ERM at the Speed of Thought:

Mitigation of Cognitive Bias in Risk Assessment

Damon Levine, CFA, CRCMP

Abstract

While there is no shortage of articles, white papers, books, and model frameworks for enterprise risk management (ERM) the majority of them share a common omission in their discussion of key risk categories: cognitive bias. The very fact that ERM is driven by risk assessments created by the human mind means that bias is naturally embedded into estimates of risk impact and likelihood. At a deep level, inconsistency and fallacy are hardwired into our brains. This paper provides an overview of those cognitive biases most often responsible for flawed risk assessments and provides practical techniques to mitigate them.

Keywords: ERM, cognitive bias, risk assessment, extreme value theory, Delphi method, black swans, stochastic models

1. The Surprising Frequency of Low Probability Events

Having just missed their train, the two logicians sit side by side on a wooden bench in a gothic train station. They have a long wait ahead. Alan and Kurt had carefully planned the excursion for their day off from the university but both had forgotten to move their clocks ahead for daylight saving time. This oversight belies two powerful intellects. They both know enough mathematics and probability to be dangerous, boring, or harmless depending on whom you ask. At the beginning of a debate between them, neither can be sure who will emerge as the persuader and who will end up being persuaded.

Alan: can you believe that earthquake in Chili?

Kurt: yes, it was unprecedented apparently

Alan: there are so many of these so-called black swans these days!

Kurt: it's not so surprising…even if we are talking very low probability events

Alan: what do you mean?

Kurt: I guess I have two thoughts. The first is more of a question I suppose…what is the probability of a 1 in 100 year event occurring in 100 years?

Alan: I think I understand what you're asking in general…given an event with annual probability 1/N, what is the probability that over a period of N years, the event occurs at least once?

Kurt: correct, the other is also a question too I just realized. Suppose we identify 20 low probability events and they are independent. Perhaps they each have annual probabilities of 1/100. In a year what is the probability of at least one of them occurring?

Alan: I need to think a bit about these

Kurt: take your time my friend…we have plenty of it

Expecting the Unexpected and an Appearance from e

Consider a very unlikely event with annual probability 1/n. Perhaps this event is regarded as a "deep tail" scenario and an event of such impact does not exist in the historical record. This means we may think of n as a large number such as 100 or 500. In a given year the probability the event does *not* occur is 1-1/n. Assuming year to year independence the probability that it does *not* occur over a period of n years is $(1-1/n)^n$. As n approaches infinity this expression approaches $1/e \approx 0.37$ and therefore, for large n, the probability the event *does* occur in n years is approximately $1 - 1/e \approx 0.63$.[1]

This convergence is relatively fast so the value of n need not be very large for the approximation to work well. This means that for an event with (annual) probability of 1/50, in 50 years the probability of the event *occurring* is approximately 1 - 1/e or about 63%. Similarly, the same can be said for an event with probability of 1/100 over a 100 year time horizon. This approximation works well for large n (e.g. n>50) and *any* "1 in n year" event with year to year independence.

We now turn to the related question. Assume we have N independent, potential events each having probability of p. In a year what is the probability of at least one of them occurring? Using similar reasoning the answer is $1 - (1-p)^N$. If N= 20 and p=1/100 (so the events are regarded as having low likelihood) the probability of at least one occurring in a single year is about 0.18 or 18%. In 10 years the probability that at least one occurs is about $1 - (1 - 0.18)^{10}$ or 86%.

The practical take-away from this is that if an ERM framework identifies 20 low likelihood risks which are independent (or "close enough" from a practical viewpoint) then in a given year we should not be very surprised if one of these "unlikely" events occurs. Furthermore, over a relatively long time horizon we should really be ready for it!

Modeling Black Swans: A Rebuttal to Nassim Taleb's Ludic Fallacy

The ludic fallacy is a term coined by Nassim Taleb in his 2007 book *The Black Swan*. He uses this term to invoke the Latin noun *ludus* referring to games or play. He views mathematical models that attempt to forecast or quantify future results as deeply flawed and doomed to fail. He goes on to say that statistical models are better left for casino gambling and other well defined games of chance. He dismisses models based on empirical data as flawed because, in his view, they are not be able to predict large impact events which have not been previously observed. In other words, he eschews nearly all mathematical business models because, he claims, they cannot model black swans. Taleb may have been overreaching on this last point.

The above is certainly true for many models but it may be due to a flaw with the model parameters or chosen modeling approach. Indeed it is possible to make statistically sound inferences about future observations which are worse than any previously seen. Two possible approaches to such modeling are the use of Extreme Value Theory (EVT) and the application of Chebyshev's inequality.

EVT is a branch of statistics dealing with extreme deviations from the median of probability distributions. Under very general conditions one of EVT's mains results, the Pickands-Balkema-de Hann theorem (PBH), describes observations above a high, fixed threshold as a generalized Pareto distribution (GPD). Given a set of historical data, one may choose a high threshold T within that data (e.g., 95th percentile) and then examine the *excesses* above T for the subset of observations greater than or equal to T. That collection of distances or excesses (positive real numbers) can be well modeled as a GPD which contains two parameters which are relatively easy to estimate. The resulting GPD is capable of modeling, in a statistically sound manner, the potential magnitude and likelihood of future observations which are worse than any previously seen. The PBH theorem applies to an extremely large family of distributions and an example of this approach applied to corporate bond returns can be found in the CIA/CAS/SOA Joint Risk Management Section newsletter.[2]

Chebyshev's inequality is another result which is powerful in that very few assumptions are needed for its application. For *any* random variable X with finite expected value μ and finite non-zero variance σ^2 we have for any real number k>0,

$$P\,(|X - \mu| \geq k\,\sigma) \leq 1/k^2$$

The inequality is only useful for k >1 because otherwise the right hand side is larger than 1 and it only says the probability is bounded above by 1. For k>1 is states that the probability of a realization of X being at least k standard deviations away from the mean is at most $1/k^2$. For example, for an arbitrary random variable with finite expected value and finite non-zero variance ("typical"), we can make the practical statement that for large sample sizes the observed portion of observations 3 or more standard deviations away from the mean is at most 1/9. This allows us to assign probabilities for observations in various tails. Many of us have intuition regarding tails that has been shaped by the familiar bell curve of the normal distribution. (The pun was only somewhat intentional.) The upper bound of $1/k^2$ in the inequality helps us understand the possibly higher likelihood of deep tail events for unspecified but typical distributions. We can therefore avoid what could be called a "Gaussian bias".

In many cases we are interested solely in the events from the left tail and the following one sided version of the inequality may be used for typical distributions and any k>0:

$$P\,(X \leq \mu - k\,\sigma) \leq 1/(1+k^2)$$

Use of either PBH or Chebyshev's inequality in risk models allows for modeling of black swans in a rigorous manner. PBH tends to provide better results when a large data set is available.

A good model is a map of sorts. It helps one to understand a particular area of interest: the "territory". The model is not meant to fully capture reality any more than a map reflects all details of the territory. The Polish-American scientist and philosopher Alfred Korzybski remarked that "the map is not the territory". Louis Carroll put things somewhat less seriously in *Sylvie and Bruno Concluded*, with his description of a fictional map that had "the scale of a mile to the mile". A character notes some practical difficulties with such a map and states that "we now use the country itself, as its own map, and I assure you it does nearly as well." The (serious) point to remember is that a model can be useful even though it does not perfectly mimic reality or perhaps comes up short in black swan prediction.

In his book, The Failure of Risk Management, Douglas Hubbard makes a good point when he cites this quote of DC based consultant Jerry Brashear: "A successful model tells you things you didn't tell it to tell you." There is nothing to prevent a model from offering surprises to its own builders even though the model assumptions, parameters, and logic were well known before the model was run. Hubbard describes a model he produced for the Marine Corps which used stochastic modeling to forecast battlefield fuel usage. The model provided the far from obvious result that road conditions on the main supply routes were much better predictors of fuel use than the chance of enemy contact. He mentions that "such revelations are even more profound (and helpful) when we use empirical measurements that themselves had surprising results."

2. Why is Probability So Hard?

The vast majority of today's physicists believe that the universe is, at its most fundamental level, largely based on chance. This idea underpins quantum physics and the study of subatomic particles, the building blocks of all matter within and around us. That being said, the probabilistic aspect of our everyday existence is well hidden from humans as we live on the macroscopic level and, strangely enough, our intuition regarding probability is very often flawed.

