2015 Enterprise Risk Management Symposium

June 11–12, 2015, National Harbor, Maryland

Into the Tails of Risk: An Intervention into the Process of Risk Evaluation

By David Ingram

Into the Tails of Risk: An Intervention into the Process of Risk Evaluation

David Ingram

Abstract

People naturally observe risk as the range of experienced gains and losses represented in statistical terms by standard deviation. Statistical techniques are used to develop values for extreme tails of the distribution of gains and losses. These processes are essentially an extrapolation from the "known" risk of volatility near the mean to "unknown" risk of extreme losses. This paper will propose a tail risk metric (the coefficient of riskiness) that can be used to enhance discussion between model builders and model users about the fatness of the tails in risk models.

Risk models all start with observations. Modelers look at the observations and the shape of a plot of the observations. From that shape, the modelers choose a mathematical formula to represent the risk driver (such as interest rates or stock market returns) or for the loss severity itself. Those formulas are known as probability distribution functions (PDF).

The most famous and most commonly used of these functions is known as the "normal" curve. Mathematicians (sometimes called quants or rocket scientists) particularly favored the use of the normal PDF because its mathematical characteristics made it particularly easy to manipulate, making rapid analysis of risk functions based upon the normal PDF possible.[1]

---

[1] Indeed, the use of the normal PDF in finance can be traced to the rediscovered 1900 thesis of Louis Bachelier, *Theory of Speculation*, trans. Mark Davis and Alison Etheridge (Princeton: Princeton University Press, 2006). Bachelier sets a standard followed by many of presenting the normal PDF as *the* basis for statistical modeling of financial risk. Bachelier may have also been the first to caution that "The calculus of probabilities can no doubt never apply to movements of stock exchange quotations."

Thanks to Daniel Bar Yaccov, Ph.D., Ian Cook, FIA, and Neil Bodoff, FCAS, for their comments. The author is solely responsible for how their excellent advice was actually used.

In the 2008 global financial crisis, we found that many financial market risk models based on the normal PDF drastically underestimated the likelihood of losses which were much, much worse than the average.

Unfortunately, for many people, expectations of extreme losses learned from business courses, media and, to some extent, risk models are drawn from the very same characteristics as the normal PDF.

The language of the normal PDF is our basic language of risk. The normal PDF is defined completely by just two terms—mean and standard deviation. We tend to expect the mean and the standard deviation to tell us "all about" any new risk, without realizing we are thereby assuming the risk is normal.

The normal PDF says we should expect about two-thirds of our observations to fall within one standard deviation of the mean and over 90 percent of the observations within two standard deviations of the mean. It also says it is extremely unlikely to have any observations beyond three standard deviations from the mean. In fact, observations should fall within three standard deviations 99.9 percent of the time for the normal PDF.[2]

And that is how we were able to confirm the normal PDF underestimated the likelihood of large deviations from the mean. David Viniar, Goldman Sachs' chief financial officer, famously observed during the financial crisis, "We are seeing things that were 25 standard deviation moves, several days in a row,"[3] which, under the normal PDF, was highly unlikely to happen even once in the time since the last ice age ended.

[2] Of course, the normal PDF actually says the 99.9th percentile observation should be 3.09 standard deviations from the mean.
[3] Peter Thal Larsen, "Goldman Pays the Price of Being Big," *Financial Times*, August 13, 2007.

The idea that risk fits a normal curve is so deeply embedded that almost all discussion of Viniar's 25-standard-deviation statement was in the form of discussion of exactly how to calculate the likelihood of a 25-standard-deviation move under the normal PDF, instead of challenging the very idea that the normal PDF might not be appropriate.[4]

Two noted exceptions to these generalizations are Benoit Mandelbrot and Nassim Taleb. Mandelbrot, in his work studying price movements in cotton markets in the 1960s, suggests there are seven states of randomness, only the first of which is properly modeled by a normal PDF.[5] Taleb, in his books, actually divides the world into two regimes—Mediocristan and Extremistan—where the normal PDF explains the first regime and a Pareto PDF explains the second.[6]

In insurance modeling by actuaries and catastrophe modelers, the use of a normal PDF is much less dominant. Other PDFs, especially the Pareto PDF, allow for quite extreme values with relatively high likelihood. In fact, with certain calibrations, the Pareto PDF allows for infinite values of metrics like variance, something that is possibly even more unrealistic than the normal PDF's low likelihood for extreme values. Alternately, some modelers who see the need for higher likelihood of extreme values with normal PDF-like features otherwise have used combinations of multiple normal PDFs to achieve the desired "fat tails."[7] Other models of a single category of risk exposures may combine two or more different PDFs. For example, a model of a property insurance line of an insurer may consist of separate models of natural catastrophe losses, losses from large exposures and losses from small and

