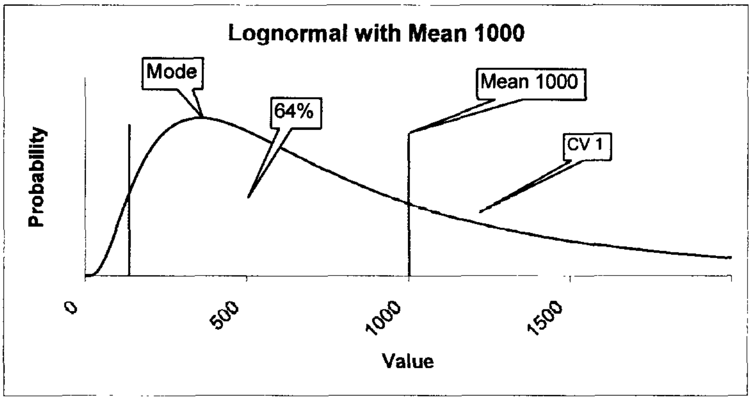

Chart 3

Table 1 shows for varying samples size averages from lognormals with different coefficients of variation the percentage of sample averages whose probability density function value is higher than the value at the mean. This is area "A" in chart 3.

Sample Size	Coefficient of Variation				
	0.5	1.0	2.0	5.0	10.0
1	47%	64%	74%	82%	86%
25	13	28	44	62	70
50	12	19	38	56	66
75	11	16	33	54	64
100	11	13	28	53	63
150	7	13	26	47	57
200	7	3	25	47	57
300	7	3	17	41	55
400	3	2	16	41	53
500	2	2	14	38	51

Table 1 Area "A" for different sample average sizes distributions and CV's

As an example, the numbers in this table says that if you are taking averages from a lognormal distribution with a coefficient of variation of 10.0, then there is better than a 50% chance that the sample average will be below the true average of the distribution if the sample size is 500 or less. I like to think that focusing on the mode makes it easy to capture a lot of this information and I hope to convince you of that with the following simulation exercises.

So what makes up a very small sample size? In order to answer this question, I simulated a single random value from a lognormal distribution with mean 1000 and varying CV's using an Excel add-in called @Risk by Palisades. The @Risk add-in has functions that will simulate random values from various statistical distributions and it has functions that will calculate statistics for the random results. Since I am sampling from a continuous distribution, it is unlikely that I would sample any single point more than once. So rather than find the most common single point, I set the program to keep track of the most common interval of width 5 as a proxy to finding the single point mode. As a check on this process and to see if @Risk actually does what it claims to do, I wanted to see if the mode for a sample size of one tracks with the formula mode of the distribution. The results in Table 2 show that the simulated results track closely to the formula mode of the distribution after 1,000,000 simulations.

CV	Simulated Mode	Formula Mode
0.5	717.91	715.54
1.0	361.67	353.55
2.0	79.75	89.44
5.0	7.86	7.54
10.0	2.51	0.99

Table 2

I then increased the sample size to 25 random values, 50 random values, 100, 200, 300, 400, 500 and 10,000 random values. I measured the midpoint of the most common interval for the average of those larger samples doing 500,000 simulations each time. (See Appendix C for additional details.) The results are shown in Chart 4 for lognormal distributions with mean of 1000 and CV's of 0.5, 1.0, 2.0, 5.0 and 10.0.

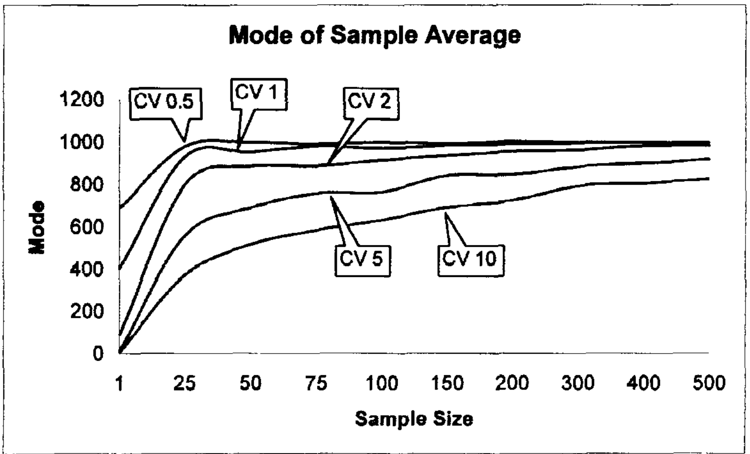

Chart 4

What does this Chart 4 tell us? For individual claim size distributions that have low CV's, the most likely value that we would see from a sample average very quickly approaches the mean of the distribution. We are getting the same result as if we were sampling from a symmetrical distribution where the mode is equal to the mean. However as the CV increases, it takes a very big sample size before the most likely value of the sample average approaches the mean of the individual claim distribution. For a distribution with a CV of 10, even at a sample size of 500 the most likely value we would see from the sample average is 85% of the distribution mean. Formal credibility formulas aside, I believe many actuaries would consider 500 homogeneous claims a fairly large database. Appendix D has charts, albeit more complicated charts, which show additional information about the entire distributions of the sample averages.