Neglect of Probability

The tendency to completely disregard probability when making a decision under uncertainty is an all too common tendency in our decision making. Events with perceived low likelihood are typically either neglected entirely or hugely overrated and outcomes with relatively large probabilities may also be somehow distorted in our view. The continuum between the extremes may be ignored.

An examination into instances of this bias in children was conducted in 1993 by Baron, J., Granato, L., Spranca, M., and Teubal, E. and included posing the following question[3]:

Susan and Jennifer are arguing about whether they should wear seat belts when they ride in a car. Susan says that you should. Jennifer says you shouldn't... Jennifer says that she heard of an accident where a car fell into a lake and a woman was kept from getting out in time because of wearing her seat belt, and another accident where a seat belt kept someone from getting out of the car in time when there was a fire. What do you think about this?

One of the study's authors noted the following exchange:

> A: Well, in that case I don't think you should wear a seat belt.
> Q (interviewer): How do you know when that's gonna happen?
> A: Like, just hope it doesn't!
> Q: So, should you or shouldn't you wear seat belts?
> A: Well, tell-you-the-truth we should wear seat belts.
> Q: How come?
> A: Just in case of an accident. You won't get hurt as much as you will if you didn't wear a seat belt.
> Q: OK, well what about these kinds of things, when people get trapped?
> A: I don't think you should, in that case.

From a rational perspective, one should weigh the probability of an accident where a seatbelt is beneficial versus the probability of an accident where the seatbelt only serves to prevent an imperative escape from the car after an accident. Clearly no such comparison is captured in the replies above. Worse yet, it seems the respondent is trying to say “yes” and “no” at the same time in a confused effort to address the binary situation.

Another subject’s response to the same question was:

> A: If you have a long trip, you wear seat belts half way.
> Q: Which is more likely?
> A: That you'll go flyin' through the windshield.
> Q: Doesn't that mean you should wear them all the time?
> A: No, it doesn't mean that.
> Q: How do you know if you're gonna have one kind of accident or the other?
> A: You don't know. You just hope and pray that you don't.

Probability is once again ignored or perhaps it is implicitly assumed that the two accident outcomes are equally likely. Even if this were true (it is definitely not) the idea of wearing the seat belt for half the time is still a dubious conclusion because wearing it for any portion of the time would be the same from a risk perspective.

A risk manager strives to incorporate the proper use of probability assessments in making decisions under uncertainty. It should be recognized that even technically minded people familiar with probability theory and practice still may have some form of probability neglect. This may take the form of underestimating small probabilities and overestimating larger ones.

Cumulative prospect theory (CPT) proposes a weighting function which links probabilities with their “distorted” view in those humans subject to the bias. The general understanding is that below a probability level of perhaps 40% or 50% one tends to inflate the probability in question in their decision making process while those probabilities above that level tend to be underweighted in the (biased) thought process. Amos Tversky and Daniel Kahneman who are credited with CPT describe this weighting function as mapping objective cumulative probabilities to subjective cumulative probabilities.[4] The graph of such a function, with objective probabilities on the X-axis and subjective on the Y-axis might be above the line $y=x$ for $x<40\%$ and below $y=x$ for $x>40\%$. The inverse of such a function can be used to stress test risk models and risk assessments that make use of probabilities.

Difficult Probability, Easier Impact

Consider the risk of losing a client or business partner where the impact might be reasonably estimated by removing the profit associated with the company that is inherent in the financial forecast. The estimation of impact is fairly straightforward, however, the probability may be difficult to assess as the situation is unique or perhaps there may be little data to suggest what a reasonable estimate would look like.

This situation might also be seen when one is trying to assess the risk of variation in forecast assumptions. If one projects a specific dollar amount of claims for an insurance product, it is straightforward to assess the impact of a deviation from that forecast claims level. Less obvious is the probability to assign to specific scenarios of deviation.

Difficult Impact, Easier Probability

A company that provides earthquake coverage in Chile can estimate the probability that a magnitude 5-6 earthquake strikes in the next year with the help of many years of data and the views of experts on earthquake activity in this region.

The impact is rather hard to estimate because so many factors and minor variations in the levels of these factors greatly affect the damage that results from a quake. (e.g., population density, location of the epicenter, soil stability, and the building codes in the affected region) Furthermore, the range of magnitude 5-6 quakes is itself very wide in terms of the strength of the quake. Consider that a quake of magnitude 6.0 versus a 5.0 would have 10 times the measured amplitude and about 31.6 times the amount of energy.

Interchangeability of Probability and Impact

In some cases there is very little difference between estimating probabilities and estimating impacts. In the practical use of a model, there are occasions where there is no difference at all between the probability and impact estimates. The following example concerns benefits paid to policyholders by an insurance company but could easily be restated to refer a wide variety of other risk assessments.

Example 1

As part of the financial planning for an insurance company one considers the forecast amount of total benefits paid in the Plan. This value is defined by the forecasting team as the statistical expectation or average per person benefit multiplied by the total number of policyholders for the next year. This average per person benefit is denoted by B and is estimated using a scenario approach partly informed by several years of data.

A total of n scenarios are defined to capture a robust range of payment outcomes from very small to very large. Each scenario specifies the payment (as a portion of the maximum annual benefit "M") and the probability of the scenario. The probabilities for the scenarios sum to 1.

There are constants determined by risk experts, all belonging to the interval [0,1], $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ such that the average benefit B is the probability weighted average of payments across the scenarios:

$$B = \Sigma\, Mx_iy_i \qquad \text{where the summation is over } i=1,2,\ldots,n$$

Consider the following questions:

1. Do revisions to an x value affect the calculation of B differently from revisions to a y value?
2. Observe that, for each i, the expression Mx_iy_i contains both the scenario payment ("impact") and the scenario probability. Making no assumptions based on the ordering of the factors in the expression, which of x_i or y_i is a probability estimate and which is the fraction of the maximum benefit M paid in the i*th* scenario?

3. From a modeling/planning viewpoint what is the difference in the effect of estimation errors (underestimates or overestimates) for probability versus that of impact?

If an initial estimate of B is altered by replacing a single x_i by $1.1x_i$ this would lead to the same revised value of B that would result from replacing y_i by $1.1y_i$. This is true for any choice of i and any revision multiple (such as the "1.1" used above). Said differently, B is symmetric with respect to the $\{x_i\}$ and the $\{y_i\}$. This may be intuitive by observing that B has a "symmetric" property: if each x_i is replaced by y_i and each y_i is replaced by x_i the value of B is unchanged. In other words, the answer to (1) is "no".

The symmetry concept above suggests that we cannot answer (2) with anything but a blind guess. Both choices yield the same result. After considering the replies to the first two questions it should be fairly clear how we can answer (3): in the method used to determine B, there is absolutely no difference between the effects of an impact estimate and a probability estimate.

Example 2

In situations where a continuous distribution is used to describe a particular risk source the risk modeling can be viewed as approximating the cumulative distribution function F. F is characterized (approximately) by specifying many ordered pairs $(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$ such that these points are on the graph of F. In other words, $y_i = P(\text{impact} \le x_i)$ for each $i=1,2,\ldots,n$.

For example, assume that the x values are potential storm surge levels for a coastal city. Estimating F for a large range of potential storm surge values (e.g. in feet) would be an important input to a decision regarding how high to build protective sea walls at various vulnerable points on the city's coastline.

One might say, "The exercise is really only about estimation of impact": one can simply list many (probability) values $y_1, y_2,\ldots,y_n$ between 0 and 1 (possibly using consistent spacing) and then estimate the related storm surge levels, i.e., $x_1, x_2, \ldots, x_n$.

Alternatively, someone might say, "the exercise is really only about estimation of probabilities": one can simply consider storm surge levels $x_1, x_2, \ldots, x_n$ and then estimate the probabilities that the surge is less than or equal to those levels, i.e., estimate $y_1, y_2,\ldots,y_n$.

Neither view is the "right" one. Pragmatically speaking, they are the same. The fundamental problem is that we do not know the probability associated with a given surge level; but this is the same thing as saying we do not know the surge level associated with a given probability.

The above situation can be re-worded to be about the pricing of excess of loss reinsurance. In this case the fundamental problem is the same: we can fix dollar loss levels and then make estimates of non-exceedance probabilities (percentiles) or we can choose percentiles and estimate the loss amounts.