[4] For example, see "How Unlucky is 25-Sigma?" by Kevin Dowd, John Cotter, Chris Humphrey and Margaret Woods.
[5] Benoit B. Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business* 36. no. 4 (1963).
[6] Nassim Nicholas Taleb, *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life* (New York: Random House, 2001); *The Black Swan: The Impact of the Highly Improbable* (New York: Random House, 2007); and *Antifragile: Things That Gain From Disorder* (New York: Random House, 2012).
[7] Mary R. Hardy, "A Regime-Switching Model of Long-Term Stock Returns," *North American Actuarial Journal* 5, no. 2 (2001).

moderate-sized exposures. Each of these submodels is often based upon a different PDF.

Each of the alternate PDFs has different characteristics that have been given names by statisticians such as skewness (which quantifies asymmetry) and kurtosis (which quantifies the sharpness of the distribution's peak). The accepted wisdom among modelers is that for someone to "understand" a model of risk, they must walk the path of the modelers: Follow the math of the PDFs and definitely understand the nuances of skewness and kurtosis.

Extreme value theory (EVT) is an explicit but highly technical approach to building statistical models that are not focused on fitting the mean or the observations near to the mean. EVT focuses on using specific PDFs that are inherently fat tailed. The EVT process is designed to be driven by the data and the axioms of EVT to analytically determine the tails, especially the values beyond the observations.[8]

There is a strong push for top managers and even board members to become active users of the outcomes of risk models and to actually participate in the process of validating the risk model. For example, the risk committee charter of one bank says that board committee will oversee:

> Model Risk, by reviewing all model-related policies and assessments of the most significant models, in each case annually, and reviewing model development and validation activities periodically.[9]

But both the mathematical approach to describing the PDFs and the process-based explanations that require simply following the modeler's thinking fail to engender either understanding or faith in the model. JPMorgan Chase & Co., the original

---

[8] See Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch, *Modelling Extremal Events for Insurance and Finance* (New York: Springer, 1997).

[9] Santander Consumer USA Holdings Inc., "Board Enterprise Risk Committee Charter," effective Dec. 8, 2014.

proponent of the value-at-risk models used extensively in banks, experienced a major loss in late 2011 and early 2012 that was in part attributed to a flawed risk model update.[10] According to Esade Business School professor Pablo Triana:

> The perception of a bank's risk should not depend on the technicalities of a mathematical model but rather on commonsensical analysis of what should and should not be acceptable.[11]

The remainder of this paper will present an alternate approach to discussing the nature of a risk model's prediction of the likelihood of an extreme deviation. This approach will not require extensive mathematical or statistical education on the part of the user, nor will it require much in the way of new vocabulary. It will work from where most people stand now in their understanding of the math of risk—with the concepts of mean and standard deviation. This approach to presenting a measure of "fatness of tails" does not replace anything currently in wide use for discussions of risk models with nontechnical users of risk models. It could be a powerful addition to the discussion of risk models with those nontechnical users and may lead to an important change in the relationship between those users and modelers by providing a basis for communication regarding a most important aspect of the models.

---

[10] Christopher Whittall, "Value-at-Risk Model Masked JP Morgan $2 bln Loss," *Reuters*, May 11, 2012.
[11] Tracy Alloway, "JPMorgan Loss Stokes Risk Model Fears," *Financial Times*, May 13, 2012.

Extrapolating the Tails of the Risk Model

The statistical approach to building a model of risk involves collecting observations and then using the data along with a general understanding of the underlying phenomena to choose a PDF. The parameters of that PDF are then chosen to a best fit with both the data and the general expectations about the risk.

This process is often explained in those terms—fitting one of several common PDFs to the data. But an alternate view of the process would be to think of it as an extrapolation. The observed values generally fall near to the mean. Under the normal PDF, we would expect the observations to fall within one standard deviation of the mean about two-thirds of the time and within two standard deviations almost 98 percent of the time. When modeling annual results, it is fairly unlikely we will have even one observation to guide the "fit" at the 99th percentile.

So, in most cases, we really are using the shape of the PDF to extrapolate to get a 99th percentile or 99.5th percentile value. But our method of describing our models presents that fact in a fairly obtuse fashion. Sometimes model documentation mentions the PDF we use for this extrapolation. Rarely does the documentation discuss why the PDF was chosen and, when this is discussed, it is almost never mentioned that it is judgment of the modeler which drives the exact selection of the parameters that will determine the extreme values via the extrapolation process.

