Mild vs. Wild Randomness: Focusing on those Risks that Matter

Benoit Mandelbrot & Nassim Nicholas Taleb

Forthcoming in, *The Known, the Unknown and the Unknowable in Financial Institutions* Frank Diebold, Neil Doherty, and Richard Herring, editors, Princeton: Princeton University Press.

Conventional studies of uncertainty, whether in statistics, economics, finance or social science, have largely stayed close to the so-called "bell curve", a symmetrical graph that represents a probability distribution. Used to great effect to describe errors in astronomical measurement by the 19th-century mathematician Carl Friedrich Gauss, the bell curve, or Gaussian model, has since pervaded our business and scientific culture, and terms like sigma, variance, standard deviation, correlation, R-square and Sharpe ratio are all directly linked to it. Neoclassical finance and portfolio theory are completely grounded in it.

If you read a mutual fund prospectus, or a hedge fund's exposure, the odds are that information will incorporate some quantitative summary claiming to measure "risk". That measure will be based on one of the above buzzwords that derive from the bell curve and its kin.

Such measures of future uncertainty satisfy our human's ingrained desire to "simplify" by squeezing into one single number matters that are too rich to be described by it. In addition, they cater to psychological biases and our tendency to understate uncertainty in order to provide an illusion of understanding the world.

The bell curve has been presented as "normal" for almost two centuries, even though its flaws have always been obvious to any practitioner with empirical sense[1]. Granted, it has been tinkered with using such methods as complementary "jumps", stress testing, regime switching or the elaborate methods known as GARCH, but while they represent a good effort, they fail to remediate the bell curve's irremediable flaws.

The problem is that measures of uncertainty using the bell curve simply disregard the possibility of sharp jumps or discontinuities. Therefore they have no meaning or consequence. Using them is like focusing on the grass and missing out on the (gigantic) trees.

In fact, while the occasional and unpredictable large deviations are rare, they cannot be dismissed as "outliers" because, cumulatively, their impact in the long term is so dramatic.

The good news, especially for practitioners, is that the fractal model is both intuitively and computationally simpler than the Gaussian. It too has been around since the sixties, which makes us wonder why it was not implemented before.

The traditional Gaussian way of looking at the world begins by focusing on the ordinary, and only later deals with exceptions or so-called outliers as

---

[1] There is very little of the "Normal" in the Gaussian: we seem to be conditioned to justify non Gaussianity –yet, it is often Gaussianity that needs specific justification. To justify it in finance by a Limit Theorem necessitates assumptions that the evidence has shown to be very restrictive, like independence or short-run memory, as well as other strictures that can rarely be verified. Likewise for the Poisson. Indeed there are many processes that generate Nonscalable randomness.

More technically, the idea that sums of i.i.d. finite variance random variables are Gaussian involves a *limit* or *asymptotic* behavior. Not only i.i.d observations are not the norm in natural life, something that was already observed by Yule in 1925, but in the long-run we shall all be dead and science is mostly concerned with pre-asymptotic results, including the speed at which the limit is reached. It is far, far slower than can be deemed acceptable in practice — see the Technical Appendix that follows this chapter. All that the Central limit Theorem asserts is that the limit is Gaussian within a narrow central band, it does not prevent nonGaussianity in the tails of the distribution.

ancillaries[2]. But there is also a second way, which takes the so-called exceptional as a starting point and deals with the ordinary in a subordinate manner – simply because that "ordinary" is less consequential.

These two models correspond to two mutually exclusive types of randomness: mild or Gaussian on the one hand, and wild, fractal or "scalable power laws" on the other. Measurements that exhibit mild randomness are suitable for treatment by the bell curve or Gaussian models, whereas those that are susceptible to wild randomness can only be expressed accurately using a fractal scale.

