Companion Volume to *Antifragility* (forth) and *The Black Swan (2007-2010,2nd Ed.)*

Technical and Academic Papers and Derivations

Nassim Nicholas Taleb
NYU-Poly

2012

Table of Contents

A Map and Simple Heuristic to Detect Fragility, Antifragility, and Model Error

N. N. Taleb
New York University -Polytechnic Institute

First Version, June 4, 2011

This Paper is to be presented at:

JP Morgan, New York, June 16, 2011; CFM, Paris, June 17, 2011; GAIM Conference, Monaco, June 21, 2011; Max Planck Institute, BERLIN, Summer Institute on Bounded Rationality 2011 - *Foundations of an Interdisciplinary Decision Theory*- June 23, 2011; Eighth International Conference on Complex Systems - BOSTON, July 1, 2011.

Abstract

The main results are 1) definition of fragility, antifragility and model error (and biases) from missed nonlinearities and 2) detection of these using a single "fast-and-frugal", model-free, probability free heuristic. We provide an expression of fragility and antifragility as negative or positive sensitivity to convexity effects, i.e., dispersion and volatility (a variant of negative or positive "vega") across domains and show similarities to model errors coming from missing hidden convexities -model errors treated as left or right skewed random variables. Broadening and formalizing the methods of *Dynamic Hedging*, Taleb (1997), we present the effect of nonlinear transformation (convex, concave, mixed) of a random variable with applications ranging from exposure to error, tail events, the fragility of porcelain cups, deficits and large firms and the antifragility of trial-and-error and evolution. The heuristic lends itself to immediate implementation, and uncovers hidden risks related to company size, forecasting problems, and bank tail exposures (it explains the forecasting biases). While simple, it vastly outperforms stress testing and other such methods such as Value-at-Risk.

Introduction:

Main practical result of this paper: *a risk heuristic that "works" in detecting fragility even if we use the wrong model/pricing method/probability distribution.* The main idea is that ***a wrong ruler will not measure the height of a child; but it can certainly tell you if he is growing*.** Since (as we will see) risks in the tails map to nonlinearities (concavity of exposure), second order effects reveal fragility, particularly in the tails (revealed through perturbation)where they map to large tail exposures.

Further, the misspecification in using thin-tailed distributions (say the Gaussian) shows immediately through perturbations of standard deviation when it appears to be unstable. Further here are results that shows how fat-tailed (powerlaw tail) probability distribution can be expressed by simple perturbation and mixing of the Gaussian.

Why the same heuristic (detection of convexity effects) can measure both fragility and model error: Where F is a valuation "model",

$$\text{Model (or Valuation) Error} = \mathbf{E_1} + \mathbf{E_2} + \mathbf{E_3}$$

where we asume that the three types of errors are orthogonal hence additive.

$\mathbf{E_1}$= linear error, the "slope", an error about the first derivative of the model with respect to a variable (equivalent of the delta for an option), say $a = \frac{F[x+\Delta x]-F[x]}{\Delta x}$. The model identifies the parameter a, but has a wrong value for such parameter in, say, a regression. One can safely believe that modelers cannot easily make such error (the results of the mistracking will be immediately visible).

$\mathbf{E_2}$= missing a stochastic variable determining F. We unfortunately do not deal with that in this paper, but have evidence (Makridakis et al, 1982; Makridakis and Hibon, 2000) that, if anything, models by overly insample fitting, include too many variables, not too few.

$\mathbf{E_3}$ (Procrustean Bed)= missing convexity effects, the "hidden gamma", that is, a) missing the stochastic character of a variable deemed deterministic (and fixed) and b) F is convex or concave with respect of such variable. The resulting bias causes misestimation of F , with undervaluation or overvaluation that maps to the nonlinearity. Such error being rare (and compounded by those rare large deviations), it is likely to be missed.

Example of $\mathbf{E_3}$. A government estimates unemployment for the next three years as averaging 9%; it uses its econometric models to issue a forecast balance B of 200 billion deficit in the local currency. But it misses (like almost everything in economics) that unemployment is a stochastic variable. Employment over 3 years periods has fluctuated by 1% on average. We can calculate the effect of the error with the following:

- Unemployment at 8% , Balance B(8%) = -75 bn (improvement of 125bn)
- Unemployment at 9%, Balance B(9%)= -200 bn

- Unemployment at 10%, Balance B(10%)= --550 bn (worsening of 350bn)

So E_3 is the convexity bias from underestimation of the deficit is by -112.5bn, since $\frac{B(8\%)+B(10\%)}{2}$ = -312.5

Further look at the probability distribution caused by the missed variable (assuming to simplify deficit is Gaussian with a Mean Deviation of 1%)

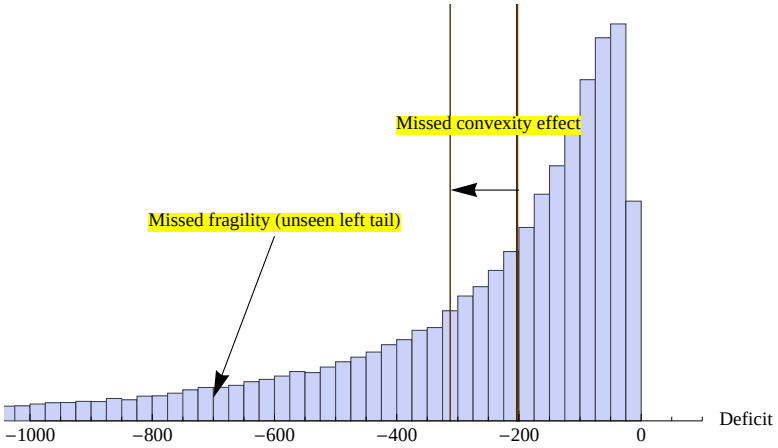


Figure 1**CONVEXITY EFFECTS ALLOW THE DETECTION OF BOTH MODEL BIAS AND FRAGILITY**. Illustration of the example; histogram from Monte Carlo simulation of government deficit as a left-tailed random variable simply as a result of randomizing unemployment of which it is a convex function. The method of point estimate would assume a Dirac stick at -200, thus underestimating both the ***expected*** deficit (-312) and the skewness (i.e., fragility) of it.

Most significant (and preventable) model errors, as we will see, arise from E_3.

Now this paper will focus on a heuristic that can both detect Fragility and E_3since our definition of fragility is grounded in nonlinearities. Further, the "fat tailedeness" of probability distributions is a straight application of E_3, the missing of a convexity effect.

Nonlinearity and Fragility: Simply, for the fragile, *shocks bring higher and higher harm as their intensity increases (up to the point of breaking)*. Another example. For a collision, forty miles per hour causes more than four times the harm of ten miles per hour. Jumping from a level 30 feet high is more harmful than jumping 10 times from 3 feet.

Every payoff one can think of in nature is nonlinear, hence subjected to some tail payoff, and some asymmetry in its distribution. And every model has some kind of Procrustean bed-style sucker problem coming with it, some error from missing the stochasticity of some variable and the nonlinear character of the payoff.

The object here is to detect fragility (and, by the same process, to detect its opposite, antifragility, ability to gain from disorder). The same method that detects fragility can detect convexity biases, or model error stemming from missing the stochasticity of a variable, as well as sensitivity to the use of the wrong probability distribution.

Our steps are as follows:

a. We define fragility, robustness and antifragility.

b. We presents the problem of measuring tail risks and show the presence of severe biases attending the estimation of small probability and its nonlinearity (convexity) to parametric (and other) perturbations .

c. We express the concept of model fragility in terms of left tail exposure, and show correspondence to the concavity of the payoff from a random variable.

d. Finally, we present our simple heuristic to detect the possibility of both fragility and model error across a broad range of probabilistic estimations.

The central Table 1 introduces the exhaustive map of possible outcomes, with 4 exhaustive mutually exclusive categories of payoffs. The end product is *f(x)*, which can be reduced to a scalar, and is the central variable of concern. We consider both the probability distribution of f(x) the payoff function, a "derivative" function of *x*, *x* being a "primitive" random variable, and the functional properties (concave, convex, linear).

We present a series of arguments that can be proved (owing to the format of the discussion, some idiot-savant "quants" might not recognize the proof, so try a bit harder to adapt to the language).