After the 2001 dot-com stock market crash, many modelers of stock market risk adopted a regime-switching model as a technique to create the "fat tails" that many realized were missing from stock market risk models.[12.]

But how fat were the tails in these regime-switching models? Would reporting the skewness and kurtosis of the resulting model help with understanding of the model?

---

[12] Mary R. Hardy, "A Regime-Switching Model of Long-Term Stock Returns," *North American Actuarial Journal* 5, no. 2 (2001).

Or is the regime-switching equity risk model now a black box that can only be understood by other modelers?

Almost every business decision-maker is familiar with the meaning of average and standard deviation when applied to business statistics. We propose that those commonly used and almost universally understood terms be used as the basis for a new metric of "fatness of tails."[13]

We use the idea of extrapolation to construct for this new proposed measure of fatness of tails. The central idea is that we will have a three-point description of our risk model, and with these three terms we can describe the degree to which we can expect a risk to have common fluctuations that will drive variability in expected earnings (mean and standard deviation) as well as a third factor that indicates the degree to which this risk might produce extreme losses of the sort we generally hold capital for.

Coefficient of Riskiness

We will add just one term to our elementary vocabulary of risk—the coefficient of riskiness (CR). This value will be the third term in describing the risk model. It is the indicator of the fatness of the tail of the risk model.

$$CR = (V_{.999} - \mu)/\sigma$$

Or, in English, the number of standard deviations the 99.9th percentile value is from the mean.[14]

[13] Many analysts rely on the coefficient of variation (CV) for comparing riskiness of different models. The CV is a good measure for looking at earnings volatility, but it does not give strong indication of the fatness of the tails. Its definition, using only mean and standard deviation, also supports a presumption of the normal PDF.
[14] The choice of 99.9th percentile is discussed in the appendix of this paper.

We used this concept above when we said *observations should fall within three standard deviations 99.9 percent of the time for the normal PDF.*

The CR can be quickly and easily calculated for almost all risk models. It can then be used to communicate the way the risk model predicts extreme losses, allowing for actual discussion of extreme loss expectations with nonmodelers. We use the mean and standard deviation in defining the CR not because they are the mathematically optimal way to measure extreme value tendency, but because they are the two risk-modeling terms already widely known to business leaders.

Potentially, the CR could become a part of the process for the initial construction of risk models, taking the position of a Baysean prior[15] in the common situation where there are no observations of the extreme values. And, if CR has been established as a common idea with nonmodelers, they could have a voice in the process of determining how the model will approach that part of the risk-modeling puzzle.

The CR value will not be a reliable indicator for models where the standard deviation is not reliable. It is instructive to identify the characteristics of such models and the underlying risks such models seek to capture.

Coefficient of Riskiness for Various Probability Distribution Functions

The CR for the normal PDF is 3.09. This is true for all models that use the normal PDF because all values of a normal PDF are uniquely determined by the mean and standard deviation.[16]

[15] A Baysean prior is an opinion that acts as a seed to the risk model at the stage of the process when there is insufficient data to fully define a mathematical model.

[16] For the reader who wishes to check this, an Excel table of values for mean, standard deviation, 99.9th percentile value and CR can easily be constructed. Mean and standard deviation would be values, 99.9th percentile value would be Norminv(.999,mean,std dev) and the CR would be 99.9th percentile value less the mean divided by the standard deviation. Try as many values for the mean and standard deviation as you wish.

Another commonly used PDF is the lognormal. The lognormal model has two characteristics that make it popular for risk models—it does not allow negative outcomes and it has a limited positive skew.[17]

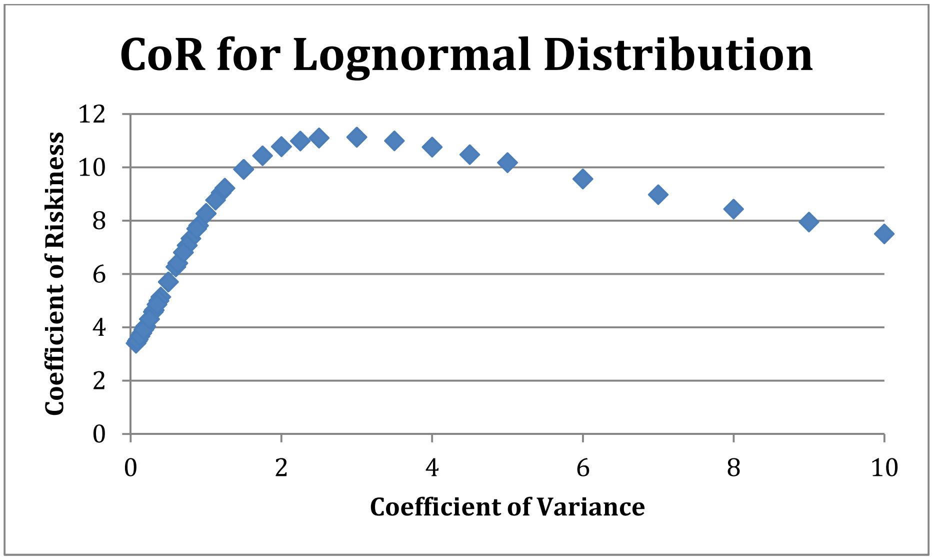
Figure 1. Lognormal PDF—CR for
various means/standard deviation combinations