Let us first turn to an illustration of mild randomness. Assume that you round up 1,000 people at random among the general population and bring them into a stadium. Then, add the heaviest person you can think of to that sample. Even assuming he weighs 300kg, more than three times the average, he will rarely represent more than a very small fraction of the entire population (say, 0.3 per cent). Similarly, in the car insurance business, no single accident will put a dent on a company's annual income. These two examples both relate to the "Law of Large Numbers", which implies that the average of a random sample is likely to be close to the mean of the whole population. The entire sample theory is based on the idea.

In a population that follows a mild type of randomness, one single observation, such as a very heavy person, may seem impressive by itself but will not disproportionately impact the aggregate or total. A randomness that disappears

under averaging is trivial and harmless. You can diversify it away by having a large sample.

There are specific measurements where the bell curve approach works very well, such as weight, height, calories consumed, death by heart attacks or performance of a gambler at a casino. An individual that is a few million miles tall is not biologically possible, but with a different sort of variable, an exception of equivalent scale cannot be ruled out with a different sort of variable, as we will see next.

WILD RANDOMNESS

What is wild randomness[3]? Simply put, it is an environment in which a single observation or a particular number can impact the total in a disproportionate way. The bell curve has "thin tails" in the sense that large events are considered possible but far too rare to be consequential. But many fundamental quantities follow distributions that have "fat tails" – namely, a higher probability of extreme values that can have a significant impact on the total.

One can safely disregard the odds of running into someone several miles tall, or someone who weighs several million kilogrammes, but similar excessive observations can never be ruled out in other areas of life.

Having already considered the weight of 1,000 people assembled for the previous experiment, let us instead consider their wealths. Add to the crowd of 1,000 the wealthiest person to be found on the planet – Bill Gates, the founder of Microsoft. Assuming that his net worth is close to $80bn, how much would he represent of the total wealth? 99.9 per cent? Indeed, all the others would represent no more than the variation of his personal portfolio over the past few seconds. For someone's weight to represent such a share, he would need to weigh 30m kg.

---

[2] A key feature of the Pareto-Lévy-Mandelbrot fractal model is the presence of "jumps." Since then, in order to capture the outliers and conserve the results of neoclassical finance, Merton (1976) has "grafted" simplified jumps onto the Gaussian. This graft is used heavily in modern finance One principal flaw is that data generated from any given probability distribution seen ex post can be immediately interpreted as a Poisson. Literally, anything. See Mandelbrot (2001) for the discussion on overfitting, and *ad hoc* superpositions. Another way to see the unfitness of the Poisson is by testing it out of sample –it fits past data rather well, but does not carry forward.

[3] Technically, many levels of wildness are distinguished in Mandelbrot (1997), ranging from the purely mild to the totally wild. Here wild randomness means scalable, or, more exactly, absence of a known characteristic scale, about which later.

Try it again with, say, book sales. Line up a collection of 1,000 authors. Then, add the most read person alive, JK Rowling, the author of the Harry Potter series. With sales of several hundred million books, she would dwarf the remaining 1,000 authors who would collectively have only a few hundred thousand readers.

So, while weight, height and calorie consumption are Gaussian, wealth is not. Nor are income, market returns, size of hedge funds, returns in the financial markets, number of deaths in wars or casualties in terrorist attacks. Almost all man-made variables are wild. Furthermore, physical science continues to discover more and more examples of wild uncertainty, such as the intensity of earthquakes, hurricanes or tsunamis.

Economic life displays numerous examples of wild uncertainty. For example, during the 1920s, the German currency moved from three to a dollar to 4 trillion to the dollar in a few years. And veteran currency traders still remember when, as late as the 1990s, short- term interest rates jumped by several thousand per cent.

We live in a world of extreme concentration where the winner takes all. Consider, for example, how Google grabs much of internet traffic, how Microsoft represents the bulk of PC software sales, how 1 per cent of the US population earns close to 90 times the bottom 20 per cent or how half the capitalization of the market (at least 10,000 listed companies) is concentrated in less than 100 corporations.

Taken together, these facts should be enough to demonstrate that it is the so-called "outlier" and not the regular that we need to model. For instance, a very small number of days accounts for the bulk of the stock market changes: just ten trading days represent 63 per cent of the returns of the past 50 years.