Note about the lack of symmetry between fragility and antifragility. By shrinking the left tail (in the presence of unbounded positive payoffs) you cause antifragility; but by increasing the right tail you don't reduce fragility.

Definition and Map of Fragility, Robustness, and Antifragility

Table 1- Introduces the Exhaustive Taxonomy of all Possible Payoffs y=f(x)

Type	Condition	Left Tail (loss domain)	Right Tail (gains domain)	Nonlinear Payoff Function $y = f(x)$ "derivative", where x is a random variable	Derivatives Equivalent (Taleb, 1997)	Effect of Jensen's Inequality on Missed Nonlinearities	Effect of Fat tails in Distribution of primitive x
Type 1	**Fragile (type 1)**	Fat	Thin	Concave	Short gamma	Lower expectation	Worsens
Type 2	**Fragile (type 2)**	Fat (regular or absorbing barrier)	Fat	Mixed concave left, convex right (fence)	Long up – gamma, short down – gamma	Invariant, or Lower expectation in case of absorbing barrier	Worsens if absorbing barrier, neutral otherwise
Type 3	**Robust**	Thin	Thin	Mixed convex left, concave right (digital, sigmoid)	Short down – gamma, long up – gamma	Invariant	Invariant
Type 4	**Antifragile**	Thin	Fat (Thicker than left)	Convex	Long gamma	Raises expectation (particularly in Type 4 *b* where trigger barriers cause ratchet – like properties)	Improves

Definition of Fragility

Fragility $\Leftrightarrow$ Left Tail $\Longleftarrow$ Concavity (Table 1)

Fragility is defined as equating with sensitivity of left tail shortfall (non conditioned by probability) to increase in disturbance over a certain threshold K

Examples

- **a.** Example: a porcelain coffee cup subjected to random daily stressors from use.
- **b.** Example: tail distribution in the function of the arrival time of an aircraft.
- **c.** Example: hidden risks of famine to a population subjected to monoculture.
- **d.** Example: hidden tail exposures to budget deficits' nonlinearities to unemployement.
- **e.** Example: hidden tail exposure from dependence on a source of energy, etc. ("squeezability argument").

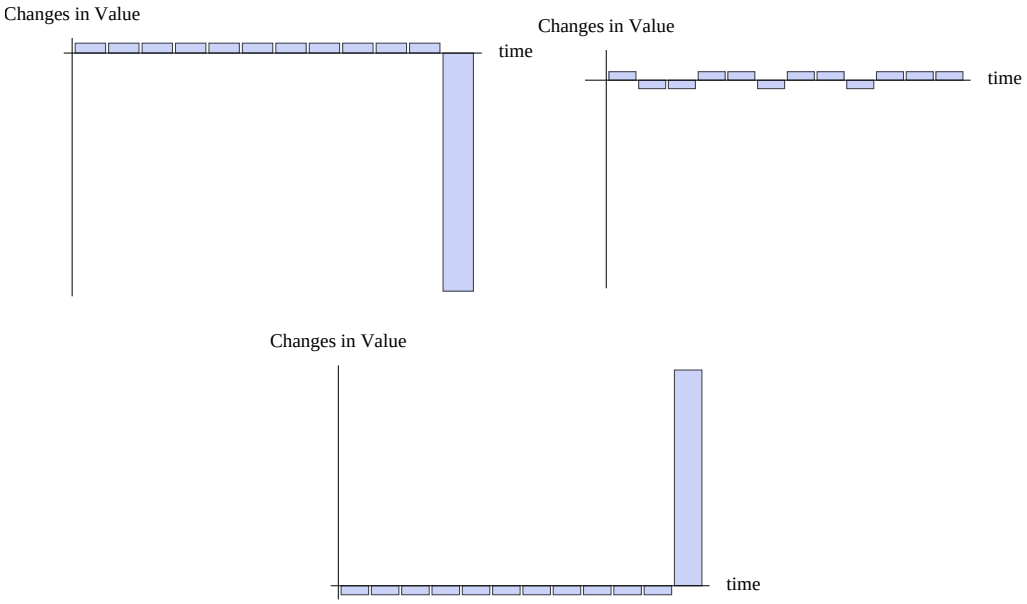


Figure 2 TYPE 1, Fragile variations through time —the horizontal axis shows time. This can apply to anything, a coffee cup, a health indicator, changes in wealth, your happiness, etc. We can see small (or no) benefits most of the time and occasional large adverse outcomes. Uncertainty can hit in a rather hard way. Notice that the loss can occur at any time and exceed the previous cumulative gains.

Figure 3 TYPE 3, the Just Robust (but not antifragile)- It experiences small or no variations through time. Never large ones.

Figure 4 TYPE 4, The antifragile system: uncertainty benefits a lot more than it hurts—the exact opposite of Figure 1.

Left Tail and Measure of Fragility

In short, fragility is negative exposure to left uncertainty as measured by some coefficient of dispersion (STD, MAD, etc.) . We will define it first and then link it to convexity.

Definition 1a (standard and monomodal distributions)*: where* y *and* z *are random variables, exposure to y is said to be more "fragile" than exposure to* z *in tail K if, for a given K in the negative (undesirable) domain,*

$$V(y,\ f,\ K,\ \Delta s) > V(z,\ g,\ K,\ \Delta s) \tag{1}$$

where f *and* g *are the respective monomodal probability distributions for y and z,*

$$V(y,\ f,\ K,\ \Delta s) \equiv \zeta\left(y,\ f,\ K,\ s + \frac{\Delta s}{2}\right) - \zeta\left(y,\ f,\ K,\ s - \frac{\Delta s}{2}\right) \tag{2}$$

$$\zeta(y,\ f,\ K,\ s) \equiv \int_{-\infty}^{K} y\, f\,(y)\, dy \tag{3}$$

s is a dispersion parameter "volatility" used by the probability distribution *f* and *g*, and Δ*s* is a set variation, a finite perturbation. The discussion in the next section on convex-concave situations shows why we rely on a finite perturbation Δ*s* instead of the infinitesimal mathematical derivative.

Sources of Fragility: *y* is a function, that is, a derivative of some primitve *x* but let us not concern ourselves with *x* for now (we will look at it when we analyze convex transformations). For now we can say that the distribution of *f* is limited to source of variation *x*, and that fragilities from other sources are not taken into account here.

For instance, *s* can be the standard deviation or mean deviation for finite moment distributions, or tail exponent for a powerlaw tailed one (tail exponents subsume mean deviations and have an inverse relationship to deviation parameter for tail exponent >1). For the rare cases of ζ not existing, say when the outcome's distribution is Cauchy, there is no need to go further as it can be deemed *infinitely* or, *unconditionally* fragile, regardless of the properties of the right tail.

Fragility is K-specific. We are only concerned with adverse events below a certain prespecified level, the breaking point. Exposures A can be more fragile than exposure B for K=0, and much less fragile if K is, say, 4 mean deviations below 0. Option traders would recognize fragility as negative "vega", or negative exposure to volatility. The use of finite Δ is to avoid situations as we will see of vega-neutrality coupled with short left tail.

Applying the measure to the examples in Figures 1 through 3: Figure 1 has negative sensitivity to dispersion, Figure 2 is neutral, Figure 3 gains from volatility.

Deal with payoff functions: *y* is a payoff function of another random variable, which might itself be symmetric and thin-tailed, but of concern

is the distribution of y, which requires transformation. For instance a call price is a funtion of another random variable, the underlying security (which itself may be a function of another r.v.), but of concern is the distribution of the call.