		Mean				
		100%	80%	40%	20%	10%
Standard Deviation	7%	3.4	3.5	3.9	4.9	6.8
	10%	3.5	3.7	4.3	5.7	8.3
	15%	3.8	4.0	5.0	7.1	9.9
	20%	4.0	4.3	5.7	8.3	10.8
	25%	4.3	4.6	6.4	9.2	11.1
	30%	4.6	5.0	7.1	9.9	11.1
	40%	5.1	5.7	8.3	10.8	10.8
	50%	5.7	6.4	9.2	11.1	10.2
	60%	6.3	7.1	9.9	11.1	9.6
	70%	6.8	7.7	10.4	11.0	9.0
	80%	7.3	8.3	10.8	10.8	8.4
	90%	7.8	8.8	11.0	10.5	7.9
	100%	8.3	9.2	11.1	10.2	7.5
	120%	9.0	9.9	11.1	9.6	6.7

As it turns out, the CR is a function of the ratio of standard deviation to mean (also known as the coefficient of variance) for the lognormal PDF.

[17] The normal PDF is exactly symmetrical and allows negative values. The positive skew of the lognormal PDF means that it is not symmetrical, extending much further on the right (positive) side of the mean than on the left (toward zero) side.

Figure 2. Lognormal—CR vs. CV



The Poisson PDF is also widely used because of its relationship to the binomial distribution. Since the Poisson PDF is fully determined by a single parameter, the CR is always approximately 3.5.

The Pareto PDF and its close cousin, the exponential PDF, are used for a variety of types of risks. These risks all have the characteristic that they are usually fairly benign but in rare instances they produce extremely adverse outcomes. Operational risks are sometimes modeled with a Pareto PDF. Risks from extreme windstorms and earthquakes are also modeled with Pareto PDFs, as is pandemic risk.

In 2006, Mandelbrot and Taleb together proposed the use of the Pareto PDF for looking at vulnerability to tail risks:

> The same "fractal" scale can be used for stock market returns and many other variables. Indeed, this fractal approach can prove to be an extremely robust method to identify a portfolio's vulnerability to severe risks. Traditional "stress testing" is usually done by selecting an arbitrary number of "worst-case scenarios" from past data. It

> assumes that whenever one has seen in the past a large move of, say, 10 per cent, one can conclude that a fluctuation of this magnitude would be the worst one can expect for the future. This method forgets that crashes happen without antecedents. Before the crash of 1987, stress testing would not have allowed for a 22 per cent move.[18]

The Pareto PDF models can produce a wide range of CR values. Standard deviation, the normal PDF concept, does not always work well for a Pareto PDF. In theory, the standard deviation (as well as the mean) can actually be infinite. The recommendation is that in place of the calculated CR value, the modeler would report that the model indicates wild randomness (WR) or extreme randomness (ER). The suggestion is explained in the appendix.

Extreme value analysis does not, by design, permit a generalized look at a statistic like CR because it is fundamentally an approach that divorces the tail risk analysis from the data regarding the middle of the distribution that make up the mean and standard deviation. However, individual risk models that blend a model of expected variation around the mean with a specific model of the extremes based upon the generalized extreme value distribution can produce values which would lead to a CR calculation.