Let us now return to the Gaussian for a closer look at its tails. The "sigma" is defined as a "standard" deviation away from the average, which could be around 0.7 to 1 per cent in a stock market or 8 to 10 cm for height. The probabilities of exceeding multiples of sigma are obtained by a complex mathematical formula. Using this formula, one finds the following values:

Probability of exceeding:
0 sigmas: 1 in 2 times
1 sigmas: 1 in 6.3 times
2 sigmas: 1 in 44 times
3 sigmas: 1 in 740 times
4 sigmas: 1 in 32,000 times
5 sigmas: 1 in 3,500,000 times
6 sigmas: 1 in 1,000,000,000 times
7 sigmas: 1 in 780,000,000,000 times
8 sigmas: 1 in 1,600,000,000,000,000 times
9 sigmas: 1 in 8,900,000,000,000,000,000 times
10 sigmas: 1 in 130,000,000,000,000,000,000,000 times
and, skipping a bit:
20 sigmas: 1 in 36,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000,000, 000 times

Soon, after about 22 sigmas, one hits a "googol", which is 1 with 100 zeroes behind it.

With measurements such as height and weight, this probability seems reasonable, as it would require a deviation from the average of more than 2m. The same cannot be said of variables such as financial markets. For example, a level described as a 22 sigma has been exceeded with the stock market crashes of 1987 and the European interest rate moves of 1992, not counting the routine devaluations in emerging market currencies.

The key here is to note how the frequencies in the preceding list drop very rapidly, in an accelerating way. The ratio is not invariant with respect to scale.

Let us now look more closely at a fractal, or scalable[4], distribution using the example of wealth.

---

[4] Technically for us a fractal distribution defined as follows: $P_{>x} = K x^{-\alpha}$ where $P_{>x}$ is the probability of exceeding a variable x and α is the asymptotic power law exponent for x large enough. This distribution is said to be scale free, in the sense that it does not have a characteristic scale: For x "large

We find that the odds of encountering a millionaire in Europe are as follows:

Richer than 1 million: 1 in 62.5
Richer than 2 million: 1 in 250
Richer than 4 million: 1 in 1,000
Richer than 8 million: 1 in 4,000
Richer than 16 million: 1 in 16,000
Richer than 32 million: 1 in 64,000
Richer than 320 million: 1 in 6,400,000

This is simply a fractal law with a "tail exponent", or "alpha", of 2, which means that when the number is doubled, the incidence goes down by the square of that number – in this case four. If you look at the ratio of the moves, you will notice that this ratio is invariant with respect to scale. If the "alpha" were one, the incidence would decline by half when the number is doubled. This would produce a "flatter" distribution (fatter tails), whereby a greater contribution to the total comes from the low probability events.

Richer than 1 million: 1 in 62.5
Richer than 2 million: 1 in 125
Richer than 4 million: 1 in 250
Richer than 8 million: 1 in 500
Richer than 16 million: 1 in 1,000[5]

---

enough," the relative deviation of $(P > x)/(P > nx)$ does not depend on x, only on n. Other distributions are non-scalable. For example, in the density $p(x) = \exp[\text{-}a\ x]$, with tails falling off exponentially, the scale will be $1/a$. For the Gaussian, the scale is the standard deviation.

The effect that is not negligible is that finite moments exist only up to the exponent α. Indeed, in order for the function $x^n\ x^{\alpha\text{--}1}$ to have a finite integral from 1 (say) to infinity, one must have $n-\alpha<0$, that is, $n<\alpha$. This does not allow the Taylor expansions required for Modern Portfolio Theory as, for scalable, higher terms are explosive. If $\alpha=3$, as we tend to observe in stocks, the third and higher moments are infinite.

[5] There is a problem of "how large" is "large". This scalability might stop somewhere –but we do not know where, so we might consider it infinite. The two statements, "very large but I don't know exactly how large" and "infinitely large" look different but are epistemologically substitutable (Taleb 2007b). There might be a point at which the distributions flip.