3. Much Ado about Zero: Probability and the Impossible

Alan: We've still got a good wait for the next train…so what's next?

Kurt: I'm thinking of a number between 0 and 1

Alan: what are you talking about!?

Kurt: a real number between 0 and 1. Take a guess!

Alan: is it 1/3?

Kurt: close! I was thinking of 1/π!

Alan: I think the odds were against me

Kurt: how so?

Alan: well, let's remove the human element. If we consider a particular real number between 0 and 1, say x, and then we randomly generate a number r between 0 and 1 what is the probability that r=x?

Kurt: the random number is a draw from a uniform distribution over that interval?

Alan: yes

Kurt: then I would say the probability is rather close to zero!

Alan: close? What do you mean precisely?

Kurt: well, it seems the event you describe is not impossible…

Alan: agreed

Kurt: so…I guess I could not really put a value on it

Alan: would you believe me if I said it's exactly zero?

Kurt: but you just agreed the event is not impossible!

Alan: what's your point?

Kurt: are you saying the event is both possible and of zero probability?

Alan: exactly. Also the complement of the event has probability equal to 1 but it is not a certainty in the precise sense of the word

Consider a fixed number x from the interval $I = (0, 1)$ and let r be a random draw from a uniform distribution over the interval $I = (0, 1)$. What is the probability that $r=x$?

If we denote by $P(E)$ the probability of an event E then $P(r=x) \leq P(x-\varepsilon < r < x+\varepsilon)$ since for sufficiently small ε the interval $(x-\varepsilon , x+\varepsilon)$ is inside I and the event that $r=x$ is a subset of the event that $x-\varepsilon < r < x+\varepsilon$. Note that $P(x-\varepsilon < r < x+\varepsilon) = 2\varepsilon$, the length of the interval $(x-\varepsilon , x+\varepsilon)$ because we assumed a uniform distribution. So we have $P(r=x) \leq 2\varepsilon$ for any positive ε. Letting ε approach zero (through positive values) we conclude $P(r=x) \leq 0$. This shows $P(r=x) = 0$. [5]

You may have come across this concept before. Statistics courses often examine continuous distributions and the probability of a realization being in any real set in the sample space is defined as the integral of the probability density function (PDF) over that set. Because the integral of a real valued continuous function over a point is zero we have the same situation described above: an outcome equal to any particular point of the sample space is of probability zero but is *not* an impossible outcome.

While the preceding may at first blush seem to be an obscure consideration for theoretical statisticians, it does in fact come up in real world situations. As an example consider a stochastic model which simulates claims (in dollars) at an insurance company as a normal distribution. From a practical point of view one might say we are really only interested in outcomes rounded to the nearest dollar (or *thousands* or *millions* of dollars) so we are actually dealing with a *discrete* random variable with a very large set of possible outcomes, i.e., a large sample space. If we are only concerned with whole dollar amounts then no distinction is made between $x and $ x + $ 0.1, so for all intents and purposes the distribution is discrete. This is a fair statement but the very large sample space means the behavior seen in the continuous setting makes an appearance in a very similar guise: any specific claims outcome will have near zero probability. Let us ignore this for the moment and boldly plow ahead with the claims modeling example.

We run the model and analyze the results of several thousand simulated claims levels. If management and the Board are not entirely comfortable with stochastic modeling (and they rarely are) it is perhaps best to highlight several scenarios that capture the "flavor" of the full range of claims results.

Suppose five scenarios (S_1 to S_5) are described as follows:

S_1: claims of $15 million

S_2: claims of $10 million

S_3: claims of $5 million

S_4: claims of $3 million

S_5: claims of $1 million

The estimated scenario probabilities are 5%, 15%, 25%, 15% and 5% respectively. The financial plan ("the Plan") assumes $4 million in claims so that, in reference to the Plan, the first three scenarios are upside or optimistic scenarios and the last two are downside or pessimistic scenarios. Though it may be surprising, we've already run into some pretty big problems. These scenarios are meant to help communicate the likelihood of various magnitudes for claims outcomes but do they help?

Consider $P(S_1) = 5\%$. It is clear that claims coming out to *exactly* $15M is extremely unlikely (near zero probability). In practice we might get around this by saying we are really estimating the probability of a result "near" the indicated level of $15M. If pressed for the meaning of the term "near" perhaps the risk modelers would say this means "within 10% of the indicated level".

Based on this "within 10%" concept the clarified scenarios are:

S_1: claims of $13.5-16.5 million

S_2: claims of $9-11 million

S_3: claims of $4.5-5.5 million

S_4: claims of $2.7-3.3 million

S_5: claims of $0.9-1.1 million

Several unintended implications still remain. Firstly, the probabilities total to 60% so the message is that there is a 60% probability that the actual claims level falls within one of the five ranges above and only 40% that is does not. This is despite the fact that values *not* falling in the scenario ranges (e.g. 11-13.5, 5.5-9, 16.5+, etc.) represent a much larger set in terms of total length! Secondly, absolutely no information is given about the 40% of the time that results fall out of the scenario ranges. Finally, the first three scenarios have probabilities which sum to 45% and they beat the Plan by between 12.5% and 313.5%; is this message intentional?

This hints at a better approach which is to describe scenario impacts (whether they refer to loss ratio, sales, expenses, etc.) as *mutually exclusive intervals* which together cover the full range of potential outcomes.

Using such an approach might look something like this:

S_1: claims of $12 or more

S_2: claims of $7-12 million

S_3: claims of $5-6 million

S_4: claims of $3-4 million

S_5: claims of $0-2.5 million

We can now attach probability estimates to each scenario which sum to 1. This captures the full range of results and the probabilities are logically consistent.

In many cases, risk scenarios may be an intuitive concept for the subject matter experts while stochastic models may leave them less comfortable. A benefit of the above approach is that the intervals described in the scenarios allow straightforward stochastic modeling, should this be desired. One may randomly select which scenario is in effect using the indicated probabilities and then sample within the "activated" scenario's interval using some distribution (e.g. uniform, triangular etc.) to simulate the observed value.

4. The Intuitive Human and WYSIATI

<u>The Bat and Ball and the Obvious</u>

In his book *Thinking, Fast and Slow* [6] Daniel Kahneman discusses the following puzzle:

A bat and ball cost $1.10.

The bat costs one dollar more than the ball.

How much does the ball cost?

In all likelihood the answer will come to you very quickly and quite naturally. Kahneman mentions that most students from Harvard, MIT, and Princeton came up with a response of 10 cents. He goes on to say "the distinctive mark of this easy puzzle is that it evokes an answer that is intuitive, appealing, and wrong." A quick check shows that the reply "10 cents" would imply the bat costs $1.10 and the total cost would be $1.20, rather than the indicated $1.10. With a little thought it's obvious the correct answer is 5 cents, so that the bat costs $1.05 and the total cost is $1.10.

This type of error is symptomatic of a tendency to jump to a conclusion that seems natural and comes to mind very quickly. The fact that clearly intelligent people make this mistake suggests that in conducting assessments of risk, setting model parameters, or producing forecasts it is worthwhile to take a step back and force oneself to think "slowly" and carefully decide if what seems "obvious" is indeed true.

Intuition is an important guide for decision making in life, the corporate world in general, and ERM in particular. We often rely on the judgment of subject matter experts when we consider business strategies, tactics, or risk-reward decisions. In many situations the "from the hip" reaction of a seasoned expert is the best answer. In other cases these fast, intuitive assessments can be deeply flawed or biased.

As an example, consider catastrophe risk modeling. Insurers often make use of proprietary models which specify the statistical distribution of catastrophe losses for a season. Risk modelers might examine the 95th percentile loss for the Atlantic hurricane season, say $X, and then communicate to management or the Board that a scenario with a loss of $X has a 1 in 20 probability. Such a statement is "obvious" to the experts but of course it's completely wrong. The probability is actually near zero and it is only true, according to the model, that the probability of *exceeding* $X of loss is 1 in 20.

A more subtle example can be found in traditional financial forecasting. Assume a single number (point estimate) is produced as the best estimate for next year's earnings at an insurance company. It is not clear in what way the estimate is "best" but if pressed the modelers at an insurance company would perhaps say for each driver of earnings (e.g. earned premium, loss ratio, expenses, etc.) they have used the value they expect to be closest to the actual value.