Effect of using the wrong distribution *f*: Comparing V(y, f, K, Δs) and the alternative distribution V(y, f*, K, Δs) , where f* is the "true" distribution, the measure of fragility is acceptable ("robust") under the following conditions:

a. that both distributions are monomodal (the condition of using *V* as a measure of fragility), or

b. that the difference between the two, that is, the bias does not reverse in sign in the tails, or

c. that the sign of higher differences $\Delta_n \neq 0$ for all orders *n* do not carry opposite signs.

$$\text{sgn}(\Delta_n) = \text{sgn}(\Delta_{n-1}) \text{ for all } n$$

where

$$\Delta_1 = \left\{V\left(y,\ f,\ K,\ \Delta s - \frac{\Delta s}{2}\right) - V\left(y,\ f,\ K,\ \Delta s + \frac{\Delta s}{2}\right)\right\} - \left\{V\left(y,\ f^*,\ K,\ \Delta s - \frac{\Delta s}{2}\right) - V\left(y,\ f^*,\ K,\ \Delta s + \frac{\Delta s}{2}\right)\right\}$$

Unconditionality of the measure of shortfall *ζ*: Many, when presenting shortfall, deal with the conditional shortfall $\frac{\int_{-\infty}^{K} y\, f(y)\, dy}{\int_{-\infty}^{K} f(y)\, dy}$; while this measure might be useful in some circumstances, its sensitivity is not at all indicative of fragility in the sense used in this discussion. The unconditional tail expectation *ζ*, $\int_{-\infty}^{K} y\, f(y)\, dy$ is more indicative of exposure to fragility. It is also preferred to the raw probability of falling below K, $\int_{-\infty}^{K} f(y)\, dy$ as the latter does not include the consequences. For instance, two such measures $\int_{-\infty}^{K} f(y)\, dx$ and $\int_{-\infty}^{K} g(y)\, dy$ can be equal over broad values of K; but the expectation $\int_{-\infty}^{K} y\, f(y)\, dy$ can be much more consequential as the cost of the break can be more severe and we are interested in its "vega" equivalent.

Exception: the case of non monomodal and truncated distributions

Another definition is necessary for non unimodal distribitions. Measures of dispersion (and higher order ones) do not work well in the presence of polarized mixture distributions. Nor do they do well with payoffs subjected to absorbing barriers.

Definition 1b (for absorbing barriers, hence bimodal and multimodal distributions in probability space)*:where* y *and z are random variables, exposure to y is said to be more "fragile" than exposure to z below "tail" K if, for a given K in the negative (undesirable) domain, the costs of hitting a barrier L below K is higher for y than z.*

So Definition 1b can be made similar as definition 1a, except that we would have to perturbate, V(y,f,K,Δp) where p is the parameter setting the distance of lower mode (lower) away from the mean (for mixed distributions, the lower mean; for barriers, the cost of hitting the barrier).

Consider the stochastic process S_t, S_0>L, with absorbing barrier L from below, so with probability one it should break, but with some distribution of stopping time. Further, there is a "cost" attached to breaking. For a coffee-cup, a silk tie, a computer, a mirror, a corporation, the value can be assumed to go to close to 0 (0 plus some minor residual). The idea also applies to debt and squeezes (death, famine, brankrupcies, etc. are absorbing barriers).

Knock In: Antifragility would be a trigger-barrier causing ratchet-like properties (biological systems with irreversibilities), for the process S_t, S_0< H, with an inverse "cost", like a benefit upon hitting the barrier.

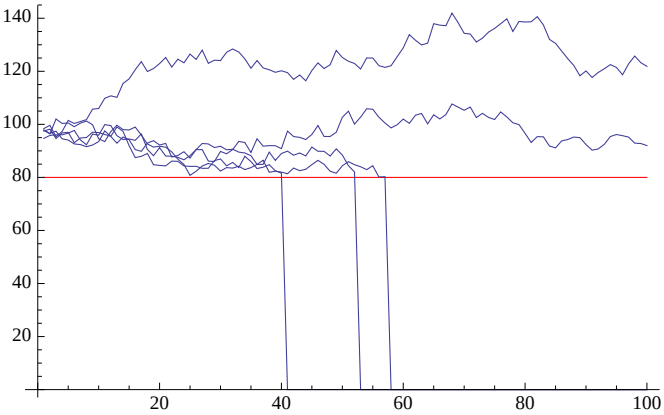
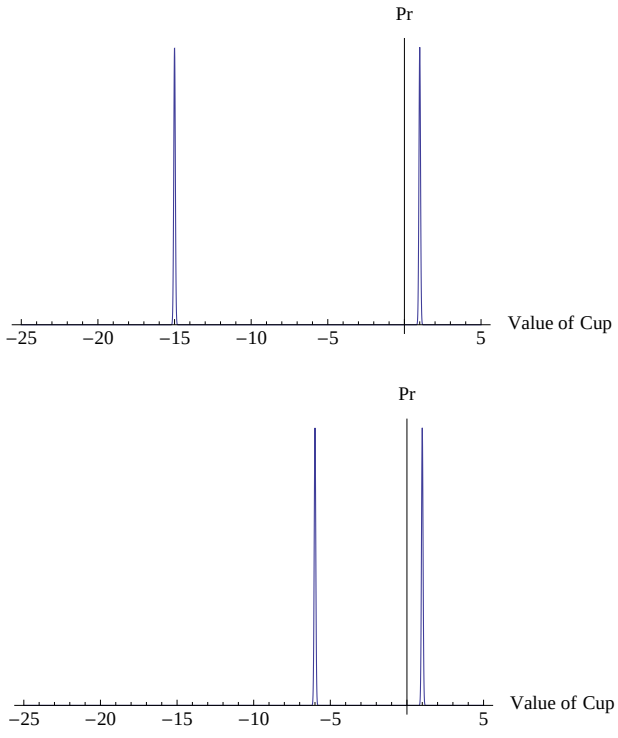


Figure 5 : L=80, S_0 = 100. Absorbing barrier (down and out) causes a special brand of left tails, when the barrier causes a collapse to a certain value (here, 0, with no residual)

The next graphs show the results of hitting a barrier in probability space, with standard bimodal, double-peaks.



Figures 6 and

Figure 7 : Bimodal Distributions (two gaussians with different means). The comparative fragility of two coffee cups, with their states as two Diracs (breaks or doesn't break). Each breaks at a given level. They both have the same probability of breaking, but a different ζ. The distribution on the left, although patently more fragile, does not respond to changes in STD They are both invariant to changes in dispersion parameter, yet the one on the left is more fragile.

Nor does kurtosis seem to matter. As Figure 6 shows, the fragile is not necessarily higher on the measure of kurtosis.

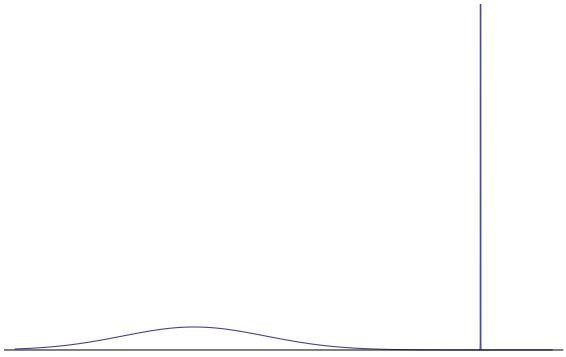


Figure 8 : A mixed distribution with two Gaussians of different means and STD (actually, the stick on the right approaches a Dirac). The most likely position is either in the “stick” or in the “breaking” section to the right as there is no mass in between. Although it appears extremely skewed (hence fragile), the Kurtosis is lower than that of a Gaussian.

The distribution in Figure 7 is vastly more common than accepted (bond returns, loans, stock mergers, etc.)

So the presence of a right tail does not matter: Jensen’s inequality will lower expectations.

Next, because *f* and *g* can be misspecified probability distributions (i.e., further from the “true” *f* and *g*):

Adding Model Error and Metadistributions: Model error should be integrated in the distribution as a stochasticization of parameters.

We will see that *f* and *g* should subsume the distribution of all possible factors affecting the final outcome (including the metadistribution of each). The so-called "perturbation" is not necessarily a change in the parameter, so much as it is a means to verify if *f* and *g* capture the full shape of the final probability distribution.

Note that, something with a bounded payoff, and a function that organically truncates the left tail at K will be impervious to all perturbations affecting the probability distribution below K.

For K=0, the measure equates to mean negative semi-deviations (more potent than negative semivariance used in financial analyses).

Definition of Antifragility

Antifragility is not the opposite of fragility, as we saw in Table 1. It requires thin left-tail (which we will define in exponential decline of probabilities) and local convexity, expressed as positive sensitivity to dispersion parameter of the probability distribution, the "long vega".