We have used the example of wealth here, but the same "fractal" scale can be used for stock market returns and many other variables –at least as a vague lower bound. In other words, this method provides an alternative qualitative method to the Gaussian.

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. Using a fractal method, it is easy to extrapolate multiple projected scenarios. If your worst-case scenario from the past data was, say, a move of −5 per cent and, if you assume that it happens once every two years, then, with an "alpha" of two, you can consider that a −10 per cent move happens every eight years and add such a possibility to your simulation.

Using this model, a −15 per cent move would happen every 16 years, and so forth. This will give you a much clearer idea of your risks by expressing them as a series of possibilities. You can also change the alpha to generate additional scenarios – lowering it means increasing the probabilities of large deviations and increasing it means reducing the probabilities. What would such a method reveal? It would certainly do what "sigma" and its siblings cannot do, which is to show how some portfolios are more robust than others to an entire spectrum of extreme risks. It can also show how some portfolios can benefit inordinately from wild uncertainty.

This will show once we look at them graphically in the appendix.

Despite the shortcomings of the bell curve, reliance on it is accelerating, and widening the gap between reality and standard tools of measurement. The consensus seems to be that any number is better than no number – even if it is wrong. Finance academia is too entrenched in the mild, Gaussian, paradigm to stop calling it "an acceptable approximation".

Let us repeat: the Gaussian (or Poisson) are _no_ approximation. Any attempts to refine the tools of modern portfolio theory by relaxing the bell curve assumptions, or by "fudging" and adding the occasional "jumps" will not be sufficient. We live in a world primarily driven by random jumps and tools designed for random walks address the wrong problem. It would be like tinkering with models of gases in an attempt to characterise them as solids and call them "a good approximation".

While scalable laws do not yet yield precise recipes, they have become an alternative way to view the world, and a methodology where large deviation and stressful events dominate the analysis instead of the other way around.

We do not know of a more robust manner for decision-making in an uncertain world.

TABLE: THE GAUSSIAN AND FRACTAL MODELS: OBSERVATIONS AND CONSEQUENCES

1 By itself, no single number can characterize uncertainty and risk but, as we have seen, we can still have a handle on it so long as we can have a table, a chart and an open mind.

2 In the Gaussian world, standard tables show that 67 per cent of the observations fall between –1 and +1 sigma. Outside of Gaussianity, sigma loses much or all of its significance. With a scalable distribution, you may have 80 percent, 90 per cent, even 99.99 per cent of observations falling between -1 and +1 sigmas. In fractals, the

standard deviation is never a "typical" value and may even be infinite[6]!

3 When assessing the effectiveness of a given financial, economic or social strategy, the observation window needs to be large enough to include substantial deviations, so one must base strategies on a long time frame. In some situations you will never see the properties.

4 You are far less diversified than you assume. Because the market returns in the very long run will be dominated by a small number of investments, you need to mitigate the risk of missing these by investing as broadly as possible. Very broad passive indexing is far more effective than active selection.

5 Projections of deficits, performance and interest rates are marred with extraordinarily large errors. In many budget calculations, US interest rates were projected to be 5 per cent for 2001 (not 1 per cent); oil prices were projected to be close to $22 a barrel for 2006 (not $62). Like prices, forecast errors follow a fractal distribution.

6 Option pricing models, such as Black-Scholes-Merton, are strongly grounded in the bell curve in their representation of risk. The Black-Scholes-Merton equation bases itself on the possibility of eliminating an option's risk through continuous dynamic hedging, a procedure incompatible with fractal discontinuities[7].

7 Some classes of investments with explosive upside, such as venture capital, need to be favored over those that do not have such potential. Technology investments get bad press; priced appropriately (in the initial stages) they can

---

[6] Even in "finite variance" cases where $\alpha > 2$, we just can no longer rely on variance as a sufficient measure of dispersion. See the Technical Appendix and also Cont (2005) for an illustration of how a cubic exponent mimics stochastic volatility and can be easily mistaken for it.