With a CV of 10 and a sample size of 10,000, the most likely value we would see is still only 96% of the mean of the distribution. William Blatcher, CFA, points out that a simulation size of 10,000 is a typical @Risk simulation size for actuaries working in reinsurance. Even at this large number of samples, there is still a downward bias of 4% from the actual average of the distribution.

Another thing to observe about these sample results is that the most likely values for the sample averages follow a pattern of rising quickly from the mode of the distribution and then hitting a fairly flat area that approaches the mean very slowly. In his book "Fooled by Randomness" [2], Nassim Taleb discusses how people are misled by skewed distributions.

He focuses on the rare extreme values in the tail of the distribution, which he calls the black swans that are usually missing from the sample results out of skewed distributions. People forget about these black swans or are unaware of them. However, for "small" samples out of skewed distributions, it is not just missing black swans that the observer can innocently miss that can cause problems. The small sample from the body of the distribution is actively misleading the observer because the mode is so much lower than the mean. It is almost as if the distribution is actively evil by feeding us misleading information from its body as opposed to passively withholding tail information from us.

What are some of the implications of this? When we are doing relatively small samples from skewed distributions, we should recognize that the most likely value of the sample average will be less than the mean of the distribution that we are trying to measure. We should adjust our sample results based on the CV of the distribution and the sample size to calculate the population mean of the sampled distribution. The correction factor should be the ratio of the population mean and the mode of the sample average. The mode of the sample average would vary by the sample size. It would equal the population mode for a sample size of one and would approach the population mean as the sample size approaches infinity. Table 3 shows the correction factors for a lognormal distribution for given sample sizes and coefficients of variation based on the simulation results. These values are just the ratio of the actual distribution mean and the mode of the sample means underlying Chart 4.

Sample Size	Coefficient of Variation				
	0.5	1	2	5	10
1	1.45	2.51	11.79	128.90	399.67
25	1.02	1.07	1.25	1.79	2.68
50	1.00	1.05	1.13	1.45	1.93
75	1.00	1.02	1.13	1.32	1.71
100	1.00	1.03	1.09	1.31	1.58
150	1.00	1.02	1.07	1.19	1.44
200	1.00	1.01	1.05	1.18	1.38
300	1.00	1.00	1.04	1.13	1.26
400	1.00	1.00	1.02	1.11	1.24
500	1.00	1.00	1.02	1.09	1.21
10000	1.00	1.00	1.00	1.02	1.05

Table 3: Correction factors

There is actually a precedent for a table of adjustment factors like this. The British

Department for Transport has recognized that planners for large public projects routinely underestimate the actual cost and time of the project [3]. To adjust for this tendency to be optimistic, project appraisers are required to make adjustments or 'uplifts" to the submitted costs, benefits and duration. The factors which depend on the type of engineering project under consideration could be up to 51% for building projects and up to 200% for IT projects. They are used to adjust the project costs to overcome this downward bias and increase them to more likely cost levels.

SMALL SAMPLE BIAS AND CREDIBILITY

When we are credibility weighting two results from skewed distributions, we should recognize that the small sample size average might be different from its population mean only because it is biased downwards. In his paper "An Examination of Credibility Concepts" [4], Stephen Philbrick presents an example of four people shooting at four different targets to help explain credibility. The diagrams in the paper show the historical results for the four shooters with their shots clustered around their four respective targets. The clustering is a simplifying assumption in order to focus on the main point of the paper. We are better able to guess who the shooter is if:

- We see more subsequent shots taken,
- The shooters are better shots or,
- The individual targets they are shooting at are moved further apart.

When the targets are widely separated and the shooters are good shots; we want to give high credibility to the hypothesis that A is the shooter when we see a subsequent shot fall near target A. This follows, in part, from the assumption that the shots are symmetrically distributed around the targets.

Now suppose a wind is blowing across the firing range affecting the results of shooter A. Most of the shots are blown away from target A and land near target B. Occasionally the wind will stop blowing and a shot will land near target A. Even more rarely, the wind will reverse direction and the shot will fall widely wide of the target on the other side. On average all the shots fall around target A. In this example, even if the means and standard deviations of the distribution of shooters has not changed from the symmetrical example, whatever standards we may have created for credibility when the shots were symmetrically clustered around the target have to be increased given that the distribution of shots is skewed. We need more shots, or the shooters have to compensate for the wind to improve their aim, or the targets have to be much further apart to achieve a given credibility standard when we are dealing with skewed distributions as opposed to symmetrical.

In this second example, if we had a small sample from this skewed distribution, we probably would have given little credibility to the idea that the shooter was A unless we really understood the process involved. If we assumed we were dealing with symmetrical distributions, we most likely would have concluded that the shooter was B since the small sample of shots would most likely have been grouped near the mode of the distribution – target B. It is important to understand what type of distribution we are working with and avoid convenient assumptions.