Antifragility requires robustness in the left tail, in addition to positive asymmetry (thicker right tail). Here we insist in limiting to a source of randomness x (or more), with y the end result a function of x (so, again, we confine robustness to a given source of variation).

Definition 2a, Left-Robustness (monomodal distribution). *A payoff y is robust below K for source of randomness x included in determining distribution f if*

$$| V(y, f, K, 2\Delta s) - V(z, g, K, \Delta s) | < e \tag{4}$$

where f *is the monomodal probability distributions for y in (3) and ζ(y,f,K,s) is the payoff below K, and e is a quantity of order deemed of "negligible utility" (subjectively), that is, does not exceed a tolerance level.*

Note that robustness is in effect impervious to changes of probability distributions. Also note that this measure robustness ignores first order variations E_1 since these are detected (and remedied) very early on.

Example of Robustness (Barbells):

- **a.** trial and error with bounded error and open payoff
- **b.** for a "barbell portfolio" with allocation to numeraire securities up to 80% of portfolio, no perturbation below K set at .8 of valuation will represent any difference in result, e=0. The same for an insured house (assuming the risk of insurance company is not a source of variation), no perturbation for the value below K will result in significant changes.
- **c.** a bet of amount B (limited liability) is robust, does not have any sensitivity to perturbations below 0.

Definition 2b, Antifragility (monomodal distribution). *A payoff y is locally antifragile over range x=L and x=H if*

y is robust below L

and

$$\lambda\left(y, f, L, H, s + \frac{\Delta s}{2}\right) - \lambda\left(y, f, L, H, s - \frac{\Delta s}{2}\right) > 0 \tag{5}$$

where

$$\lambda(y, f, L, H, s) \equiv \int_L^H y\, f(y)\, dy \tag{6}$$

We will see futher how antifragility benefits from Jensen's inequality.

Philostochaticity: Biological, Economic, and Political Systems Starved of Variations

Our definition is based on philostochasticity, love of variations (fragile is stochastophobe). Positive sensitivity to variation**s** is not part of the common vocabulary. Further, the idea of a system can be "starved of variation", i.e. weakens under absence of stressors is absent from the discourse. This is the method used in *Antifragility* (Taleb, manuscript) and applied to both biological , economic, and political systems.

How Concavity of Payoff Leads to Fragility, Convexity to Antifragility

Under monomodal distribution, Left-Concavity of payoff (for a function) $\Longrightarrow$ Left tailedness (in probability space) $\Longrightarrow$ Fragility

(where left-concavity means concavity of payoff in the loss domain, taking payoff as a "derivative", that is a function of a symmetric random variable following a set of typical classes of distributions)

Skewness: For finite moment monomodal distributions, more negative skewness implies more fragility, as V take more negative value and assymmetric distributions increase in asymmetry, positive or negative, with increase in dispersion. (We will see why this only works for monomodal distributions.) Further, the inequality between "raw" shortfalls maps into skewness, but not the reverse, particularly that there is a variety of mathematical measures of skewness, most of which depend on strong sets of assumptions.

The arrow nonlinearity -> skewness (hence fragility) is obvious, in addition there is a proof mentioned in the next section. There is no need for proof that skewness has for sole origin nonlinearity since left-tail is fragility by definition not skewness.

But my interest in skewness is only as a in comparative measurement of tails (bounded one side, unbounded the other).

Purely Concave or Purely Convex Transformations (Types 1 and 4)

For finite moment distributions and functions twice differentiable, for all values of x in the convex case and <0 in the concave one, a convex (concave) function of x is a random variable with positive (negative) odd moments, (again, Van Zwet, 1964).

Note that skewed distributions increase in skewness (positive or negative) from the increase in variance (in the case of finite moments), or their dispersion coefficient.

Example 1: A vanilla (standard) option payoff off any underlying has a positive skewness for the owner (negative for the seller) that increases with the dispersion of the underlying security.

Example 2: Take the concave function g(x)= 1- $e^{-a\,x}$ (typically used with utility models) and x Gaussian(μ,σ) distributed . With a=1, the distribution becomes $v(X) = -\frac{e^{-\frac{(\mu+\log(1-X))^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma\,(X-1)}$

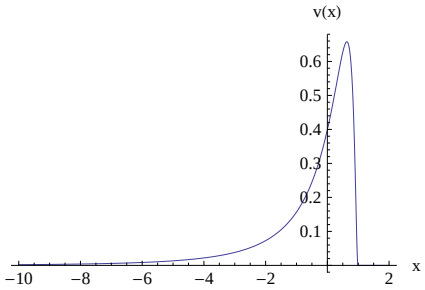


Figure 9- Left tails, pure skewness from concave transformation 1- e^{-x} of a Gaussian variable x .

Example 3: A convex transformation of the same variable x, f= Exp[x] yields the Lognormal distribution with skewness $\sqrt{e^{\sigma^2}-1}\,\left(e^{\sigma^2}+2\right)$ which, clearly, increases with dispersion.

Mixed Transformation I, Concave-Convex (Fat tail, Type II)

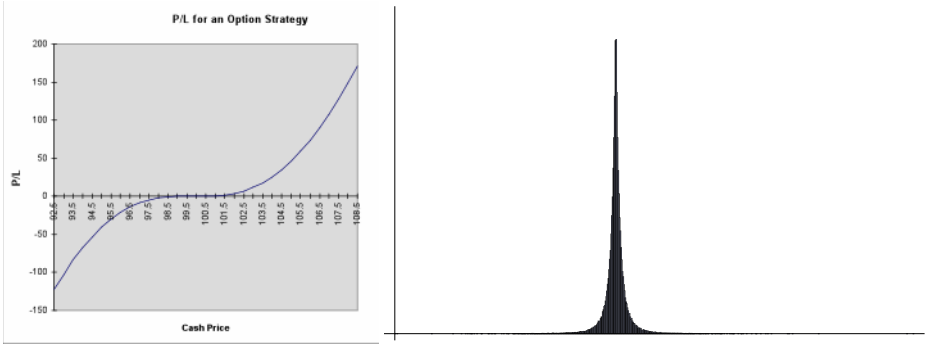


Figure 10 A "Fence" Payoff. From Dynamic Hedging, Taleb(1997), (short series of OTM puts long series OTM call, delta neutral). The shape can be expressed with $f = e^{x}$ if $x > 0$, $2 - e^{-x}$ otherwise.

Figure 11- Distribution of the payoff in Figure 5, fat tails, even with Gaussian underlying

Notes:

- **a.** This is for a monotonically increasing function (hence unbounded on both ends). For a decreasing funtion, convex-concave would have the same effect.
- **b.** In the center, the "fence payoff" has mathematical derivatives first, second, third (with respect of the variable in the horizontal axis) all at 0, but not for a set finite Δp for which they are positive.

Mixed Transformation II, Convex-Concave (Robust, Type III)

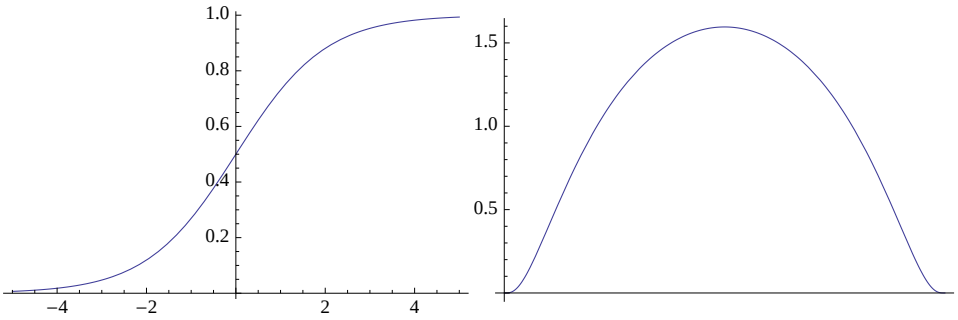


Figure 7- The "digital" payoff, opposite style to the "fence", closer to a 1st or 3rd quadrant payoff (Taleb, 1998), that is, bounded on both sides. Its properties can be captured with the sigmoid function $\frac{1}{1+e^{-a\,x}}$

Figure 8- The distribution of the payoff in Figure 6: thin tails both sides when one starts from the middle

Notes:

- **a.** This is for a monotonically increasing function (hence bounded on both ends). For a decreasing funtion, concave-convex would have the same effect.
- **b.** Note that prospect theory (Kahneman and Tversky) has this shape in the utility function --since happiness has "thin tails", does not experience extremes (it is bounded upwards and downwards), even if the underlying variable is wealth which is unbounded.