This is equivalent to viewing the drivers as having statistical distributions and using their statistical expectations (i.e., means) as the values in the forecast. This is a natural approach but tacitly assumes that the expectation of earnings is found by setting each driver equal to its expectation and then using insurance and accounting logic to derive the earnings result. When certain dependencies or correlations exist between drivers such an assumption is not valid.

If paid claims is calculated as premium times loss ratio it is tempting to evaluate E[premium * loss ratio] as E[premium]*E[loss ratio] but such a relation implicitly assumes independence between premium level and loss ratio. Concepts such as economic morbidity and anti-selection demonstrate that such an assumption is often unsupported.

Update to the Risk Committee of the Board

The Board of Directors is expected to understand a company's risk profile and risk mitigations. Board members often must form quick opinions based on what amounts to just a slice of the full body of knowledge produced by an ERM program. Consider a meeting which focuses on a single line of business (LOB) and shows a risk ranking for this LOB complete with expected effects on earnings, capital, company value, and probabilities for various scenarios.

Suppose the risk-reward analysis shows very low volatility in earnings, little chance of needing a capital infusion, and forecasts a modest profit. The Board may be tempted to say the LOB should be taking on more risk. They have forgotten to consider what is happening in the rest of the company, and in doing so, they are ignoring the portfolio or aggregate view that is so sought after in the ERM world. Perhaps it is this stable and low risk business that tempers aggressive risk taking at other lines in the company. Alternatively, there may be common risks across the companies' LOBs that only appear significant when the aggregate view is taken and the concentrations reveal themselves. The Board has fallen into the "what you see is all there is" or WYSIATI mode of thinking.

The catastrophe risk modeling mentioned on the prior page offers another example. Even after getting all the jargon and technical details right (and communicating them well) we are still portraying the model as having a lot of precision. But we've seen that these proprietary models can be drastically different from one year to the next (especially after a bad season) so it's hard to believe they are capturing the "true distribution" in any one year. When mentioning probabilities associated with catastrophe seasons or events we should consider them only as a data point subject to the usual model risks; they are not "all there is" and they should be stress tested and supplemented with plausible scenario analysis. Such scenarios may mimic past hurricanes to analyze the effect on the current block of business.

5. On Expecting Value from Expected Value

Kurt: How much money you got on you?

Alan: Why do you ask?

Kurt: Care to play a game of chance? Maybe you'll make some money. The pot will start at $2 and I will flip this ordinary coin. The pot is doubled each time a head appears. At the first tail the game ends and you take what's in the pot at that time.

Alan: So I win $2 if you flip tails right away and $4 if you flip heads and then tails on the second toss. Two heads followed by tails would get me $8.

Kurt: Yes, you've got it. So if the game ends after k tosses you win 2^k *dollars. How much will you pay to play?*

Alan: Good question. I guess I can think about the expected value of the game. For the game to consist of k tosses than means we observe k-1 heads followed by tails and this has probability $1/2^k$ *since we're dealing with a fair coin. We are using a fair coin right?*

Kurt: Of course

Alan: so the expected value is $1/2 \cdot 2 + 1/4 \cdot 4 + 1/8 \cdot 8 + \ldots$ which is $1 + 1 + 1 + \ldots$ or infinity

Kurt: well you're a good friend so I'll let you play for…hmmm…let's say $1000?

Alan: Ha! I don't think I'd offer more than $10 or $20 at most, and that's if I'm feeling generous!

Kurt: But you just told me the expected value is infinite!

Alan: Your bankroll is not though

Kurt: Fair enough…but if we, for the sake of argument, assume an infinite bankroll what is the most you'd pay to play?

Alan: I guess I'd still not offer much more than $10 or $20

Kurt: What about the infinite expected value?

Alan: I guess that's not how a wise person should evaluate the decision of how much to pay to play…at least I wouldn't do it that way.

The St. Petersburg Paradox

The apparent disconnect between what one might pay to play such a game versus the expected value is often referred to as the St. Petersburg Paradox. Many try to avoid the apparent paradox by sidestepping the issue. Some do this by bringing up the same point about the bankroll being finite but if the assumption of an infinite bankroll is made they still cannot explain the discrepancy between the value they would assign to opportunity to play the game (e.g. something under $20) versus the game's infinite expectation.

A physical limitation can be pointed out in relation to time: because our lives are finite we can only observe a finite number of coin flips. Again, this is merely an evasion. If the assumption is made that we can simply simulate the number k such that the k*th* toss is the first occurrence of tails, then we again are up against the seeming paradox of the infinite expectation. (In a rather interesting thought experiment we may use a computer to randomly choose a value in the interval (0,1) and associate its infinite *binary expansion*, which consists solely of 0s and 1s, with a random sequence of coin flips by identifying 0s with "heads" and 1s with "tails")

Expected utility theory seems to provide a more substantial resolution. The Swiss mathematician Daniel Bernoulli suggested that the determination of the value of an item must not be based on price but rather on the utility it provides. He essentially argues that $1000 is worth much more to a poor person than to a millionaire. Another part of his attempt at a resolution is the concept of diminishing marginal utility.

Bernoulli described the utility as the natural logarithm of the amount of money one possesses, his "wealth" or w. In other words we may define the utility U of w dollars as $U(w) = \ln(w)$. Note that this definition meets the diminishing marginal utility condition:

$$\text{if } d > 0 \text{ and } 0 < x < y, \text{ then } U(x + d) - U(x) > U(y + d) - U(y)$$

Then we consider the expected utility for playing the game. Assuming we pay P to play and our wealth right before the first toss is w, the expected utility of the game is:

$$E(U) = \sum [\ln(w + 2^{k-1} - P] - \ln(w)]/2^k$$

where the sum is over all natural k = 1, 2, 3, … The sum is the probability weighted average of the changes in utility for games of length 1, 2, 3, … and it converges for any finite choices of w and P. One should be willing to pay up to any value of P that yields a positive expected utility. For example, a millionaire (w = 1,000,000) should be willing to pay up to $10.94.

This might seem to resolve the paradox but if the game is changed to have the higher payoff of e raised to 2^k , that is $\exp(2^k)$, then the sum E(U) diverges to infinity and we've probably resolved nothing. Expected utility theory has been modified in an attempt to better predict human behavior in the face of such choices. In one of these theories, cumulative prospect theory, the paradox may still occur with a concave utility function if it is not bounded.[7]

The preceding is somewhat involved and rather theoretical. A simpler attempt at resolving the apparent paradox might be to consider how one could win a "large" amount in the game. We probably would not be too surprised to win any amount up to $32 or $64 perhaps but what would it take to win a really big amount, say more than a million? If in the game there are 19 heads followed by a tail (total number of tosses is k=20) then we win 2^{20} which is 1,048,576. The problem is that this payout has probability of $1/2^{20}$ or about 0.00000095. Similar calculations will show in general we can be very confident in a *small* reward. For example the probability that tails is not flipped in the first 7 tosses is $1/2^7$ or about 0.008. This implies that with probability of about 99.2% the game ends on or before the 7th toss and winning at most 2^7 or $128. This fact alone implies most people would not pay (anything near) $128 to play.

So how can we resolve the purported paradox? Strangely enough a somewhat vague concept from ERM comes to the rescue: risk appetite.

For simplicity let us consider a simpler and free to play game where with 80% probability we win a billion dollars and with 20% probability we lose $1 million. The expected value is nearly $800 million but would you play with your personal wealth? Clearly the answer depends on how much risk exposure you are willing to assume in pursuit of your goal of winning and this varies from person to person. Perhaps you are a billionaire and you can afford to play the game a hundred times. Most people, who are not extremely wealthy, would not be willing to take a 20% chance of losing a million dollars regardless of the possibility of the rather impressive upside. They cannot "afford" to play this free game!

Reducing Information Loss

Risk managers must also be careful not to be blinded by expected value calculations: they often lead to the wrong decision. In many circumstances the downside outcome or set of scenarios represent a risk exposure you are not willing to accept. In other circumstances, the balance between the potential reward and the potential of a dire outcome may be acceptable if the downside outcome has an estimated probability which is sufficiently small. This might mean, for example, an annual probability of 1% or 0.1%. It is crucial to look at the ranges of possible outcomes for key metrics and their associated probabilities. Such information is well communicated by approximating the distribution (i.e. probability density function or cumulative density function) of results.