OUTLIERS AND THE ART OF THE DEAL

James MacGinnitie in his Address to New Fellows at the November 2006 Casualty Actuarial Society annual meeting stated that the world is not normal and warns against unexamined use of the bell curve or normal distribution as a model. In his book "The (Mis)Behaviour of Markets" [5], Benoit Mandelbrot of Chaos Theory fame discusses problems with assuming the financial markets behave according to the normal distribution. Why was the normal curve used in the first place? Mandelbrot states that at one time all of nature was assumed to behave according to the bell curve -- that is why it is called the normal curve. Independent observations from a normal curve are not an appropriate model for many financial market situations even though modelers have historically used normal curves. Actual observations come from more highly skewed correlated distributions. Nevertheless, many financial modelers say, "So what?" They argue that the normal curve is a convenient approximation and as long as this assumption does not cause any problems, then we should just ignore any theoretical refinements. MacGinnitie and Mandelbrot claim that assuming normal, independent observations does cause problems by underestimating the true risk associated with skewed processes. Mandelbrot demonstrates that Chaos Theory yields a better model of the financial market's behavior. He says that you cannot make money with this insight but he does assert that it allows you to understand better the risk that is involved which could help you avoid losing money.

If someone makes a statement it is good to check it if we are able. Can we underestimate the risk by using standard statistical techniques on small samples? Table 4 shows the most likely indicated CV from different sample sizes from a lognormal simulation with varying CV's.

Sample Size	Coefficient of Variation of Sampled Distribution				
	0.50	1.00	2.00	5.00	10.00
25	0.46	0.79	1.02	1.63	1.81
50	0.47	0.85	1.31	1.82	2.21
75	0.48	0.89	1.40	2.04	2.66
100	0.48	0.89	1.41	2.16	3.00
150	0.48	0.92	1.5	2.27	3.00
200	0.50	0.93	1.56	2.44	3.01
300	0.49	0.94	1.61	2.55	3.02
400	0.49	0.96	1.63	2.97	3.17
500	0.50	1.00	1.65	2.98	3.67
10,000	0.50	1.00	1.91	3.82	6.00

Table 4 Mode of the Sampled CV's

For small samples from skewed distributions, the most likely value for the CV underestimates the CV of the actual distribution since we are missing the values from the tail of the distribution. If we used these numbers as parameters for our models, then we would underestimate the risk of the situation we are modeling.

There is another reason to avoid assuming the normal distribution either consciously or unconsciously. Besides worrying about the tail of the distribution, we also have to worry about the body. In actuarial and financial work, we have to avoid assuming that "small" sample averages from skewed distributions will give unbiased indications of the true mean of the underlying distribution as they would if we were sampling from a normal distribution.

Actuaries sometimes go out of their way to create problems, for example, by creating smaller and smaller data samples as opposed to maintaining larger groups for large samples. I have seen reserve studies that will take a book of business, split it into 34 different rating groups, and then split each of those into three different currencies for over 100 different groups to study. By doing so many splits of data, we end up creating small sample averages where the results will be biased low. A few groups may have large claims but those are ignored as aberrations as opposed to being recognized as the extreme values from a skewed distribution. If only they were combined with the other sample values in larger groups, then there would be a better chance of yielding a more appropriate estimate of the true population mean.

Is there a psychological explanation for why people disregard outliers? One of the explanations for our tendency to disregard outliers has to do with our training. Students are taught that,

> *"An outlier is a point which your data set is better off without. If you can prove your point better by ignoring some small portion of your data, why not ignore it? It's probably just a blunder on the part of the person collecting data, or some special, irrelevant circumstance that we needn't investigate in detail."* [6]

I hope everyone is appropriately shocked with this advice and will acknowledge they have never followed it in the past nor will they ever follow it again. MacGinnitie and Mandelbrot strongly recommend not ignoring outliers. Taleb and others argue that properly accounting for outliers is how to win or lose the big money. Helping customers deal with outliers is what insurance is all about.

Ignoring outliers could be instinctual. As the herd moves on, the weak, the old and the sick fall behind becoming outliers to be picked off by the wolves. The clustering illusion is an identified psychological bias where people will pick out patterns even when none exist [7]. Because of this bias professionally designed standardized tests do not have long runs of a particular multiple choice answer. Students would feel such a pattern is unlikely and then feel pressured to answer incorrectly just to break the run.

Certainly actuaries make their living by finding and identifying patterns. Once a theory is formed about a particular pattern the confirmation bias in psychology is the tendency for people to search for or interpret information to confirm one's preconception [7]. Outliers don't fit the pattern and they don't support the basic idea that's being proposed, so ignore them. What is even worse, the more outrageous the outlier the more likely we are to throw it out of the sample. We can put all these biases all together to explain why people ignore outliers and call it the "Simon and Garfunkel Bias" – still the man hears what he wants to hear and disregards the rest [8].

Certainly, there may be business reasons for a person to leave an outlier unexamined when pricing a deal. There are no absolute rules. And for certain parties in the transaction, it is in their best interest to deemphasize the outliers. The negotiation skills and dedication of the brokers and market makers influence the final price. The best dealmakers that I have seen in action are those that continually work on the ego of the person they are trying to sell. The skilled insurance broker will set up a situation where the rejection of the proposed deal at the suggested price implies the insurance underwriter lacks cojones; they are not a real player. The broker will threaten that they have other underwriters at other companies – real business people – that are ready and willing to do the deal. Eventually, the ego driven underwriter will be dying to do the deal in order not to appear weak in front of the broker.

It doesn't matter about outliers. The broker now has the underwriter working for them in finding a way to do the deal rather than the underwriter working for their employer. This whole process is a beautiful thing to behold. You really have to admire a good broker at work.