Example : Where $u = f(x) = \frac{1}{\exp(-a\,x) + 1}$, x is Gaussian distributed (μ, σ) on the real line, then expressing g the payoff function :

$$g(X) = -\frac{e^{-\frac{\left(a\,\mu+\log\left(\frac{1}{X}-1\right)\right)^2}{2\,a^2\,\sigma^2}}}{\sqrt{2\,\pi}\;a\,\sigma\,(X-1)\,X}, \quad 0 < X < 1$$

Local Antifragility: Note that at some scale, the transformation can bring a thicker right tail than the left one. This is common with biological systems.

The Convex-Concave can produce some degree of antifragility when positioned at the left end and fragility on the right side, both with truncated tails.

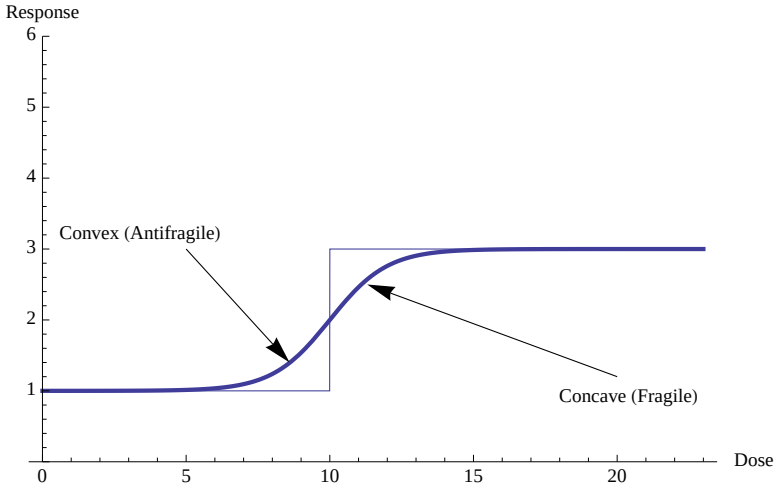


Figure 13- Dose response in biological systems (similar to digital payoffs). The mixture regularizes say, caloric consumption.

Warning on pseudo - convexity

This is the case where asymptotic properties diverge from the visible ones or the locally analytically derived ones. Typically, systems tend to push risks in the tails.

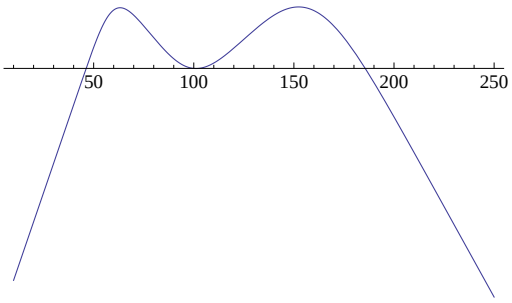


Figure14 : Many risks are hidden under local convexity and tails concavity. From *Dynamic Hedging* (1997). There are simply remedied with slighly larger size of perturbations.

Model Error and Semi-Bias as Nonlinearity from Missed Stochasticity of Variables

Model error, as we saw:

- **a.** E_2 missing the existence of a random variable that is significant in determining the outcome (say option pricing without credit risk). We cannot detect it using the heuristic presented in this paper but as mentioned earlier the error goes in the opposite direction as models tend to be richer, not poorer, from overfitting.
- **b.** E_3missing the stochasticity of a variable or underestimating its stochastic character (say option pricing with nonstochastic interest rates or ignoring that the "volatility" σ can vary), see previous argument.

Missing Effects: The study of model error is not to question whether a model is precise or not, whether or not it tracks reality; it is to ascertain the first and second order effect from missing the variable, insuring that the errors from the model don't have missing higher order terms that cause severe unexpected (and unseen) biases in one direction because of convexity or concavity. In other words, whether or not the model error causes a change in ζ.

Example: Uncertainty and Delays. How many times have you crossed the Atlantic —with a nominal flying time of 7 hours— and arrived 1, 2, 3, or 6 hours late? Or even a couple of days late, perhaps owing to the irritability of some volcano. Now, how many times have you landed 1, 2, 3, 6 hours early? Clearly we can see that in some environments uncertainty has a one way effect: extend expected arrival time. So here missing stochasticity of variables ***lengthens*** arrival time (effect of Jensen's inequality) but it also increases ζ when taken as an economic outcome as a function of arrival time.

Example: Small Probabilities. Another application explains why I spent my life making bets on unlikely events, on grounds of incompleteness of models. Assume someone tells you that the probability of an event is 0. But you don't trust his computation. Because a probability cannot be lower than 0, your expected probability should be higher, at least higher than the expected error rate in the computation of such probability. Model error increases small probabilities in a disproportionate way (and accordingly decreases large probabilities). The effect is only neutral for probabilities in the neighborhood of .5

Convexity Effects and Jensen's Inequality

Define a convex function as one with a positive second derivative, but this is a mathematical construct that does not translate well into practice (as it requires twice-differentiability). Recall from Figure 10 the "flipping" of the exposure from convex in the body of the distribution to severely concave in the tails. So, more practically, convexity over an interval 2 Δx satisfies the following inequality:

$$\frac{1}{2}\left(f(x+\Delta x)+f(x-\Delta x)\right) \geq f(x)$$

Why economics as a discipline made the monstrously consequential mistake of treating estimated parameters as nonstochastic variables and why this leads to fat-tails even while using Gaussian models.

The average of expectations is not the expectation of an average. For f convex across all values of $\{X_i\}$,

$$\sum w_i\, E\, f(X_i) \geq E\Big[\sum f(w_i\, X_i)\Big]$$

For example, take a conventional die (six sides) and consider a payoff equal to the number it lands on. The expected (average) payoff is $\frac{1}{6}\sum_{i=1}^{6} i = 3.5$. Now consider that we get the squared payoff, $\frac{1}{6}\sum_{i=1}^{6} i^2 = \frac{91}{6} \simeq 15.67$, while $\left(\frac{1}{6}\sum_{i=1}^{6} i\right)^2 = 12.25$, so, since squaring is a convex function, the average of a square payoff is higher than the square of the average payoff.

Model Bias and Second Order Effects

Having the right model (which is a very generous assumption), but being uncertain about the parameters will invariably lead to an increase in model error in the presence of convexity and nonlinearities.

As an generalization of the defiict/employment example used in the introduction, say we are using a simple function:

$$f\left(x \,\middle|\, \bar{\alpha}\right) \tag{7}$$

where $\bar{\alpha}$ is supposed to be the average expected rate, where we take ϕ as the distribution of α

$$\bar{\alpha} = \int \alpha\, \phi(\alpha)\, d\alpha \tag{8}$$

The mere fact that α is uncertain (since it is estimated) might lead to a bias if we perturbate from the outside (of the integral), i.e. stochasticize the parameter deemed fixed. Accordingly, the convexity bias is easily measured as the difference between a) f integrated across values of potential α and b) f estimated for a single value of α deemed to be its average. The convexity bias ξ becomes:

$$\xi \equiv \int f(x \mid \alpha)\, \phi(\alpha)\, d\alpha - f\left(x \,\middle|\, \int \alpha\, \phi(\alpha)\, d\alpha\right) \tag{9}$$

Example: A Call Option expanding on the convexity biases of the Bachelier-Thorpe equation:

As an example let us take the Bachelier-Thorp option equation (often called in the literature the Black-Scholes-Merton formula), an equation I spent 90% of my adult life fiddling with. I use it in my class on model error at NYU-Poly as an ideal platform to explain the effect of assuming a parameter is deterministic when in fact it can be stochastic .
A call option (simplifying for absence of interest rate) is the expected payoff:

$$C(S,\ K,\ \bar{\sigma},\ t) = \int_K^\infty (S - K)\, \phi\left(S \mid \mu,\ \bar{\sigma}\ \sqrt{t}\ \right) dS \tag{10}$$

where Φ is the Lognormal distribution, So is the initial asset price, K the strike, $\bar{\sigma}$ the mean expected standard deviation, and t the time to expiration. Only S is stochastic within the formula, all other parameters are considered as descending from some higher deity, or estimated without estimation error.