[7] Taleb (2007a) shows how Ito's lemma no longer applies and how we can no longer perform the operation of dynamic hedging to compress the risks.

deliver huge potential profits, thanks to the small, but significant, possibility of a massive windfall[8].

8 Large moves beget large moves; markets keep in memory the volatility of past deviations. A subtle concept, fractal memory provides an intrinsic way of modelling both the clustering of large events and the phenomenon of regime switching, which refers to phases when markets move from low to high volatility[9].

TABLE: COMPARISON BETWEEN SCALABLE AND NONSCALABLE RANDOMNESS

Non scalable	Scalable
The most typical member is mediocre	The most 'typical" is either giant or dwarf, i.e. there is no typical member
Winners get a small segment of the total pie	Winner-take-almost-all effects
Example: Audience of an opera singer before the gramophone	Today's audience for an artist
More likely to be found in our ancestral environment	More likely to be found in our modern environment
Subjected to gravity	There are no physical constraints on what a number can be
Corresponds (generally) to physical quantities, i.e. height	Corresponds to numbers, say wealth
Total is not determined by a single instance or observation.	Total will be determined by a small number of extreme events.
When you observe for a while you can get to know what's going on	It takes a long time to know what's going on
Tyranny of the collective	Tyranny of the accidental
Easy to predict from what you see to what you do not see	Hard to predict from past information
History crawls	History makes jumps

TECHNICAL APPENDIX
LARGE BUT FINITE SAMPLES
AND PREASYMPTOTICS

Ever since 1963, when power law densities first entered finance through the Pareto-Lévy-Mandelbrot model, the practical limitations of the limit theorems of probability theory have raised important issues. Let the tail follow the power-law distribution defined as follows: $P_{>x} = K x^{-\alpha}$ where $P_{>x}$ is the probability of exceeding a variable x and α is the asymptotic power law exponent for x large enough. If so, a first partial result is that the largest of n such variables is given by an expression ("Fréchet law")that does not depend on α. This maximum is well-known to behave like $n^{1/\alpha}$. A second partial result is that the sum of n variables is given by an expression that — to the contrary — does depend on the sign of $\alpha-2$.

If $\alpha > 2$, the variance is finite — as one used to assume without thinking. But what does the central limit theorem really tell us? Assuming EX=0, it includes the following classical result: EX is finite and there exists near EX a central bell region in which the sum is increasingly close to a Gaussian whose standard deviation behaves asymptotically like $n^{1/2}$. Subtracting nEX from the sum and combining the two partial results, one

---

[8] The exact opposite applies to business we call "concave" to the large deviation, such as banking, catastrophe insurance, or hedge-fund arbitrage of the type practiced by Long Term Capital Management.

[9] This power-law decaying volatility differs markedly from the exponentially declining memory offered by ARCH methods.

finds that the relative contribution of the largest addend behaves like $n^{1/\alpha-1/2}$. In the example of $\alpha=3$, this becomes $n^{-1/6}$. Again asymptotically for $n\to\infty$, this ratio tends to 0 — as expected — but the convergence is exquisitely slow. For comparison, examine for $EX\neq0$ the analogous very familiar ratio of the deviation from the mean — to the sum if the former behaves like the standard deviation times $n^{1/2}$. The latter — assuming $EX\neq0$ — behaves like nEX. Therefore these two factors' ratio behaves like $n^{-1/2}$. To divide it by 10, one must multiply n by 100, which is often regarded as uncomfortably large. Now back to $n^{-1/6}$: to divide it by 10, one must multiply n by 1,000,000. In empirical studies, this factor is hardly ever worth thinking about.

Now consider the — widely feared — case $\alpha<2$ for which the variance is infinite. The maximum's behavior is still $n^{1/\alpha}$, but the — subtracting nEX —sum's behavior changes from $n^{1/2}$ to the "anomalous" $n^{1/\alpha}$. Therefore, the relative contribution of the largest addend is of the order $n^{1/\alpha-1/\alpha}=n^0$. Adding all the bells and whistles, one finds that the largest addend remains a significant proportion of the sum, even as n tends to infinity.