This is where an objective actuary can be a valuable asset in these negotiations. Actuaries typically get their ego kicks from doing a thorough analysis and beating other people (either in competitive exams or in doing accurate forecasts). If another actuary has arrived at an estimate that is lower than your estimate but cannot give a satisfactory reason for why your answer is too high, then you will stubbornly stick to your result. This can be a valuable sanity check for the underwriter when evaluating a deal. Whether the deal finally gets done or not at a particular price will depend on many things. Ego is involved in a complex interaction with many different forces; forces that will vary from company to company. The actuary can be a big assistance in providing a quantitative estimate that takes into account all the available information.

Actuaries are also subject to ego problems and can be a liability to the process. Forecasts of indicated prices are the appropriate combination of all available information including outliers. As a deal is negotiated, very often new information is introduced that was not available when the first price forecast was produced, for example, a legitimate explanation of the outlier. That new information could cause the forecasted price to go up or it could cause the forecasted price to go down. It has been my experience that actuaries are more willing to allow their prices to go up rather than to go down based on new information. Part of this might be the natural reluctance to lower a price based on the suspicion that only good information is being shared and none of the bad information. Some of it might just be misplaced pride in that changing an answer somehow implies that the original forecast was wrong. Some of it might just be psychological.

There have been experiments done asking people to guess a particular number when they have no idea what the appropriate answer is. For some reason the first number that people hear sets the magnitude of the perceived correct answer whether or not it is anywhere close to the correct answer. This is known as the anchor effect [7]. All future answers will be judged against this initial answer. For example, what is the population of Brazil? Someone might throw out a guess that the population is 40 million people. It sounds like a reasonable number. From that point on people will be evaluating future answers to the question, including the correct one, based on this initial guess. And future guesses will tend to fluctuate in the neighborhood of this initial guess. (What is your guess for the population of

Brazil?*)

Are actuaries subject to the anchor effect? Are they more likely to be subject to it when they are the author of the original forecast or guess? Everyone else is affected by it so why not actuaries? If you have been around long enough, you have definitely seen this process in action. An early number sets the value for a deal, a transaction or an acquisition. Based on that early number a decision is made to do the deal or not. From that point on it doesn't matter what new information is brought forth and how the numbers change. A decision has been made and a course of action is in motion. The first numbers that are released are very important because those may be the last numbers anyone pays attention to.

The actuary has to bring a forecast to the table that reflects all the information available including outliers. A forecast is different than a prediction of the future. If data might have been withheld that can influence the answer or even if appropriate data are not available to do a proper forecast such as with a small sample from a skewed distribution, then that inadequacy of the data has to be built into the pricing of the deal. Exactly how the price is adjusted for the lack of the data is a judgment call. But that judgment call is made by those responsible for the deal. The actuary has to be upfront with the indicated forecast based on the information available and also explicit about any additional loadings in the price that are due to the quality of the data.

BLESSED ARE THE LOSERS

Economists are concerned with a problem called the winner's curse [9]. In this problem several bidders are competing for an item in an auction and the winner will be the highest bidder. This item is worth the same to all the bidders. The bidders only have incomplete information about the true value of the item and they have to make estimates about this value to prepare their bids. The average of all the bids is assumed to be around the true value of the item. If this is the case, the winner will tend to lose money since they will bid more than the item is worth – the winner's curse. Savvy bidders will avoid the winner's curse by bid shading or quoting a price below what they believe is the value of the item. The bidder who follows this strategy will lower their chance of winning a particular auction but increase their expected return over time. This is the ideal situation as described by economists.

For the reality of the insurance world we have to make some changes as the problem is more complicated. One change is the winner is the lowest bidder not the highest. Another

* The population of Brazil was reported to be 188 million people in 2006.

difference is that a particular account might be more valuable to one insurer than to another insurer for a variety of possible business reasons. An account might have value outside of its expected profit. For example, a company might have written premium goals and they will set the price to win the account as opposed to setting the price at the estimated value of the account. The winning bid does not necessarily lose money even if the bid is below the long term average cost of the account because the annual cost of an account is not fixed. An insured might have a good year and experience low losses for the year of the auction. Another difference is that the result of the traditional auction problem is known immediately but for an insurance contract, it could take years for the actual result to be known. Because there is this lagging feedback mechanism, inappropriate pricing might persist for a period of time as opposed to being corrected immediately. I would think these differences would tend to increase the monetary losses of the winner. A difference that would tend to ameliorate the losses of the winner is that all insurance companies are not all created equal. In real life, the insurance buyer bases their purchase decision on more than just price. They may go with a higher priced policy if they expect to get better service from a particular insurer.

Participants in the traditional auction problem are assumed to have incomplete advanced information about the value of the item being auctioned. This is certainly true for insurance. We also have the potential additional problem associated with small samples from skewed distributions. Gary Blumsohn points out that the more skewed the distribution, the more likely it is that bidders will be quoting prices based on downwardly biased sample averages and thus the winner's curse will be compounded.

Bidders in this case should build into their decision process how skewed the loss process is and how much actual loss information is available to price the account. The more biased the actual losses or the smaller the pool of available information on the particular account or similar accounts, the more we should be concerned about adding a charge to our bid to compensate for any potential biases in the available information. In some extreme situations, we might just want to quote a "go away" price to the risk or broker. That would be a situation where we are very uncomfortable with the risk involved and/or the deficiencies in the available pricing information. If we are concerned about pricing it too low, then we should just quote a price high enough so that it is unlikely to be accepted but high enough so that we still feel OK if for some reason we are the winning bidder.