In some cases it is feasible to develop the (approximate) PDFs of results (e.g., based on reward metrics such as earnings, sales, or ROE) for each of several options in a business decision. If we denote the reward metric of interest as x, then the following are examples of quantities which help to compare the various options in a risk-intelligent manner:

- Mean [x] / standard deviation [x]
- Mean [x] / semi-variance [x]
- Median [x] – 10^{th} percentile CTE [x]
- Plan value [x] / expected shortfall at 20^{th} percentile [x]

The above represent just a fraction of the possibilities. One might also compute some version of a Sharpe or Sortino ratio or create an efficient frontier of risk-reward tradeoffs. These approaches capture more of the key risk information embedded in a PDF than simply using the mean alone.

6. Frenzy, Forecasts, and Fear

<u>The Availability Heuristic and Cascade or Why Ebola Scared You More Than it Should</u>

Estimation of future sales, earnings, and after tax profits are of primary importance in the corporate world. Forecasts are, or course, important in many other fields as well, and in late 2014, the Centers for Disease Control (CDC) produced a pessimistic scenario which projected that in late January 2015 the number of cases of Ebola (mainly in West Africa) would be close to 1.4 million. As it turned out, by March of 2015 the total number of cases was thought to be under 25,000. How is it that these experts could make a forecast which overestimated the number of cases by more than *55 times*?

As pointed out in the Economist, "for a start, the models relied on old and partial figures. These were plugged into equations whose key variable was the rate at which each case gave rise to others. But this "reproduction number" changed as outside help arrived and those at risk went out less, avoided physical contact and took precautions around the sick and dead. So difficult are such factors to predict that epidemiologists modeling a disease often assume that they do not change at all."[8]

The projected case count was so far removed from the levels seen in all previous Ebola outbreaks but many of us, including experts on infectious disease, were very willing to accept the dire prediction. Many of us were even worried for our personal safety despite the very low likelihood of contracting the disease outside of the affected regions in West Africa. One of the reasons for this was we could clearly picture how one of us or a loved one would become infected and suffer the horrific symptoms. It is this ability to "tell a story" to ourselves and picture how this could come to pass that distorts our view of the probability. For the majority of us, it is much more likely to be in a severe traffic accident than to contract Ebola but the constant imagery on the 24 hour news cycle blinded us to this fact, at least on an emotional level.

The term *availability heuristic* refers to a mental shortcut that is based on the ease with which one can recall examples when evaluating a specific notion or decision. The availability heuristic assumes that if something comes to mind easily it must be more frequent or important than something that is harder to recall or picture. As a result, our memory of recent events can distort our perception of what is important and, in the case of a risk, can lead to distorted views of its likelihood.

A study asked participants to consider pairs of causes of death and choose which one has the higher frequency.[9] Some of the findings are as follows:

- 80% of respondents believed accidental death to be more likely than death by stroke (in reality, death by stroke was almost twice as frequent)
- Death by lightning was felt to be less likely than death from botulism (death by lightning was 52 more times likely)
- Tornadoes were thought to kill more than asthma (asthma caused 20 times more deaths)
- Death by disease was viewed as about as likely as accidental death (death by disease was about 18 times as frequent)

It is clear that our perceptions can be warped by media coverage and the assumption that ease of recall is somehow proportional to likelihood.

Daniel Kahneman speaks of an *availability cascade* as “a self sustaining chain of events, which may start from media reports of a relatively minor event and lead up to public panic and large-scale government action.”[10]

In risk assessments we often make use of scenarios which include impact and probability estimates. We must be careful not to attach too much significance to the ability to picture a future event or recall a past one. In many cases an optimistic business forecast is a result of the ease in which we can picture the perfect execution of our plan. In some cases such a rosy projection flies in the face of data or statistics to the contrary.

<u>Base Rate Neglect and So-called Expert Forecasts</u>

An article written for Morningstar[11] begins with the following quote attributed to Yogi Berra: “It’s tough to make predictions, especially about the future.” The incredibly inflated Ebola forecasts certainly support this but that dire projection was likely made in part to raise awareness and generate funding and support. The article cites several other examples of forecasts gone wrong, some of which are described below.

“Stocks have reached what looks like a permanently high plateau.”

This one comes from Irving Fisher, the man whom legendary economist Milton Friedman called "the greatest economist the United States has ever produced". Unfortunately he said this three days before the Black Thursday crash in 1929 sent the United States into the Great Depression.

“We’re going to reach a point where stocks are correctly priced, and we think that's 36,000 ... It’s not a bubble. Far from it. The stock market is undervalued.”

James Glassman, author of the unfortunately-titled Dow 36000, made this call when the Dow was hovering near 11,500, in late 1999. It took until 2006 for it to close above 12,000 and a close above 18,000, merely half of his bold prediction, was first achieved in December of 2014.

"At this juncture, however, the impact on the broader economy and financial markets of the problems in the subprime market seems likely to be contained."

It is staggering to imagine that this quote is attributable to the Chairman of the Federal Reserve, Ben Bernanke, who in a testimony to the Congressional committee in March 2007 greatly misjudged the imminent and long-term economic disaster the subprime debt crisis was about to unleash.

In January of 2008, a few months before the global financial crisis, Bernanke added: "The Federal Reserve is not currently forecasting a recession."

As a society it seems we have a special fondness for expert forecasts. In many cases such predictions are wrong to an embarrassing extent and often could have been greatly improved by strong consideration of the normal or “base rate” seen in the available data. It is the temptation of too many experts (and laymen for that matter) to ignore the generic situation once a particular case comes under analysis. As an example, an executive pushing for a new business acquisition might project that all sales growth, profit and synergy projections will be met or exceeded over the next ten years while the historic record would paint a much more somber picture of post-acquisition performance.

Neglect of the base rate can appear in an explicit probabilistic environment as well. Consider the following scenario and your gut response to the question.[12]

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. (The Green company cabs are indeed all green and all Blue cabs are blue) You are given the following data:

- *85% of the cabs in the city are Green and 15% are Blue.*
- *A witness identified the cab as Blue. The court tested the reliability of the witness under the circumstances that existed on the night of the accident and concluded that the witness is 80% accurate in color identification (so is wrong 20% of the time).*

What is the probability that the cab involved in the accident was Blue?

Bayes’ theorem is the correct approach to solving this question and gives 41%. The “base rate” information in this problem is the statement about the percentage of each cab in the city. With no other assumptions we must infer from it that Green cabs are in 85% of the accidents (hit-and-run or otherwise). Bayes’ theorem is the right approach to combine that information with the witness testimony. If the base rate is ignored, and it is by many people, one comes up with a reply of 80%, which is very far from the correct answer of 41%.

In many situations the base rate is simply a historical average or trend. It may also be a company's or individual's track record of success for forecasting profits, picking winning stocks, or choosing attractive acquisition targets. One should be wary of projections or outlooks that are very different from base rates. At the very least some type of weighted average of base rate and the specific forecast should be brought up as an alternative to the stand-alone forecast.

If 70% of acquisitions destroy value, take it with a grain of salt when the M&A guy down the hall pitches the target as a "slam dunk" with expected sales growth of 20% each of the first five years and a profit margin higher than any previously seen at the company.

7. Decisions: Delphi, Deal Making, and Dating

Alan: Looks like my wife and I will be in the market for a new car soon

Kurt: Sticking with the usual?

Alan: No way…not after that debacle with the ignition switch problem. I just can't trust them. It seems even high ranking employees did not explain the deadly situation to their superiors. Come on…there's a problem where a car can slip out of drive and freeze up…no steering or brakes, airbags disabled… while you're driving it…and they knew this for ten years before they do a recall!

Kurt: When everyone is responsible no one is accountable

Alan: Very true. The independent report basically said their safety culture was defective…people were afraid to raise an unpopular issue and no one would accept responsibility for an issue. They had plenty of warnings and chances to address the problems with the ignition switch but failed to act

Kurt: My son is a risk manager. He makes sure each key risk has an owner…a real person, not a whole team or department. They understand what their roles are for monitoring and communication. I wonder if most companies spell out such things or if they assume everyone just gets it

Alan: Apparently the product safety litigation lawyers, including their most experienced attorney, knew about the ignition switch problem. He personally investigated it, but failed to tell his boss… the General Counsel. As a result the senior management was never made aware of the issue by the attorneys!