Conclusion: In the asymptotic regime tackled by the theory, n^0 altogether differs from $n^{-1/6}$, but in the preasymptotic regime within which one works in practice — especially after sampling fluctuations are considered — those two expressions are hard to tell apart. In other words, the sharp discontinuity at $\alpha=2$, which has created so much anguish in finance — is replaced in practice by a very gradual transition. Asymptotically, the Lévy stability of the Pareto-Lévy-Mandelbrot model remains restricted to $\alpha<2$ but preasymptotically it continues to hold if α is not far above 2.

FIGURES

The next two figures show the representation of the scalable in the tails.
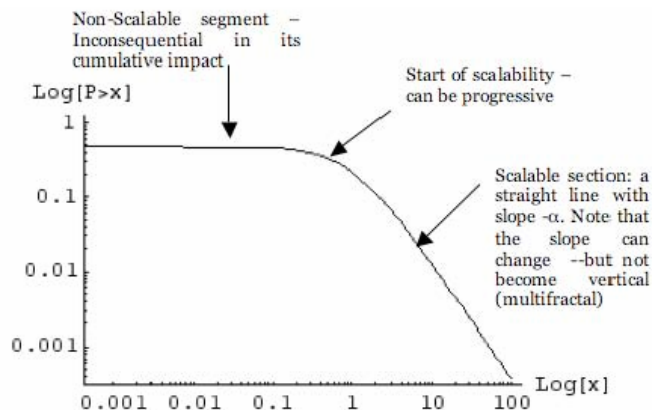
Figure 1 Looking at a distribution. Log P>x = $-\alpha$ Log X $+C^t$ for a scalable. When we do a log-log plot (i.e., plot P>x and x on a logarithmic scale), as in Figures 1 and 2, we should see a straight line in the asymptote.
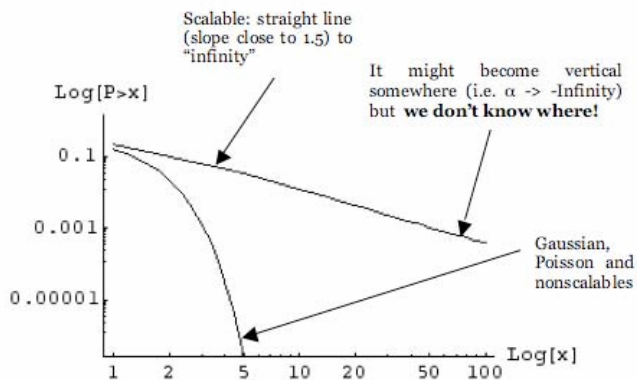
Figure 2 The two exhaustive domains of attraction: vertical or straight line with slopes either negative infinity or constant negative α. Note that since probabilities need to add-up to 1 there cannot be other alternatives to the two basins, which is why we narrow it down to these two exclusively –as we said the two paradigms are mutually exclusive.

REFERENCES

Cont Rama & Peter Tankov, 2003, *Financial Modelling with Jump Processes*, Chapman & Hall / CRC Press, 2003.

Mandelbrot, Benoit, x

Mandelbrot, Benoit, 1967, Sur l'épistémologie du hasard dans les sciences sociales: invariance des lois et vérification des hypothèses, *Encyclopédie de la Pléiade* (Gallimard). *Logique et Connaissance Scientifique* (Dirigé par J. Piaget), 1097-1113.

Mandelbrot, Benoit, 1997. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. New York: Springer-Verlag. Reproduces earlier classic papers.

Mandelbrot, Benoit, 2001. Stochastic volatility, power-laws and long memory. *Quantitative Finance*: **1**, 558-559.

Mandelbrot, Benoit & Richard L. Hudson, 2004. *The (Mis)behavior of Markets: A Fractal View of Risk, Ruin, and Reward*. New York: Basic Books.

Taleb, Nassim Nicholas, 2007, "Scale Invariance in Practice", Working Paper.