The easy way to see the bias is by producing a nested distribution for the standard deviation σ, say a Lognormal with standard deviation V then the true option price becomes, from the integration from the outside:

$$\xi \equiv \int_0^\infty \int_K^\infty (S - K)\, \phi(S,\ \ \sigma)\, dS\, d\sigma - \int_K^\infty (S - K)\, \phi\left(S \mid \mu,\ \bar{\sigma}\ \sqrt{t}\ \right) dS \tag{11}$$

(assuming independence between the distribution of S and that of σ, it equals the simpler to calculate

$$\xi = \int_K^\infty C(S,\ K,\ \sigma,\ t)\, f(\sigma)\, d\sigma - C(S,\ K,\ \bar{\sigma},\ t) \tag{12}$$

(assuming independence between the distribution of S and that of σ).

The convexity bias is of course well known by option operators who price out-of-the-money options, the most convex, at some premium to the initial Bachelier-Thorpe model, a relative premium that increases with the convexity of the payoff to variations in σ.

Simplifying, using *a* as a perturbation magnitude for σ about one mean deviation away

$$\xi \simeq \frac{C\,(S,\ K,\ \sigma\,(1 + a),\ t) - C(S,\ K,\ \sigma\,(1 - a),\ t)}{C\,(S,\ K,\ \sigma,\ t)} \tag{13}$$

For options struck 6 σ away from the money, with a= 1/5, the relative bias approches 5000% in option value. Note that a= 1/5 is "mild compared to the variations we see, with a often = 4/5.

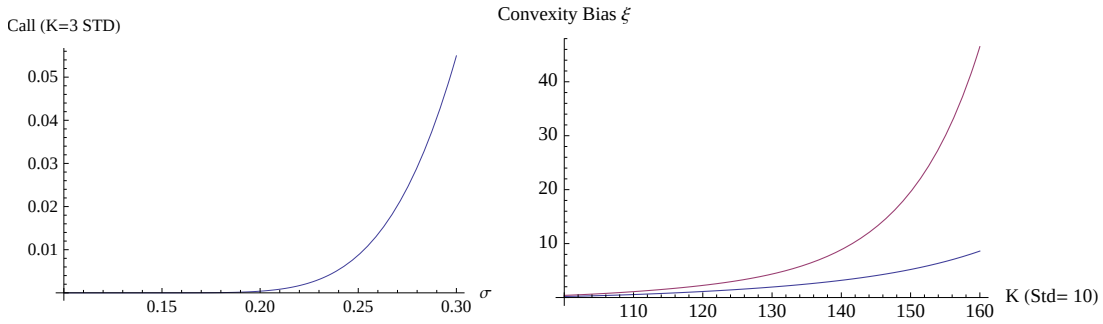


Figure 15 The value of a call as a function of σ; out of the money options are extremely convex to σ.

Figure 16 The convexity bias for options, with K strike prices away from the money, where S is 100 (an increment of 10 is one STD, as K ($\sigma\ \sqrt{t}$) =10). a= $\frac{1}{5}$ and $\frac{1}{10}$

By stocasticizing the possible values of σ assuming a Lognormal distribution we end up with skewed values of C (positive for the holder, negative for the seller)

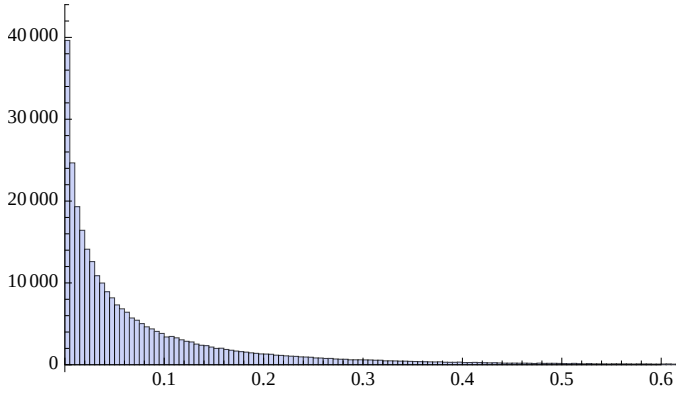


Figure 17 Distribution of the value of the option in previous example across stochastic σ [n=300K simulations]. Much more skewed.

Model Bias and Small Probabilities

Argument 1: Incomputability of small probability. The smaller the more incomputable, hence the most fragile. This comes from the effect of nonlinearity increasing biases in the tails of the distribution.

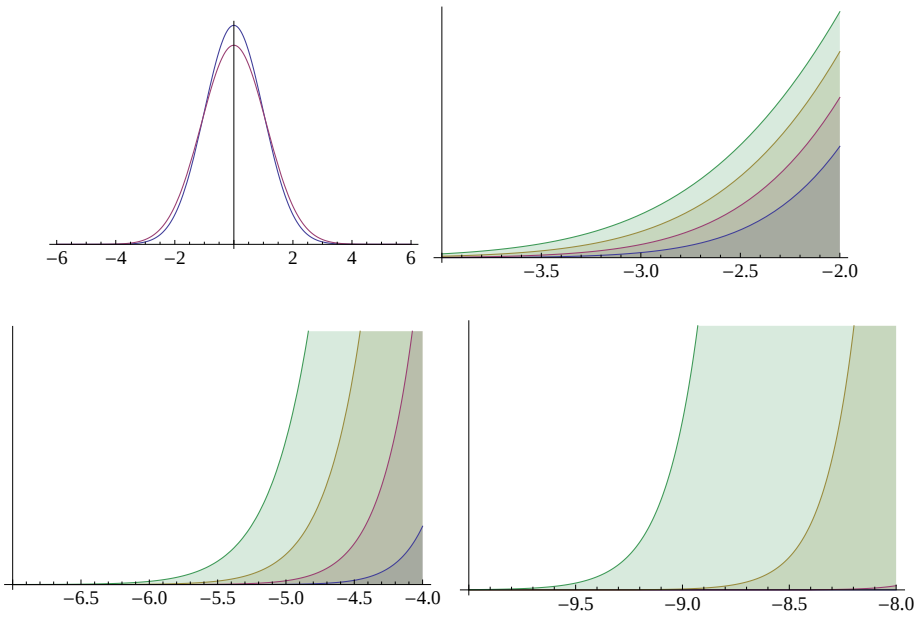
The convexity bias for distribution in wich p is the "determining" parameter and a the magnitude of relative perturbation, P> the excess probability

$$\xi \equiv \frac{P > K \mid p\,(1+a) + P > K \mid p(1-a)}{P > K \mid p} \tag{14}$$

One layer might not be enough: in my paper "The Future Has Thicker Tails than the Past: Model Error As Branching Counterfactuals", I perturbate a with $\frac{1}{n}$ a(1)(1± a(2)) (1± a(3))...(1±a(n)) for additional convexity effects.

Example, assume P is Gaussian,ξ for different values of K (remotenss of event) which can be monstrous in the tails, meaning one needs more and more certainty about σ for calulations involving remote events. (Branching counterfactuals are not even needed here for the argument).

Argument 2: Convexity to Gaussian parameter σ as indicative of wrong probability distribution; stochasticizing it leads to fat tails. Using the Gaussian as a wrong probability distribution is equivalent (for monomodal distributions) to missing the stochastic character and nonlinear effect of a parameter in the Gaussian (say the error in using a Gaussian in place of power-law arises from adding a nested random character to the standard deviation σ, etc.)



Figures 18 How successive increases of STD by 10% fill-in deeper and deeper into the tails, with thicker and thicker areas under the curves in the tails. The vertical axis represents probability, the horizontal one represents standard deviations.

Other Applications

Corporate Finance: In short, corporate finance seems to be based on point projections, not distributional projections; thus if one perturbates cash flow projections, say, in the Gordon valuation model, replacing the fixed —and known— growth by continuously varying jumps (particularly under fat tails distributions), companies deemed "expensive", or those with high growth, but low earnings, would markedly increase in expected value, something the market prices heuristically but without explicit reason.