Kurt: It's unfortunate that something as important as risk management depends so heavily on a concept as soft as risk culture!

Alan: I've seen my students in group projects…I always think, if you get enough people together there's a risk that they'll barely have a brain between them!

<u>On Board the Flying Bank</u>

The term groupthink refers to a spectrum of cognitive biases but generally includes some notion of the desire for harmony or conformity in the group leading to irrational or dysfunctional decision-making.

Group members wary of conflict (or out of some sense of loyalty) may reach a consensus decision without analysis of other options, by avoiding disagreement, and by discounting external influences.

Aaron Hermann and Hussain Rammal cite groupthink as a contributing factor in the collapse of Swissair, a Swiss airline company that was thought to be so financially stable that it earned the title the "Flying Bank." The authors highlight two symptoms of groupthink: the belief that the group is invulnerable and the belief in the morality of the group. They also point out that before its failure, the size of the company board was reduced, subsequently eliminating industrial expertise. The authors go on to say, "with the board members lacking expertise in the field and having somewhat similar background, norms, and values, the pressure to conform may have become more prominent." [13]

A simple way to avoid several of the symptoms of groupthink is to, if possible, have members of the group indicate preferences, provide assessments, or estimate key quantities in private before conducting any group discussions on the way forward. Besides looking at individual results to assess majority views, one can also use the spread or range of numerical risk assessments as a measure of understanding or volatility of the risk in question. In other words, a wide variety of replies may indicate there is not a great understanding of the risk or it may simply be a risk with much volatility or uncertainty.

The Delphi method is a term which is sometimes used to refer to a group of experts reaching a common conclusion through collaborative discussion and analysis. Perhaps a better definition is "a procedure to obtain the most reliable consensus of opinion of a group of experts . . . by a series of intensive questionnaires interspersed with controlled opinion feedback."[14] Group discussion can be very useful but it is important to structure it in a way that avoids the various biases that naturally creep into such situations.

Gene Rowe and George Wright looked at the Delphi method as a forecasting tool and stated, "Delphi is not a procedure intended to challenge statistical or model-based procedures, against which human judgment is generally shown to be inferior: it is intended for use in judgment and forecasting situations in which pure model-based statistical methods are not practical or possible because of the lack of appropriate historical /economic/ technical data, and thus where some form of human judgmental input is necessary."[15]

Of Sunken Costs and Anchors

Risk quantification and financial forecasting often begin with a single initial estimate. The source for the value might be last year's value, a result of deep analysis, or could be based on intuition. Regardless of the source, once the initial value is seen, it is very hard to mentally move far away from it in subsequent consideration or estimation of alternatives.

A professor of management science at MIT, Dan Ariely, conducted a mock auction with his MBA students. A CFO magazine article summarizes his behavior experiment: "He asked students to write down the last two digits of their Social Security numbers, and then submit bids on such items as bottles of wine and chocolate. The half of the group with higher two-digit numbers bid "between 60 percent and 120 percent more" on the items…"People don't know how much something is worth to them," he comments. He adds that "people are good at setting relative values" but "it's very hard to figure out what the fundamental value of something is whether it's an accounting system, a company's stock, or a CEO."[16] This is the *anchor effect* and can be seen in financial forecasting, impact estimates, and many other situations. The first value proposed anchors our subsequent estimates in a small neighborhood.

Another cognitive bias tends to keep a person's view or decision tethered to the initial estimate or conclusion. In his article "Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action", Barry Staw describes this behavior in the investment world: "It is commonly expected that individuals will reverse decisions or change behaviors which result in negative consequences. Yet, within investment decision contexts, negative consequences may actually cause decision makers to increase the commitment of resources and undergo the risk of further negative consequences."[17]

This bias or a similar behavior is sometimes referred to as the sunk cost fallacy or escalation of commitment. If one has invested a lot of time, money, or effort into a project or goal, there is a tendency to view it as worthy of pursuit. This is sometimes seen in the M&A world. Suppose a target business is valued by the potential buyer at $10 million and negotiations and due-diligence occur over a two year period. When it comes time to close the deal the seller suddenly threatens to pull out if he does not get $12 million. At this point it is very tempting for the buyer to think of reasons the deal is still worthwhile; after all, so much time and effort has been invested in pursuing this deal that walking away is almost unthinkable. One might start to imagine all the synergies they will "surely" achieve that were not baked into the valuation.

<u>Don't Go Away Mad, Just Go Away</u>

Former CEO of Citibank Sandy Weill says "Knowing when to get out of the game is a critical consideration…I've been in situations where we've got an agreement, and as time goes by the other side sees you getting anxious and raises the price," he says. "You have to be disciplined at that point, and it isn't easy. Deal making is akin to dating and falling in love. If you don't think the behavior of the other party is something you can live with from a cultural point of view, you have to grit your teeth and simply say 'No. We're done.'"[18]

To draw this line in the sand, Jerre Stead, CEO and chairman of IHS Inc. offers this advice: "Have a walk-away price from the start." Following this counsel prevents "a potential acquisition from becoming an emotional decision," he explains. "Set a fair and full value upfront that you know you can stand behind, and stick to it. If it's not acceptable to all parties, be comfortable walking away from the deal."[18]

8. Small Numbers and the Sophomore Slump

Risk analysis often involves numerical data and our intuition about behavior of samples. Consider a hypothetical study which looks at rate of incidence for a certain disease. The national rate, R_N, is 8 in 1000 or 0.80%. The study looked at rates for many counties (named "A", "B" etc.) and a portion of the data is shown below:

County	Rate	Rate/R_N
A	0.00%	0.0%
B	3.41%	426.1%
C	2.73%	340.9%
D	0.00%	0.0%
E	0.40%	50.0%
F	0.84%	104.9%
G	0.81%	101.6%

Certain results stand out. Counties A and D show a 0% incidence and county E shows a rate of one half the national rate. Counties B and C show very large rates which are more than 4 and 3 times the national rate, respectively. The only counties that have rates close to the national rate are F and G.

Suppose that research reveals that counties B and C are poor, rural, low population counties. It is tempting to rationalize this outcome by suggesting these areas may have less access to quality medical care. This seems to make intuitive sense but it is also discovered that the counties with the very low rates, A, D, and E, are also poor, rural, low population counties.

The following chart shows the same results with the population and case counts added:

County Population	Number of Cases	Rate	Rate/R_N
48	0	0.00%	0.0%
88	3	3.41%	426.1%
110	3	2.73%	340.9%
129	0	0.00%	0.0%
250	1	0.40%	50.0%
15011	126	0.84%	104.9%
49590	403	0.81%	101.6%

This "data" is not referring to any real situation but was modeled stochastically. The county populations were chosen arbitrarily and then the number of cases was determined by assigning each "person" in the county a "healthy" or "diseased" state by using a random digit from (0,1). A random number less than 0.8% was interpreted as "diseased"; otherwise the person was considered healthy. For example, in the largest county, 49,590 random numbers were generated to arrive at the 403 case count. Observe that this method is consistent with the national rate, R_N, of 0.80% and assumes no correlation relating to cases, counties, or any other variables.

We are dealing with binomial distributions. For higher populated counties (i.e. larger samples) we would naturally expect less variance from the average rate of 0.80%. Conversely, a small sample can often show a large departure from the average rate. This is the primary reason that samples should be as large as feasible when one is attempting some sort of statistical inference, such as political polling.

The "law of small numbers" refers to this misleading pattern which drops out as we reach samples of sufficient size. Most risk managers know the problems with small samples but still may attach significance to the behavior seen in small samples. Many reading the above "data" would very quickly start to believe in their mental images of poor people with low standards of living who are far away from doctors who might have been able to provide them with proper guidance, care, and support.

Another type of numerical behavior that may throw us off is found in a wide range of situations from sports performance to stock picking. Consider an average or typical "C student" who receives an A+ on an exam. His parents praise him and buy him a new phone. His next grade is a C+. A few months later he disappoints them with an "F" and is severely punished. The next grade he gets is a C. What's happening here? Is positive reinforcement a bad idea and is it only punishment that motivates us?