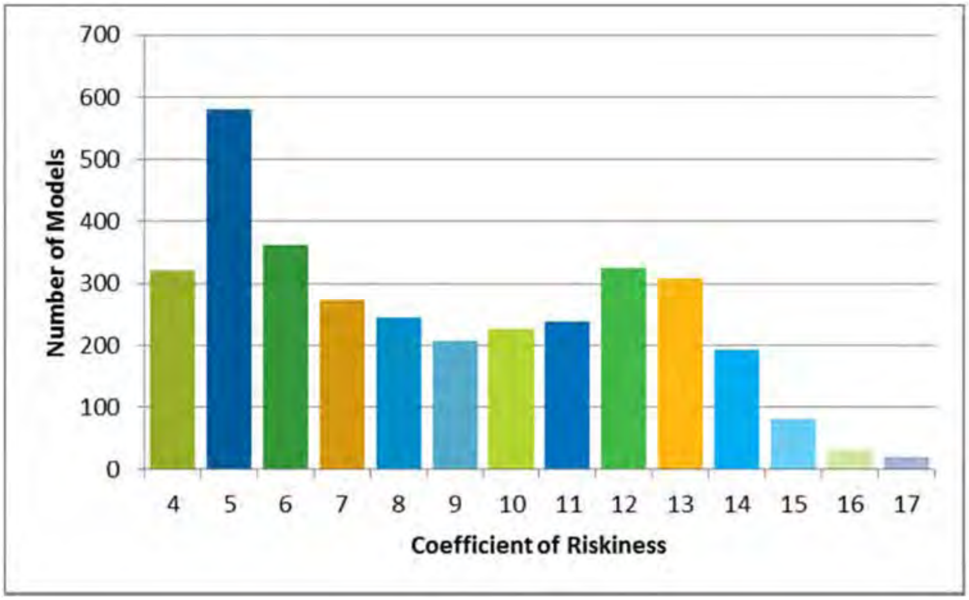
Examples from Insurance Risk Models

The author has obtained summary information from approximately 3,400 models of gross (before reinsurance) property and casualty insurance risks performed between 2009 and 2013 by actuaries at Willis Re.

---

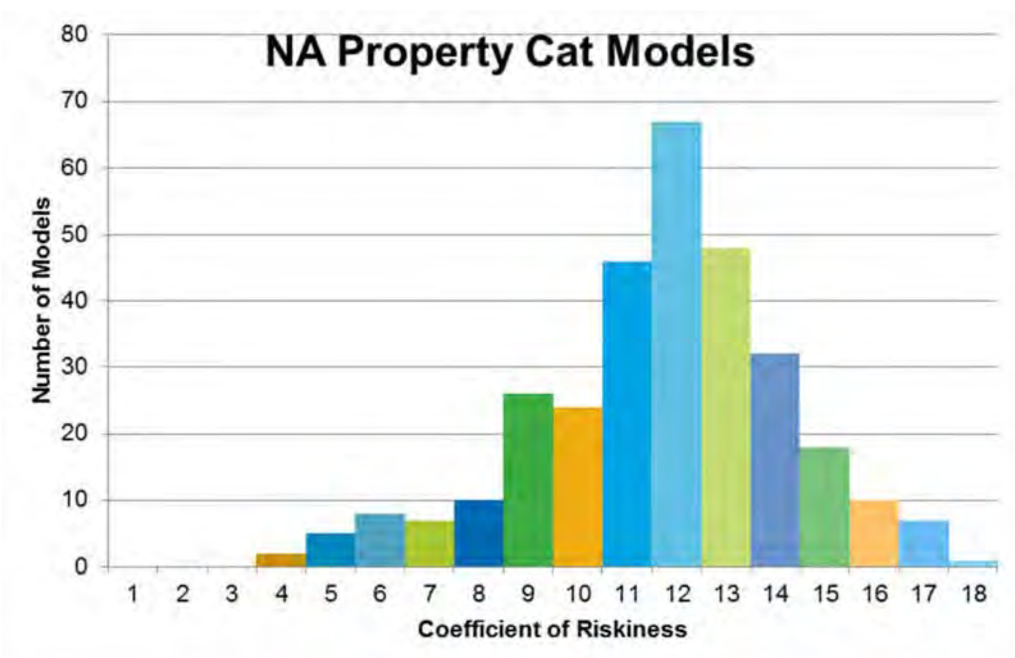
[18] Benoit Mandelbrot and Nassim Taleb, "A Focus on the Exceptions That Prove the Rule," *Financial Times*, March 23, 2006.

Figure 3. 3,400 insurance risk models[19]



In addition, we have obtained summary output from standalone natural catastrophe model runs for property insurance.

Figure 4. 400 natural catastrophe models



[19] For figures 3 and 4, the CR of 4, for example, indicates a value between 3 and 4.

It is interesting to note that none of these models showed a 99.9th percentile result that was 25 standard deviations. But, as you see, the natural catastrophe models did produce CR values as high as 18.

What you can see from these examples is that CR does seem to be bounded for these actual models into the range of 3 to 18 and that existing processes for modeling insurance risks do already produce a range of CR values.

A Simple Binomial Model

Some insight to the dynamics of CR can be reached by looking at models of small groups of independent risks that have low frequency.