If one is faced with the risk of an event that is likely to be a small sample from a highly skewed loss distribution, then there are a few things that can be done to improve the situation. The first is to increase the sample size and combine that risk with other risks to take advantage of the law of large numbers. If you are a single insured, you buy a policy from an insurance company who does the combining of risks for you. If you are an insurance company you can expand your writings until you have sufficient volume to produce stable results. As an insurance company, if it is impossible to combine the presented risk with a sufficiently large number of other independent risks, then the other alternative is to reduce the skew of the risk distribution. You can reduce the limits on the policy, restrict policy terms, or buy some form of appropriately priced reinsurance. Finally, the last thing that you can do is reduce the probability of the event to zero by not writing the risk at all.

A SMALL SAMPLE OF SKEWED RANDOM THOUGHTS

More than once, I have heard a story at a luncheon at a Casualty Actuarial Society meeting about either a start up company or a new branch of operation where the initial loss experience is good. The stories deal with heroic battles between an actuary and the naïve management team. The actuary wants to hold surplus and maintain high rate levels in anticipation of losses yet to come. Management wants to cut rates or pay out large dividends based upon the small but exceptional experience to date. In the stories that I have heard, either the actuary wins out or the company barely survives its first few years. Because those are the only endings that I have heard, I have to assume that there is a survivorship bias in these results – only the survivors are happy to share their stories.

The heroes of these stories recognized that skewed distributions give biased results not just due to small sample sizes but also because the mode is seen before the average result. Incremental claim reports follow skewed patterns. Once people in a company see incremental claim reports from a particular accident or policy year declining after the mode of the distribution has passed, they might think the worst is over and that claim reports will drop off as fast as they appeared. However, the tail that follows the mode could stretch out for years. Actuaries who have recognized this and have convinced their colleagues of claim reports yet to come have the right to boast [10]. Actuaries think accident year; everyone else thinks calendar year.

Speaking of start up companies, another lunchtime conversation has to do with the strategy of starting a reinsurance company devoted to catastrophes. The question is whether the company will be able to build capital by surviving its first year without sustaining a catastrophe. These companies are insuring events from highly skewed distributions. The most likely loss that they will see is zero. Chance is in their favor that they will survive the first year. This same type of thinking could explain why some investors are willing to rush in and refinance catastrophe reinsurance companies after they suffer a particularly bad season. The hope of stock price recoveries from big premium increases following the cat loss also has something to do with it.

In case the traditional age to age development method has not been beaten up enough in recent papers, what about the way we average development factors? In the past, I have used methods that take the average of the most recent x development factors excluding the high and the low value. I have also heard of methods that use x binomial factors as weights to apply to the most recent x development factors ranked from highest to lowest to get a weighted average estimate. It certainly sounds like these methods are making an implicit assumption that we are sampling from a symmetric distribution if not a normal one. However, if these unquestionably small sample averages of individual loss development factors are from a skewed distribution then these methods are throwing out or downplaying important information.

We become complacent about our safety or survival from repeated exposure to threatening situations that do not actually happen. Psychologists call this habituation [11]. When I worked in Jersey City, NJ, I would pass a good example of habituation in action every day when I went to work. Each morning I would drive past Nunez Restaurant on the corner of Montgomery Street and Jordan Avenue. The owner of the restaurant had put a couple of plastic owls on the ledge of his building to scare away the pigeons and save his customers untold embarrassing problems. However, the pigeons had become habituated to the owls. Initially the model of their natural predator would have scared the pigeons away but when nothing threatening ever happened, the pigeons learned to suppress their natural instinct to be afraid. So many pigeons have become habituated at Nunez Restaurant that on some mornings the corner could be used for the Jersey City/pigeon remake of Alfred Hitchcock's "The Birds". The pigeons have become conditioned to the mode of nothing happening and suppressed their fear of the extreme event of being attacked by the stationary plastic owls. (Some people might argue that my driving to work every day down Montgomery Street in Jersey City is a good enough example of habituation.)

People are not to be outdone by pigeons. Here is a possible explanation for riding a motorcycle without a helmet. People are probably encouraged to do any high risk activity because they get something out of it and, most likely, not suffer any consequences for at least the first few times. The loss distributions associated with any high risk activities are skewed – driving without a seat belt, cave diving, painting the outside of a house without securing the ladder. For a small number of trials, any individual is most likely to experience the mode of the distribution and not suffer any consequences. Based upon the lack of any immediate losses, the individual grows complacent, ignores any warnings, and continues with the activity. Yet the possibility that the individual will experience a loss increases over time as the sample size increases.

The difference between the mode and the mean in skewed industry loss distributions might be a contributing factor to the insurance cycle. The distribution of annual industry results appears skewed to me. If that is the case, the most common result for the industry during the cycle period will be something closer to the mode of the distribution. All other things being equal, competition will keep pricing levels below the mean of the distribution as people grow complacent and the sky doesn't seem to be falling as constantly predicted by actuaries. Every once and a while there will be a major industry loss event. The industry will feel the cash flow shock because it was pricing below the long-term average and it will overcorrect above the long-term average when it reacts. Mix in some skewed distributions associated with the asset side of the balance sheet and away we go.