Portfolio Theory: The first defect of portfolio theory and every single theory based on "optimization" is absence of uncertainty about the source of parameters --while these theorists leave it to the econometricians to ferret out the data, not realizing the inconsistency that an unknown parameter has a stochastic character. Of course the second defect is the use of thin-tailed idealized probability distributions.

We will revisit the method after the derivation of a heuristic and apply it to various: deficits, traffic delays, company size, etc.

The Fragility-Bias Detection Heuristic

Example 1 (Detecting Tail Risk). *A bank issues a so-called "stress test" (something that has never worked in history), off a parameter (say stock market) at -15%. We ask them to recompute at -10% and -20%. Should the exposure show negative asymmetry (worse at -20% than it improves at -10%), we deem that their risk increases in the tails. There are certainly hidden tail exposures and a definite higher probability of blowup in addition to exposure to model error.*

Note that it is somewhat more effective to use our measure of shortfall in Definition 1a, but the method here is effective enough to show hidden risks, particularly at wider increases (try 25% and 30% and see if exposure shows increase). Most effective would be to use powerlaw distributions and perturbate exponent.

Example 2 (Detecting Tail Risk in Overoptimized System). *Raise airport traffic 10%, lower 10%, take average expected traveling time from each, and check the asymmetry for nonlinearity. If asymmetry is significant, then declare the system as overoptimized.*

Example 3.(Detecting Model Bias). *A government computes government budget (deficit) off an unemployement forecast of 8%, as a point estimate. We ask the governement to recompute expected deficit using the exact same method, but with unemployment at 7% and 9%, and check if there is an unfavorable asymmetry (deficit improves less at +1% increase in employment than it worsens at 1% decrease). If there is unfavorable asymmetry, deficit estimate is likely to be underestimated according to Jensen's Inequality. It is maily deemed fragile to estimation error.*

Note again that the method works even if the government has a wrong model.

Example 4. *A corporation laden with debt issues point-estimated of profit forecasts using a collection of parameters, from energy costs to demand. Perturbate the main factors and see the odds of the company going bust.*

The same procedure uncovers both fragility and consequence of model error (potential harm from having wrong probability distribution, a thin-tailed rather than a fat-tailed one). As a trader (and see Gigerenzer's discussions, Gigerenzer and Brighton (2009), Gigerenzer and Goldstein

(1996)) playing with second order effects of simplistic tools can be more effective than more complicated and harder to calibrate methods. See also the intuition of fast and frugal in Derman and Wilmott (2009), Haug and Taleb (2011).

***The Heuristic (Application to Model Error Detection)**:*

Simply, for function f(x) calculated at x_0 the calculate model error over finite range Δp *discretely compute the convexity bias ξ, that is* $f(x|\bar{p}) - \frac{1}{2}\{f(x|p+\frac{\Delta p}{2}) + f(x|p-\frac{\Delta p}{2})\}$ *for every parameter that needs to be estimated or can be subjected to measurement error. Any significant difference implies model error. Model bias can be even calculated when* $\frac{\Delta p}{2}$ *approximates the mean absolute error for parameter* p.

***The Heuristic (Application to Risk Detection)**:*

1- First Step (first order). Measure the sensitivity to all parameters p determining V over finite ranges Δp. If materially significant, check if stochasticity of parameter is taken into account by risk assessment. If not, then stop and declare the risk as grossly mismeasured (no need for further risk assessment). (Note that Ricardo's wine-cloth example example miserable fails the first step upon stochasticizing either).

2-Second Step (second order). For all parameters p compute the second order $H(\Delta p) \equiv \frac{V'}{V}$, *where* $V'(\Delta p) \equiv \frac{1}{2}\left(V\left(p+\frac{1}{2}\Delta p\right) + V\left(p-\frac{1}{2}\Delta p\right)\right)$.

3- Third Step. Note parameters for which H is significantly > or < 1

***Properties of the Heuristic**:*

*i- **Fragility**: V' is a more accurate indicator of fragility than V over Δp when p is stochastic or subjected to estimation errors with mean deviation Δp*

*ii- **Model Error**: A model M(p) with parameter p held constant underestimates the fragility of payoff from x under perturbation Δp if H>1,*

iii- if H=1, the exposure to x is robust over Δp and model error over p is inconsequential.

iv- if H<1, the exposure to x is antifragile over Δp (since antifragility is immediately obtained from bounding the left payoff while keeping the right one open)

v- if H remains ≥ 1 for larger and larger Δp, then the heuristic is broad (absence of pseudoconvexities)

We can apply the method to V in Equation 1, as it becomes a perturbation of a perturbation, (in *Dynamic Hedging* "vvol", or "volatility of volatility" or in later lingo vvol), H= $\frac{V(x,f,K,\Delta s+1/2\Delta s)+V(x,f,K,\Delta s-1/2\Delta s)}{2V(x,f,K,\Delta s)}$ where K is the fragility threshold, x is a random variable describing outcomes, Δp is a set perturbation and f the probability measure used to compute ζ.

Note that for K set at ∞, the heuristic becomes a simple detection of model bias from the effect of Jensen's inequality when stochasticizing a term held to be deterministic.

The heuristic has the ability to "fill-in the tail", by extending further down into the probability distribution as Δp increases. It is best to perturbate the tail exponent of a power law.

Remarks:

- **a.** Simple heuristics have a robustness (in spite of a possible bias) compared to optimized and calibrated measures. Ironically, it is from the multiplication of convexity biases and the potential errors from missing them (i.e., again, Jensen's Inequality) that calibrated models that work in-sample underperform heuristics out of sample.
- **b.** It is not necessary to have the right probability distribution for the heuristic to be accurate, since we are measuring second order effects and potential tail exposure. Even wrong distributions, wrong methods show the right tail exposure through perturbation. This is where most of the strength lies.
- **c.** It allows to detection of the effect of the use of the wrong probability distribution without changing probability distribution (just from the dependence on parameters).
- **d.** It outperforms all other commonly used measures of risk, such as cVar, "expected shortfall", stress-testing, and similar methods have been proven to be completely ineffective.
- **e.** It does not require parametrization beyond varying Δp.

Examples:

- **i.** It detects fragility to forecasting errors in projection as these reside in convexity of duration/costs to uncertainty
- **ii.** Bank portfolios
- **iii.** Example: hidden tail exposures to budget deficits' nonlinearities to unemployement
- **iv.** Example: hidden tail exposure from dependence on a source of energy, etc. ("squeezability argument")

Comparison of the Heuristic to Other Methods

CVar & VaR: these are totally ineffective, no need for further discussion here (or elsewhere) as they have been shown to be so empirically and mathematically.

Stress Testing. The author has shown where these can be as ineffective owing to risk hiding in the tail below the stress test. See Taleb (2009) on why the level K of the stress test is arbitrary and cannot be appropriately revealed by the past realizations of the random variable. But if stress tests show an increase in risk at lower and lower levels of stress, then the position reveals exposure in the tails.

Note that hidden risks reside in the tails as they are easy to hide there, undetected by conventional methods and tend to hide there.

Detection of How Optimization Leads to Hidden Fragility, Future Works

In parallel works, applying the ***"simple heuristic"*** allows us to detect the following "hidden short options" problems by merely perturbating a certain parameter *p*:

- **a.** Size and peudoeconomies of scale.
  - **i.** size and squeezability (nonlinearities of squeezes in costs per unit)
- **b.** Specialization (Ricardo) and variants of globalization.
  - **i.** missing stochasticity of variables (price of wine).
  - **ii.** specialization and nature.
- **c.** Portfolio optimization (Markowitz)
- **d.** Debt
- **e.** Budget Deficits: convexity effects explain why uncertainty lengthens, doesn't shorten expected deficits.
- **f.** Iatrogenics (medical) or how some treatments are concave to benefits, convex to errors.
- **g.** Disturbing natural systems

Acknowledgments

Bruno Dupire, Raphael Douady.