Figure 5. CR for small binomial groups

n	p	Mean	Stdev	1 in 1000	COR
20	0.005	0.1	0.32	2.00	6.02
50	0.005	0.25	0.50	3.00	5.51
100	0.005	0.5	0.71	4.00	4.96
200	0.005	1	1.00	5.00	4.01
300	0.005	1.5	1.22	6.00	3.68
400	0.005	2	1.41	8.00	4.25
500	0.005	2.5	1.58	9.00	4.12
750	0.005	3.75	1.93	11.00	3.75
1000	0.005	5	2.23	13.00	3.59

If we start with looking at a group of 200 independent risk exposures that each have a likelihood of five in 1,000 of happening separately, the expectation is for one loss. The standard deviation would be one as well. The 99.9th percentile result would be for five losses, resulting in a CR of 4. That is slightly higher than the expected CR for the Poisson PDF of 3.5, and you see that as the group size gets larger, the CR gets closer to 3.5.

Figure 6. CR for a lower incidence rate

	Mean	Std Dev	CR	Economic Capital
Interest rate	12.6M	6.0M	4.5	6.9M
Equity	5.5M	10.0M	3.5	22M
Credit	2.5M	1.5M	6.0	3.5M
Underwriting: Property	20.0M	8.0M	12.2	36.8M
Underwriting: Auto	6.0M	2.5M	3.2	0.5M
Underwriting: Health	10.0M	8.0M	3.8	13.2M
Underwriting: All other	2.0M	0.7M	4.0	0.1M
Reserves	0.0M	$12.0M	4.3	37.8M
Operational	0.0M	0.1M	6.0	0.4M

Figure 6 shows it is possible to achieve somewhat higher CR with a group with a lower mean.

One hypothesis that could explain these simple calculations is that a risk which has a higher CR is susceptible to an extreme loss for a large fraction of the exposures when the expected loss is for a small fraction. You could say there is a concentrated exposure to the extreme event. Due to the concentrated exposure to the large event (hurricane or earthquake), in that event, their book of insurance contracts acts like a very small group of exposures. So the binomial view of these very small groups may well reproduce the experience of a large group with concentration.

Communicating Extreme Risk Inherent in Risk Models

Just walk a mile in his moccasins
Before you abuse, criticize and accuse.
If just for one hour, you could find a way
To see through his eyes, instead of your own muse.

(Mary T. Lathrap, 1895)[20]

All too often, the explanation for a model will be to identify the data used to parameterize the model. Sometimes, the result of the selection of PDF is mentioned, sometimes not. Rarely is there any discussion of the process for selecting the PDF used or the implications of that choice.

As mentioned above, nontechnical managers are usually familiar with the ideas of mean and standard deviation as the defining terms for statistical models. The coefficient of riskiness described here is proposed as a substitute for a discussion of the characteristics and implications of the selection of PDF that, in general, is needed but is not taking place.

The CR, if adopted widely, could come to be used similarly to the moment magnitude scale for earthquakes or the Saffir-Simpson Hurricane Wind Scale. If you were presenting a model of hurricanes or earthquakes and mentioned that you had modeled a 2 as the most severe event, everyone in the room would have a sense of what that meant, even if they do not know anything about the details of the modeling approach. They will have an opinion about whether a 2 is the appropriate value for the most severe possible hurricane or earthquake. They can easily participate in a discussion of the assumptions of the model on that basis.

The CR could become a similar tool for broad communication of model severity. If you believe that Viniar's comment about 25 standard deviations was actually based upon a measurement (rather than a round number exaggeration to make a point), then you would doubtless reject the validity of the model with a CR of 3 or 4. If nontechnical users of a risk model gained an appreciation of which of the company's

[20] *The Poems and Written Addresses of Mary T. Lathrap With a Short Sketch of her Life*, ed. Julia R. Parish (Michigan: The Women Christian Temperance Union, 1895).

risks have CR of 3 and which have 12, it would be a large leap of understanding of a very important characteristic of the risks.

So, as an illustrative example, an enterprise risk model might be described as follows:

Figure 7. Enterprise risk model: illustrative values only
(these do not represent any actual model)

	Enterprise Risk Model			
	Mean	Std Dev	CR	Economic Capital
Interest rate	12.6M	6.0M	4.5	6.9
Equity	5.5M	10.0M	3.5	22
Credit	2.5M	1.5M	6.0	3.5
Underwriting: Property	20.0M	8.0M	12.2	36.8
Underwriting: Auto	6.0M	2.5M	3.2	0.5
Underwriting: Health	10.0M	8.0M	3.8	13.2
Underwriting: All other	2.0M	0.7M	4.0	0.1
Reserves	0.0M	$12.0M	4.3	37.8
Operational	0.0M	0.1M	6.0	0.4
All risk (after diversification)	60.4M	37M	5.0	69.1

Then the discussion of the risk model can focus on the three sets of facts presented—the projected mean, the projected standard deviation and the fatness of the tail. These three facts about the model can be compared to similar facts about the past experience. What was the mean experience for each risk? What was the range of that experience as stated by the standard deviation? What is the historical fatness of the tail?[21] The discussion can then be all about why the model does or does not match up with past experience.