Along these lines, Ted Kelly, CEO of Liberty Mutual was quoted in the November 27, 2006 [12] issue of the National Underwriter warning about pricing levels in the 2006 property market. Property insurance prices had increased dramatically in 2006 because of the losses associated with Hurricane Katrina in 2005 and presumably due to the early predictions by the hurricane forecasters of severe hurricanes for 2006 and beyond. He said, "The lack of catastrophes this year will create its own set of problems, including accusations that we cried wolf when we raised rates and are now price gouging." He joked that, "It's like saying someone who survives Russian roulette faced no risk just because the gun didn't go off, when we all know there is still a bullet in the chamber, and if you play the cat game long enough, it's going to go off." In my opinion, using the best estimate of the loss over the period in which the policy is exposed would be the correct way to fund for catastrophes. Currently, all the market forces seem to produce a collective behavior that is influenced by the results of small sample averages and then plays catch up after a major shock loss. If nothing else, funding at the best loss estimate for the exposure period would identify the costs that that market is facing. That being said, what is the loss distribution and what is the best cost estimate are among the difficult questions that all the participants in this market have to answer.

The only cure for complacency is a conscious effort to take measures guarding against extreme events. Insurance companies exist to help our customers guard against extreme unexpected financial consequences of life. As actuaries and managers of insurance companies, we have to make sure we are forecasting the true long-term results and acting appropriately to account for extreme events so that our companies will be there to pay the losses of our customers. We have to avoid complacency bred by constant exposure to the mode of distributions.

A quote attributed to former first lady Barbara Bush is, "Bias has to be taught." She was speaking about prejudices and that children learn prejudices from adults as opposed to being born with any biases. If we speak about the statistical bias of small sample averages, there are whole hosts of places where subtly we are being misled. No one is teaching us these things. We are forming theories based on small samples and then forgetting or not realizing that those theories might be wrong since they are based on small samples. Is that not a definition of prejudice – developing a theory and then forgetting that it is a theory and assuming that it is fact? It would have been better if Barbara Bush had said, "Bias has to be fought".

There is an old insurance joke that says an insurance company is a car being driven down the road by the blindfolded president of the company. The head of marketing is stepping on the gas, the underwriter is stepping on the brake and the actuary is looking out the rearview mirror yelling which way to turn. To paraphrase the warning label that appears on the passenger side rearview mirrors of US cars, those loss estimates that the actuary sees are larger than they appear.

In that joke, the actuary is the only person in the car who is looking at any section of the road. When working with small samples from skewed distributions, we should keep in mind that it might take many samples in order to get an average that provides a good estimate of the true average of the underlying distribution. We have to understand the loss process we are trying to model along with the limitations of our data samples, and make forecasts and recommendations accordingly.

Acknowledgements

I would like to thank William Blatcher and Larry Schober for their valuable feedback on earlier drafts of this paper. Also I would like to give many thanks to the reviewer of this paper Gary Blumsohn.

Appendix A

Counter Examples Where the Mode of a Distribution is Larger than the Mean of a Distribution

There is more than one definition of the skew of a distribution. The skew of a distribution is usually calculated as the third central moment of the distribution.

$$\frac{1}{n}\sum\left(\frac{\left(x_i-\overline{x}\right)}{s}\right)^3$$

In this formula, n is the sample size, $\overline{x}$ is the sample mean and s is the sample standard deviation. A positively skewed distribution has a longer tail to the right.

In Excel, the formula for the skew of a distribution is the following:

$$\frac{n}{(n-1)(n-2)}\sum\left(\frac{\left(x_i-\overline{x}\right)}{s}\right)^3$$

A distribution may have more than one mode but for this discussion I am going to assume that we are dealing with distributions with only one mode.

Using the Excel definition, it is easy to construct a discrete distribution that is skewed to the right and the mode is greater than the mean [13].

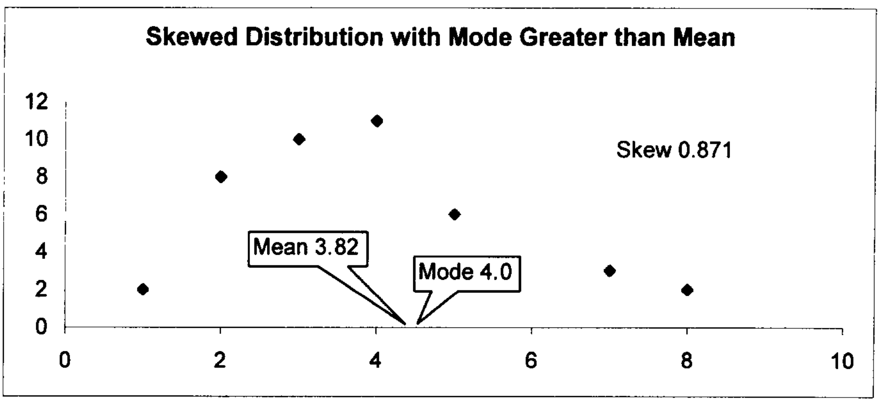

Chart 4

One way to remove the counter intuitive examples is to define them away. The Pearson mode skewness of a distribution is defined as:

$$\frac{\text{mean - mode}}{\text{standard deviation}}$$

Using this definition, a positively skewed distribution would always have the mean higher than the mode.