If the student has a C average we must assume that an A+ is an aberration and so is the F. It is not common to see such large departures from the average; it is more likely to see a value closer to the average. If a baseball player is a .300 hitter and has a month where he bats .450 it is likely that the next month's average will not be such a large deviation and will be closer to .300. If he has a bad month where he bats .200, the next month will probably be better because in general we should not be "not too far" from the average. This is *not* to say that there is some type of auto-correlation or some "force" that pulls something back to the average; it is simply that in many situations a value closer to the average is more likely than one far from the average. Some refer to this as the Sophomore Slump. A fantastic result in a first attempt or first season is often followed by a less impressive performance.

The bottom line is this: in most real world situations, a value that is very far from the average is less likely than one which is closer to the average. If one of the "typical", near-the-average values follows one that was extreme (far from the average) this will seem like some type of *reversion to the mean* and this is a common term for this behavior. In sports, it is common to see a rookie's stellar first year followed by a more "normal" level of performance; in Australia they call this the "second year syndrome".

If a bond manager has an incredible year where he was among the top 1% of managers across the country, there's a good chance the next year will be less "extreme" and will be closer to the average. A business unit that has their best year ever will probably not do as well the next year. No matter what name is used for this behavior, risk managers should understand it and not assume it's a result of corrective action, praise, or year to year (negative) correlation.

9. Uncertainty, Exclusion, and Confusion

Kurt: My son had an interesting story…at an ERM seminar his company sent him to he heard someone from another company say "we don't estimate probability…we just estimate impacts". He asked the man why and the reply was "we have no idea what the probabilities are."

Alan: What did he think of that?

Kurt: I don't think he was impressed

Alan: Didn't he agree?

Kurt: He put it this way: in some situations it's easier to discuss impact than it is to think about probability, while it other cases the opposite is true. In some situations they are nearly interchangeable! He also said some companies don't use real "numerical probability" but instead like to use plain language phrases like "unlikely" or "very likely". He wasn't a big fan of those either.

Alan: at least with probability you can't be off by more than one!

Kurt: well it's a big spectrum in between…anyway, I asked him if they use probabilities in portfolio management. He once told me a lot of ERM was motivated by that field.

Alan: I dabbled in investments bit. I think, for example, a bond manager does not have much use for probability estimation. They tend to use risk adjusted rates, credit spreads, or give a haircut to a yield.

Kurt: Really? What's the idea with a credit spread?

Alan: they first consider a risk free rate…think CDs…for the same time horizon of a potential bond investment they're looking at. They then add a number to the risk free yield as a premium that they expect to earn over the risk free yield as compensation for the fact that the issuer of the bond may not make all the interest or principal payments on time or in full

Kurt: So if the risk free rate is 3% and I'm looking at a risky bond over the same time horizon I might demand, or at least hope, that the expected yield on the bond is a few percent above that level?

Alan: right…if the credit spread is 2% or "200 basis points in the jargon" then the bond is priced to earn 3% + 2% or 5%.

Kurt: and higher spreads for larger perceived credit risk?

Alan: yes, all other things being equal

Kurt: how is the spread determined?

Alan: the market basically prices it through their demand for the bond…that effectively sets the price and the spread pops out of the implied yield, which is a function of the market price

Kurt: and their assessment of default is baked in?

Alan: exactly

Kurt: so they somehow consider probability of default of the various coupon payments and return of principal…do they use probabilities?

Alan: I think it's somehow equivalent…it may be a sort of kick the can game. Instead of assigning probabilities to each future cash flow you just do something similar in one fell swoop by deriving the credit spread

Kurt: that almost sounds harder! I wonder if there is some mathematical equivalence or translation…

Restating Probability in Terms of Rates

Suppose an investment is scheduled to make a stream of 10 annual payments each in the amount of \$1000, with the first payment due one year from today. Assume the investment carries credit risk and this is the only risk that affects the valuation. Because of this credit risk, the value of the investment should be less than its present value based on a risk free rate. That is, it should be *less than*:

$$1000[1/(1+r) + 1/(1+r)^2 + \ldots + 1/(1+r)^{10}] \text{ where r is the annual risk free rate.}$$

A valuation for this investment can be made as the expected present value which is the sum of the probability-weighted discounted cash flows. This is, of course, the same as the statistical expectation.

If we write the probability of the n*th* payment as p_n, then the value would be:

$$1000[p_1/(1+r) + p_2/(1+r)^2 + \ldots + p_{10}/(1+r)^{10}]$$

Because of some properties of these continuous functions there is a constant s such that:

$$1000[p_1/(1+r) + p_2/(1+r)^2 + \ldots + p_{10}/(1+r)^{10}] =$$

$$1000[1/(1+r+s) + 1/(1+r+s)^2 + \ldots + 1/(1+r+s)^{10}]$$

Of course "s" is usually called the credit spread and it captures the probabilities p_1, p_2, …p_{10} in "one fell swoop". Though easy to determine using a computer, the value of s is a fairly complicated function of the probabilities. It reduces the problem of making 10 estimates of probabilities to the estimate of a single constant but s does depend on each of the probabilities; it is an average of sorts. This is a sort of "kick the can": the spread, which is added to the risk free discount rate, does not remove the need for estimation of probability; it just rephrases it mathematically into a rate adjustment.

The same idea is used when requiring higher "hurdle" rates for riskier businesses. The riskiness of the business could be captured in probabilities assigned to cash flows or earnings flows but one may choose instead to try to reflect them all at once in a spread which is added on to a risk free discount rate.

We Only Estimate Impacts, Not Probabilities

Some ERM functions eschew estimation of probability because, they say, "we have no idea what the probability is." It seems they are comfortable modeling impact of a scenario but choose to remain silent on how likely such a scenario might be. Section 2 explained why this view makes little sense.

If a scenario's impact is carefully modeled and then presented to management what is the message? Is it that "here is the impact of some event which we believe is not out of the realm of possibility"? Clearly they spent time on impact estimation because there is some view that it has a legitimate chance of happening or they want to be prepared for it. Risk is the interplay of probability and impact; can a robust risk assessment be performed by only looking at one of these two entangled concepts?

Stochastic models are sometimes used to estimate the volume of a new oil field. The modeling might include ranges (intervals) for the area of the field, the depth, the porosity of the rock, the water content, etc. When the model is run, the output shows a range of possible values for how much oil is in the field. But when it comes to modeling one of the most uncertain variables, the price of oil, they sometimes don't use ranges. For the price of oil they may use an exact point. In Douglas Hubbard's view, this means "that when management is looking at the output of a risk of an oil exploration project, they really aren't looking at the actual risks. They are looking at a hybrid of a proper risk analysis based on ranges and an arbitrary point estimate. They undermine the [model's] entire purpose."[19]

Hubbard provides other examples of risk managers who develop very detailed Monte Carlo (stochastic) models but intentionally exclude any modeling for certain key variables they regard as "too uncertain" to model. He goes on to say "why leave out something because it is uncertain? The whole point of building a Monte Carlo model is to deal with uncertainties in a system. Leaving out a variable because it is *too* uncertain makes about as much sense as not drinking because you are thirsty." [20]

<u>The Probable Damage of Probability Phrases or How to Add a Layer of Confusion</u>

In some companies probability is regarded as too technical a concept for many subject matter experts. As an attempt to get away from this challenge they may use plain language phrases instead of (numerical) probabilities. This concept includes terms such as "likely" or "very unlikely".

Though we typically cannot estimate probability with a lot of precision, the advantage of using the true numerical version (rather than the phrases) is that at least we understand what they mean and avoid introducing the subjective interpretations that the phrases invariably create.

A study by David Budescu, Stephan Broomell, and Han-Hui Po of the University of Illinois on the use of these probability phrases made use of a report by the Intergovernmental Panel on Climate Change (IPCC). Subjects were asked to read sentences from the IPCC report which used phrases such as "likely" and "very unlikely" and then assign a numerical probability to the statement. For example, the report has the statement "It is *very likely* that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent." The subjects would read this sentence and assign an equivalent probability to this event. For example, a subject might read this statement and estimate that "There is *95% probability* that hot extremes, heat waves, and heavy precipitation events will continue to become more frequent." [21]

The study demonstrates a wide variety of interpretation of the phrases despite the fact that the subjects were given specific guidelines which mapped the phrases to probability levels. Budescu feels that use of the phrases creates an "illusion of communication." It may seem people are in agreement on the likelihood of a certain event when they in fact have very different views. The exhibit on the next page shows a summary of the results.