[21] The historical coefficient of riskiness can be defined as the historical worst case less the historical mean divided by the historical standard deviation. Since you will almost never have enough historical

The hope is that by turning away from the technical, statistical discussion about choice of PDF and parameterization, the discussion can actually tap into the extensive knowledge, experience and gut feel of the nontechnical management and board members. Perhaps the CR can become like the moment magnitude scale of risk models. Few people understand the science or math behind the moment magnitude scale, but everyone in an earthquake zone can experience a shake and come pretty close to nailing the score of that event without any fancy equipment. And they know how to prepare for a 4 or a 5 or a 6 quake. The same goes for the Saffir-Simpson scale.

Conclusion

> "Would you tell me, please, which way I ought to go from here?"
> "That depends a good deal on where you want to get to," said the Cat.
> "I don't much care where—" said Alice.
> "Then it doesn't matter which way you go," said the Cat.
> "—so long as I get SOMEWHERE," Alice added as an explanation.
> "Oh, you're sure to do that," said the Cat, "if you only walk long enough."
> (Lewis Carroll, *Alice in Wonderland*, 1865)

People naturally observe risk in the form of the range of experienced gains and losses. In statistical terms, those observations are represented by standard deviation. Statistical techniques that have long been applied to insurance company risks to develop central estimates are being used to calculate values in the extreme tails of the distribution of gains and losses. These processes are essentially an extrapolation from the "known" risk of volatility near the mean to "unknown" risk of extreme losses.

To date, there is no established language to talk about the nature of that extrapolation. The coefficient of riskiness described here is an attempt to bridge that gap. The CR can be used to differentiate risk models according to the fatness of the

experience to calculate a 99.9th percentile frequency, this discussion will always be about how much worse we each think it can get in the extreme.

tails and could become a standard part of our discussion of risk models. With the use of a metric like the CR, we believe the knowledge and experience of nontechnical management and board members can be brought into the discussions of risk model parameterization. The end result of such discussions will both ultimately improve the models and increase the degree to which they are actually relied upon for informing important decisions within a risk-taking enterprise.

Appendix

1. The Exponential Risk Model Problem

It was stated above that some exponential risk models will not fit with the CR calculation. That is a possible problem. The problem arises because in some models, the variance and perhaps the mean value is infinite.

Mandelbrot describes seven states of randomness

1. Proper mild randomness (the normal distribution)
2. Borderline mild randomness (the exponential distribution with $\lambda = 1$)
3. Slow randomness with finite and delocalized moments
4. Slow randomness with finite and localized moments (such as the lognormal distribution)
5. Prewild randomness (Pareto distribution with $\alpha = 2 - 3$)
6. Wild randomness: infinite second moment (variance is infinite; Pareto distribution with $\alpha = 1 - 2$)
7. Extreme randomness (mean is infinite; Pareto distribution with $\alpha \leq 1$)[22]

To solve that problems, some models use truncated exponential models. Truncated exponential models will have finite variance but might still have unstable sample values at the 99.9th percentile and therefore unstable CR.

Such extreme values as the 99.9th percentile are mainly used by insurers that use the tail value at risk (TVaR) as their primary risk metric. But, after saying that, those firms must have solved this problem in order to calculate the TVaR.

So we conclude there is a solution to this problem for any risk where the TVaR can be calculated. But we suggest extreme caution to any modelers dealing with wild

[22] Benoit B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (New York: Springer, 1997).

randomness (WR) or extreme randomness (ER). The CR is not calculable. But if their firm is really exposed to wild or extreme randomness, their problems are much larger than the reliability of their tail risk measure. We recommend that in a situation where the risk model does indicate wild or extreme randomness, the CR be reported as WR or ER. We also presume that a report with those indications will lead to very intense discussions of the risks being modeled.

2. Other Uses for CR

Risk modeling is a difficult and time-consuming process. If we develop a language around a tail risk metric such as CR, it would be possible to estimate risk model-type tail results by identifying the likely level of the CR for a risk and then combining that with the observed mean and standard deviation of actual experience. By turning risk calculation into a three-parameter problem where one parameter assures us that the tails will be appropriately "fat," then our risk model results can be easily and quickly estimated.