Appendix B
Some Formulas for the Lognormal Distribution

As a reference or reminder, the probability density function for a lognormal distribution with parameters $x < 0$, $-\infty < \mu < \infty$, and $\sigma > 0$ is the following [14]:

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

The mean is equal to:

$$e^{\left(\mu+\sigma^2/2\right)}$$

The variance is equal to:

$$e^{\left(2\mu+\sigma^2\right)}\left(e^{\sigma^2}-1\right)$$

And the coefficient of variation (CV), the standard deviation of the distribution divided by the mean, is equal to:

$$\sqrt{\left(e^{\sigma^2}-1\right)}$$

Formulas that actuaries are probably not familiar with are the formulas for the median:

$$e^{\mu}$$

The formula for the mode:

$$e^{\left(\mu-\sigma^2\right)}$$

And the formula for the skew of the distribution is the following:

$$\sqrt{\left(e^{\sigma^2}-1\right)}\left(2+e^{\sigma^2}\right)$$

The lognormal is positively skewed for all values of σ.

The formula for the ratio of the mode to the mean as a function of the CV is the following:

$$\frac{\text{mode}}{\text{mean}}=\left(CV^2+1\right)^{-\frac{3}{2}}$$

Appendix C
Simulation of Results

For these simulations, I used @Risk Version 4.05. I did different sample averages for different sizes 1, 25, 50, 75, 100, 150, 200, 300, 400, 500 and 10,000. For each sample size I took a sample average of that many random values of the lognormal function and for each different sample size I used different random values. I did not generate 10,000 random values and then take the average of the first 25, the first 50, etcetera.

At each simulation, I selected as an output the RiskMod function to measure the mode of the simulated sample averages. Because we are dealing with a continuous distribution, I checked for the midpoint of the most common interval of width 5 as opposed to the most common single value. Excel/@Risk formulas for sample size 5 are below in Figure 1.

	A	B
1		
2	Mean	Standard Deviation
3		
4	1000	2000
5		
6	1	=RiskLognorm(A4,B4)
7	=+A6+1	=RiskLognorm(A4,B4)
8	=+A7+1	=RiskLognorm(A4,B4)
9	=+A8+1	=RiskLognorm(A4,B4)
10	=+A9+1	=RiskLognorm(A4,B4)
11		
12		Average 5
13	=RiskOutput() + TRUNC(AVERAGE(B6:B10)/5,0)*5+2.5	=+RiskOutput()+RiskMode(A13)
14		SD 5
15	=RiskOutput() + TRUNC(STDEV(B$6:B10)/5,0)*5+2.5	=RiskOutput() + RiskMode(A15)
16		CV 5
17	=RiskOutput() + TRUNC((+A15/A13)/0.01,0)*0.01+0.005	=RiskOutput() + RiskMode(A17)
18		

Figure 1

Appendix D
Contour Maps of Sample Average Distributions

In the paper, I showed charts on how the mode varies with the CV and how the mode varies in relation to the mean as the sample size varies. I wanted to get a feel for how the distribution of the lognormal sample average is shaped at different sample sizes. Even though the mode of the distribution is the most likely result, I was interested in whether it is really more likely than other values or just marginally more likely. Here are the results using standard output from @Risk and Excel graphing routines.

Charts 5 and 6 are contour maps of the distributions of sample averages of lognormal distributions with CV's of 1 and 5 as the sample size varies. These are two-dimensional representations of three-dimensional surfaces. If you are familiar with topographical maps, then think of these charts in the same way.

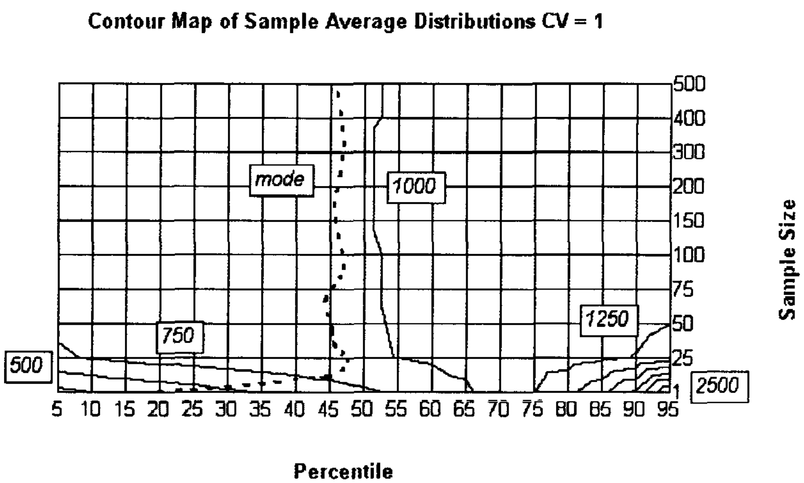

Chart 5

Reading across the gridline on Chart 5 for a sample size of 25 from left to right, the 5th percentile point is on a slope between the 500 and 750 contour lines. At the 10th percentile point, a wide gentle plateau begins going from 750 up to 1250 at about the 85th percentile point. The mode appears approximately the 47th percentile point and the mean is close by at the 54th percentile point. After we pass the 85th percentile, the distribution approaches a steeply rising area signified by the contour lines getting closer and closer. The 95th percentile point looks like it is just under a value of 1500. If we are sampling from a distribution with a

CV of 1 and the sample size if 25 or above, it is not going to be a problem that the mode of the average is lower than the mean. The values are close together straddling the median value at the 50th percentile line. It looks like we could get away with an assumption of a symmetrical distribution.