Not only may there be inconsistency in the use of such phrases by different people; in some occasions the same person may use the phrases to express very different probability levels depending on the context. Douglas Hubbard cites an interesting, if somewhat unsettling, example [see note (21)]:

> *I was talking to a client about a scoring method he had applied to risk related to a large project portfolio…I asked one manager, "what does it mean when you say this risk is 'very likely'?". I pointed to a particular risk plotted on his 'risk matrix'. With little hesitation, he said, "I guess it means there is about a 20% chance it will happen." One of his colleagues was surprised by this response. When he asked for clarification the first manager responded, "Well, this is a very high impact event and 20% is too likely for that kind of impact." A roomful of people looked at each other as if they were just realizing that, after several tedious workshops of evaluating risks, they had been speaking different languages all along.*

Highlights from the "Probability Phrases" Study

Probability Phrase	IPCC Guideline for Meaning	Interpreted Meaning According to Subjects (Distribution of Actual Responses): Minimum of All Responses	Interpreted Meaning According to Subjects (Distribution of Actual Responses): Maximum of All Responses	Percent of Responses Violating Guidelines
Very Likely	> 90%	43%	99%	58%
Likely	> 66%	45%	84%	46%
Unlikely	< 33%	8%	66%	43%
Very Unlikely	< 10%	3%	76%	67%

Source: David V. Budescu, Stephen Broomell, and Han-Hui Po, University of Illinois at Urbana-Champaign

To avoid asking about probabilities, others may use a numerical likelihood rating or scale (e.g. 1-5) that does not explicitly discuss numerical probabilities. Perhaps they initially have a discussion and evoke some of the typical probability phrases above. If a 5-point scale is used, how can one interpret a "4" versus a "2"? Does this indicate an event having twice the probability as the other? Is there some linearity or proportional assumption implicit in the rating or will such characteristics be assumed by management when viewing these ratings?

Another symptom of this "disease" is the use of a color scale. Does "red" mean the event is "more likely than not" as the FASB would describe? Or does it mean a probability of more than 67% or 80%? There is great potential for confusion and inconsistency. It's worth mentioning that these problems can also occur if the color rating or scale is applied to impact estimates. Does a "yellow" mean the same thing to a small business unit that it does for a larger one? How can one aggregate the various colors to get the holistic view that ERM is meant to provide? Even if one addresses these issues with clear definitions or instructions it is quite possible they will be ignored, misconstrued or forgotten, as was the case in the University of Illinois study described above.

It is important to aim for using true, numerical probabilities any time the audience is willing to do so. Of course they will typically be rough estimates but this situation is better than adding an additional layer of confusion and inconsistency by introducing probability phrases. While everyone will understand what an 80% probability is, no one knows what "very likely" means, even if you tell them!

10. Conclusion

ERM is, to use a hackneyed expression, "both an art and a science". In many circumstances we strive to quantify impacts or probabilities of potential events which do not come with extensive histories or data sets. Even when there is a large amount of information on a possible risk event, it still may not be clear if our models are making reasonable assumptions. An important benefit of quantitative modeling is that it helps bring our views out into the open and helps to express them in an internally consistent manner. That transparency and the "apples to apples" comparisons it enables are important even though our models are surely imperfect.

ERM operates in the *fog of war*. Uncertainty is in the risk exposures we *believe* we face and it is in our understanding of the exposures, the expected impacts, and interrelationships. Indeed, without uncertainty the concept of risk makes little sense. The way we deal with this uncertainty in our modeling and in our decision making is much of the "art" of ERM. Though this aspect is unavoidable (and perhaps even intellectually stimulating) there must still be a vigilant pursuit of internal logic, consistency, and, where possible, accuracy.

Even a comprehensive and robust ERM framework is fraught with uncertainty and incomplete information. A risk manager must not exacerbate this situation by allowing cognitive biases to have free rein. Proper mitigation can reduce the ability for these biases to distort our risk assessments and will only lead to more informed business decisions.

Notes

1. The limit of $(1-1/x)^x$ as x approaches infinity can be derived in several ways. A relatively straightforward approach is to let $y = (1-1/x)^x$ so that ln y = x ln (1-1/x) and then use the Taylor expansion for the natural logarithm. After distributing the first "x" it is easy to see ln y approaches -1 as x approaches infinity so the original expression, y, approaches 1/e.

2. See "Modeling Tail Behavior with Extreme Value Theory" available online at: http://www.soa.org/library/newsletters/risk-management-newsletter/2009/september/jrm-2009-iss17-levine.pdf

3. "Decision-making biases in children and early adolescents: exploratory studies" Baron, J., Granato, L., Spranca, M., & Teubal, E. (1993). Merrill-Palmer Quarterly, 39(1), 22-46.

4. See http://psych.fullerton.edu/mbirnbaum/psych466/articles/Tversky_Kahneman_JRU_92.pdf

5. Let r be a random draw from a uniform distribution over the interval I = (0, 1) and let P(E) be the probability of an event E. Then for any numbers a and b from I, with a < b, we have P(r ε [a,b]) = b – a, the length of the interval [a, b]. The symbol ε is shorthand for "belongs to" or "is an element of"; in other words r ε [a, b] means a ≤ r ≤ b.

   Consider a fixed, real number z from I and let d=min (z, 1-z). So d represents the distance from z to the closer of zero and one, and therefore d<1. For each n=1, 2, 3,… let J_n = [z-d/(2n), z+d/(2n)]. Then each J_n is a subset of I and z ε J_n for each n = 1, 2, 3… so that for the random draw r we have P(r=z) ≤ P(r ε J_n) for each natural number n.

   Because P(r ε J_n) = length of J_n = d/n this implies, using the inequality above, that P(r=z) ≤ d/n for each n=1,2,3,… Because any probability is non-negative and d<1 we have 0 ≤ P(r=z) ≤ d/n <1/n for each n. But the decreasing sequence {1/n} approaches zero: given any positive number θ there is a natural number k such that 1/n < θ for all n ≥ k. We then have 0 ≤ P(r=z) ≤ 0 and therefore P(r=z) = 0.

6. Thinking, Fast and Slow by Daniel Kahneman describes a large number of cognitive biases and was the winner of the National Academy of Sciences Best Book Award in 2012.

7. "Cumulative prospect theory and the St. Petersburg paradox". Rieger, Marc Oliver; Wang, Mei (August 2006). Economic Theory 28 (3): 665–679.

8. Available online at: http://www.economist.com/news/international/21642242-why-projections-ebola-west-africa-turned-out-wrong-predictions-purpose

9. "Facts and Fears: Societal Perception of Risk", Paul Slovic, Baruch Fischhoff, and Sarah Lichtenstein (1981), NA - Advances in Consumer Research Volume 08, eds. Kent B. Monroe, Ann Arbor, MI : Association for Consumer Research, Pages: 497-502

10. See (6), Kahneman, p. 142.

11. “Financial Forecasts Gone Wrong”, by Nazin Khan. Online at: http://www.morningstar.in/posts/12556/financial-forecasts-gone-wrong.aspx

12. See (6), Kahneman, p. 166. One can use Bayes’ theorem in the form: $P(A|B) = P(B|A)P(A)/[\ P(B|A)P(A) + P(B|A')P(A')$ where A' is the complement of A. The left hand side can be taken as P(cab was blue | witness said “Blue”) to arrive at approximately 41.4%.

13. “The grounding of the “flying bank”” Hermann, A., Rammal, H. G. (2010).. Management Decision 48 (7): 1051.

14. “An experimental application of the Delphi method to the use of experts” Dalkey, N. C., & Helmer, O. (1963). Management Science 9, 458-467.

15. “The Delphi technique as a forecasting tool: issues and analysis”. Rowe, Gene & Wright, George. International Journal of Forecasting 15 (1999) 353–375.

16. “Avoiding Decision Traps”, by Edward Teach. Online at: http://ww2.cfo.com/human-capital-careers/2004/06/avoiding-decision-traps/

17. “Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action”, by Barry Staw. Organizational Behavior and Human Performance, Volume 16, Issue 1, June 1976, Pages 27–44.

18. “Making the Most of M&A Deals”. Online at: http://chiefexecutive.net/making-the-most-of-manda-deals

19. The Failure of Risk Management, Douglas Hubbard (2009), p. 192.

20. Hubbard, p. 192.

21. Hubbard, p. 126-128.

Disclaimer: The views expressed in this paper are my own and not necessarily those of my employer, Assurant Inc.