These quick estimates can be used for ready risk estimates and also for model validation. The validation can be to check the CR for each submodel against CR for other models of similar risks; the quick estimates described above can be an independent calculation. The model validator can develop a tolerance for deviation of actual model results from the quick estimate as a trigger for more in-depth examination of particular submodels.

Another possible area of application is for very quick estimates of economic capital model outcomes based upon aggregated historical experience. If experience with CR measures tell us, for example, that the CR for a certain line of business is usually in the range of 3.5 to 4.5, we can estimate the 99.9th percentile value from the historical standard deviation of results for that line in aggregate and then interpolate to get a 99th percentile or 99.5th percentile value. This might be useful in public data evaluations of insurers.

3. Choice of Metric

There is no specific science to the selection of the 99.9th percentile value at risk as the basis for the CR. We went beyond the common VaR points of 99.5 percent or 99 percent following the concept that surveyors use to find a point. They usually make their line to a point beyond the spot to be determined to reduce the chance of very local errors in the region of the desired point. This calculation could just as easily be made with 99th percentile or 99.5th percentile values. Some practitioners have suggested that such alternate values might be more stable with the number of actual simulation runs that are made for some of the risk models. The thinking behind the selection of 99.9 percent was that a one in 1,000 was definitely "in the tail" and, to look at fatness of tail, it might be better to look further out than the model values being used "all of the time." The choice was also influenced by the fact that the normal PDF value for 99.9 percent produced a CR of approximately 3, rather than the 2.575829304 at 99.5 percent or 2.326347874 at 99 percent. It seemed easier to talk about a metric that ran from 3 to 18, than for one that went from 2.576 to whatever.

Using a real value rather than an index has an advantage as well. First of all, if we think of the one-in-1,000 event as a worst-case event, then with using CR, we start by reminding folks the worst case is at least three times the standard deviation. This is important because often people are lulled into a false sense of security when some time goes by without the experience of any tail events. Then when we say a risk has a CR of 6, that means the worst case is simply six times the standard deviation worse than the mean. So if we expect things to mostly fall within one or two standard deviations, then the CR gives a sense of how much worse the worst case can be, without complicated multistep calculations.

The question of unstable values can be resolved. If a 99.9th percentile CR becomes a standard value, then occasionally the risk models can be left to run for more scenarios to produce stable values at that return period. But if that is not a viable solution because the models are simply unstable at that return period, then that is probably a limitation which needs to be understood by the users.

Finally, the thinking was that if CR would be used eventually to judge the reasonableness of another point in the tail such as 99.5 percent, then that validation was more powerful if it could be stated that the model results were reasonable out past that value and that the 99.5th percentile value was consistent with the 99.9th percentile value. By focusing solely on the 99.5th percentile value, modelers and model users run the risk that their models are not even viable at 99.51th percentile. And a focus on just that single metric is itself dangerous.[23]

However, even if these arguments for 99.9th percentile were compelling, it is highly likely that some models might adopt the idea but not the calibration. So we suggest that when the CR is calculated to anything other than 99.9th percentile, that be made clear with a subscript which states the percentile (i.e., $CR_{99.5\%}$).

4. Further Research

This paper is meant to be the introduction of one possible metric for fatness of tails. If there is sufficient interest in using this metric, then it should be tested against various standards of robustness for risk metrics, for example, the criteria for coherent risk measures.[24]

There could also be research into the range of CR values for different models of similar risks. How wide is the range of CR? What are the drivers of higher or lower CR values within a class of risks? How to predict the CR without actually modeling a risk?

[23] David Ingram, "Risk and Light," paper, 2010, http://ssrn.com/abstract=1594877.
[24] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. "Coherent Measures of Risk," *Mathematical Finance* 9, no. 3 (1999).

With that sort of information in hand, additional research might be undertaken to see if there is a reasonable way to take the three parameters we might have separate from a risk model—mean, standard deviation and CR—and create a simple distribution of gains and losses. That method might well be different for different levels of CR. The range of expected CR values determined independently of a model might also be a good piece of information to drive the actual selection of PDF for the risk model.

Very preliminary views of CR for full enterprise risk models that include both independent and interdependent risks suggest the effect of diversification is a smaller CR. Further research could be done to look at how the CR performs for combinations of independent and interdependent PDFs to see if there is any predictable reduction in CR from the combination.

Ultimately, this further research might lead to the conclusion that there is a better measure of tail risk. But it would be a good result if some tail risk measure that can be widely understood is widely adopted.