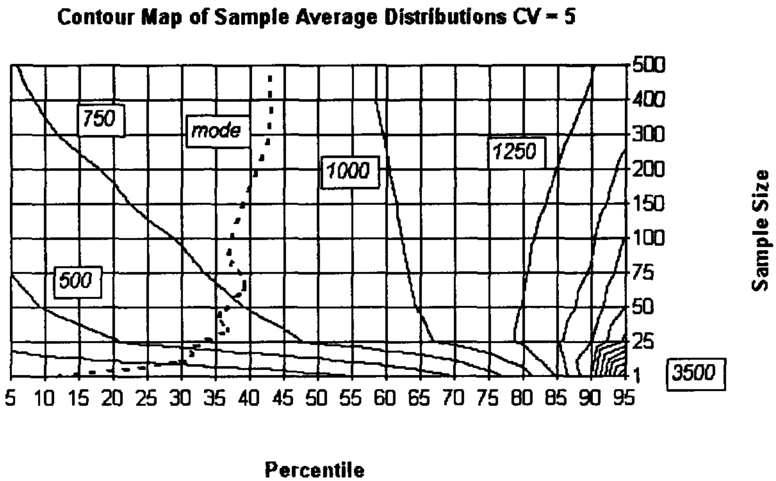

Chart 6

When the CV is 5 as in Chart 6, the map gets to be a little more interesting. If we look across the 25-sample size gridline again, the fifth percentile point is very close to the 250 contour line. Now rather than a gentle plateau stretching across the graph, we have a steadily increasing slope going across the chart up to the 1250 contour line at the 78th percentile. Here the surface starts to increase more steeply reaching just above the 2250 point at the 95th percentile. On this chart, the mode and the mean are widely separated for a sample size of 25. Approximately 67 percent of all the values of the distribution are below the mean. The mode looks like it is situated right in the middle of those values at the 33rd percentile. There are definitely higher values that are likely to occur out on the higher percentile part of the curve.

As the sample size increases, the situation is not as clear-cut that we can get away with a symmetrical distribution approximation as it was when the CV was lower. Here the 1000 contour line stays above the 60th percentile until the sample size reaches 250. Even at a sample size of 500, the contour line for 1000 is above the 57th percentile while the mode is at 43rd percentile.

These charts also show something interesting about company funding and size. Suppose we are collecting exactly the expected losses from each insured and we want to have enough surplus in the first year so that we have a 90% chance that surplus will not go negative just due to loss fluctuations. Using Chart 5, we can see for a small company producing only 25 claims it would need an additional 750 of surplus for each claim resulting in a premium to surplus ratio of approximately 1.33. A larger company producing 500 claims could get by with a premium to surplus ratio of approximately 4.0.

REFERENCES

[1] Hewitt, Charles C. Credibility for Severity Proceedings of the Casualty Actuarial Society Casualty Actuarial Society 1970 Vol. LVII pp.148-171

[2] Taleb, Nassim Nicholas. Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life. New York: Texere LLC 2001

[3] Flyvbjerg, Bent, "Procedure for Dealing with Optimism Bias in Transport Planning." [Online] Available http://flyvbjerg.plan.aau.dk/0406DfT-UK%20OptBiasASPUBL.pdf 2007

[4] Philbrick, Stephen W. An Examination of Credibility Concepts. Proceedings of the Casualty Actuarial Society, 1981 Vol. LXVIII pp. 195 – 212

[5] Mandelbrot, Benoit, and Richard L. Hudson. The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin and Reward. New York: Basic Books, 2004

[6] Mittledorf, Joshua. Response to, 'What is the definition of outlier?' [Online] Available http://www.mathforum.org/library/drmath/view/52801.html 2007

[7] Wikimedia Foundation. List of Cognitive Biases [Online] Available http://en.wikipedia.org/wiki/List_of_cognitive_biases 2007

[8] Simon, Paul. "The Boxer" 1968 [Online] Available http://www.lyricsfreak.com/s/simon+and+garfunkel/the+boxer_20124664.html 2007

[9] Wikimedia Foundation. Winner's Curse [Online] Available http://en.wikipedia.org/wiki/Winner's_curse 2007

[10] Weissner, Edward W. Estimation of the Distribution of Report Lags by Method of Maximum Likelihood. Proceedings of the Casualty Actuarial Society, 1978 Vol LXV pp. 1 – 9

[11] Wikimedia Foundation. Habituation [Online] Available http://en.wikipedia.org/wiki/Habituation 2007

[12] Friedman, Sam. "Top Dogs Barking." National Underwriter P&C, 27 November 2006, p. 5.

[13] von Hippel, Paul T., "Mean, Median, and Skew: Correcting a Textbook Rule." [Online] Available http://www.amstat.org/publications/jse/v13n2/vonhippel.html 2007

[14] Hogg, Robert V. and Stuart A. Klugman. Loss Distributions. New York: John Wiley & Sons 1984