References

Derman, E. and Wilmott, P. (2009). The Financial Modelers' Manifesto, SSRN: http://ssrn.com/abstract=1324878

Gigerenzer, G. and Brighton, H.(2009). Homo heuristicus: Why biased minds make better inferences, *Topics in Cognitive Science*, 1-1, 107-143

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 46(2):171–185.

Jensen, J. L. W. V. (1906). "Sur les fonctions convexes et les inégalités entre les valeurs moyennes". *Acta Mathematica* 30

Haug, E. & Taleb, N.N. (2011) Option Traders Use (very) Sophisticated Heuristics, Never the Black–Scholes–Merton Formula *Journal of Economic Behavior and Organization*, Vol. 77, No. 2,

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, R. Parzen, and R. Winkler (1982). "The Accuracy of Extrapolation (Time Series) Methods: Results of a Forecasting Competition." *Journal of Forecasting* 1: 111–153.

Makridakis, S., and M. Hibon (2000). "The M3-Competition: Results, Conclusions and Implications." *International Journal of Forecasting* 16: 451–476.

Taleb, N.N. (1997). *Dynamic Hedging: Managing Vanilla and Exotic Options*, Wiley

Taleb, N.N. (2009). Errors, robustness and the fourth quadrant, *International Journal of Forecasting*, 25-4, 744--759

Taleb, N.N. (2011). *Antifragility*, Manuscript

W.R. Van Zwet (1964). *Convex Transformations of Random Variables*, Mathematical Center Amsterdam, 7

p

The Future Has Thicker Tails than the Past: Model Error As Branching Counterfactuals

Nassim N Taleb
NYU-Poly Institute

May 31 ,2011, 3rd version

PRESENTED IN HONOR OF BENOIT MANDELBROT'S AT HIS SCIENTIFIC MEMORIAL

Yale University, APRIL 29, 2011

Abstract

Ex ante forecast outcomes should be interpreted as counterfactuals (potential histories), with errors as the spread between outcomes. We reapply measurements of uncertainty about the estimation errors of the estimation errors of an estimation treated as branching counterfactuals. Such recursions of epistemic uncertainty have markedly different distributional properties from conventional sampling error, and lead to fatter tails in the projections than in past realizations. Counterfactuals of error rates always lead to fat tails, regardless of the probability distribution used. A mere .01% branching error rate about the STD (itself an error rate), and .01% branching error rate about that error rate, etc. (recursing all the way) results in explosive (and infinite) moments higher than 1. Missing any degree of regress leads to the underestimation of small probabilities and concave payoffs (a standard example of which is Fukushima). The paper states the conditions under which higher order rates of uncertainty (expressed in spreads of counterfactuals) alters the shapes the of final distribution and shows which *a priori* beliefs about conterfactuals are needed to accept the reliability of conventional probabilistic methods (thin tails or mildly fat tails).

KEYWORDS: Fukushima, Counterfactual histories, Risk management, Epistemology of probability, Model errors, Fragility and Antifragility, Fourth Quadrant

Introduction

Intuition. An event has never shown in past samples; we are told that it was "estimated" as having zero probability. But an estimation has to have an error rate; only measures deemed *a priori* or fallen from the sky and dictated by some infaillible deity can escape such error. Since probabilities cannot be negative, the estimation error will necessarily put a lower bound on it and make the probability > 0.

This, in a nutshell, is how we should treat the convexity bias stemming from uncertainty about small probabilities. Using the same reasoning, we need to increase the raw "estimation" of small probabilities by a margin for the purposes of future , or out-of-sample projections. There can be uncertainty about the relationship between past samples and future ones, or, more philosophically, from problems attending inductive inference. Doubting the reliability of the methods used to produce these probabilities, the stability of the generating process, or beliefs about the future resembling the past will lead us to envision a spate of different alternative future outcomes. The small probability event will necessarily have, in expectation, i.e., on average across all potential future histories, a higher than what was measured on a single set. The increase in the probability will be commensurate with the error rate in the estimation. It, simply, results from the convexity bias that makes small probabilities rise when we are uncertain about them. Accordingly, the future needs to be dealt with as having thicker tails (and higher peaks) than what was measured in the past.

Incoherence in Probabilistic Measurements. Just as "estimating" an event to be of measure 0 is lacking in coherence, it is equally inconsistent to estimate anything without introducing an estimation error in the analysis and adding a convexity bias (positive or negative). But this incoherence (or confusion between estimated and *a priori*) pervades the economics literature whenever probabilistic and statistical methods are used. For instance, the highest use of probabiltiy in modern financial economics is in portfolio theories resulting from the seminal Markowitz (1952), which has the derivations starting with assuming E and V (expectation and variance) for certain securities. At the end of the paper the author states that these parameters need to be estimated. Injecting an estimation error in the analysis would entirely cancel the derivations of the paper as they are based on immutable certainties (which explains why the results of Markowitz (1952) have proved unusable in practice).

Regressing Counterfactuals

We can go beyond probabilities and perturbate parameters of probability distributions used in practice and, further, perturbate the rates of perturbation. There is no reason to stop except where there are certainties lest we fall in a Markowitz-style incoherence. So this paper introduces two notions: treating errors as branching counterfactual histories and regressing (i.e., compounding) the error rates. By counterfactual I mean in the Ferguson (1997) and Parker and Tetlock (2006) sense of alternative historical events. An error rate about a forecast can be estimated (or, of course, "guessed"). The estimation (or "guess"), in turn, will have an error rate. The estimation of such error rate will have an error rate. (The forecast can be an economic variable, the future rainfall in Brazil, or the damage from a nuclear accident).

What is called a regress argument by philosophers can be used to put some scrutiny on quantitative methods or risk and probability. The mere existence of such regress argument will lead to series of branching counterfactuals three different regimes, two of which lead to the necessity to raise the values of small probabilities, and one of them to the necessity to use power law distributions. This study of the structures of the error rates refines the analysis of the *Fourth Quadrant* (Taleb, 2008) setting the limit of the possibility of the use of probabilistic methods (and their reliability in the decision-making), based on errors in the tails of the distribution.

So the boundary between the regimes is what this paper is about -what assumptions one needs to have set beforehand to avoid radical skepticism and which specific *a priori* undefeasable beliefs are necessary to hold for that. In other words someone using probabilistic estimates should tell us beforehand which immutable certainties are built into his representation, and what should be subjected to error -and regress --otherwise they risk falling into a certain form of incoherence: if a parameter is estimated, second order effects need to be considered.

This paper can also help setting a wedge between forecasting and statistical estimation.

The Regress Argument (Error about Error)

The main problem in probabilistic risk measument is the limited understanding of model (or representation) error, and, for those who get it, a lack of understanding of second order errors (about the methods used to compute the errors) and by a regress argument, an inability to continuously reapplying the thinking all the way to its limit (***particularly when they provide no reason to stop***). Again, there is no problem in stopping the recursion, provided it is accepted as a declared *a priori* that escapes quantitative and statistical methods.

Regress Arguments: Probability professionals quantitative risk professionals ("quants") do not include in the probabilistic measurement itself an error rate about, say, the estimation of a parameter provided by an expert, or other uncertainties attending the computations. This would only be acceptable if they consciously accepted such limit. Just reapplying layers of uncertainties may show convexity biases, and, fortunately, it does not necessarily kill probability theory; it just disciplines the use of some distributions, at the expense of others --distributions in the $\mathcal{L}^2$ norm (i.e., square integrable) may no longer be valid, for epistemic reasons. *Without understanding errors, a measure is nothing* and one should take the point to its logical consequence that *any measure of error needs to have its own error taken into account*.

The epistemic and counterfactual aspect of standard deviations: The standard deviation of a distribution ***for future outcomes*** (and not the sampling of some properties of existing population), the measure of dispersion, needs to be interpreted as the measure of uncertainty, distance between counterfactuals, hence *epistemic*, and that it, in turn, should necessarily have uncertainties (errors) attached to it (unless he accepted infallibility of belief in such measure). One needs to look at the standard deviation -or other measures of dispersion -as a degree of ignorance about the future realizations of the process. The higher the uncertainty, the higher the measure of dispersion (variance, mean deviation, etc.)

Such uncertainty, by Jensen's inequality, creates non-negligible convexity biases. So far this is well known in places in which subordinated processes have been used --for instance stochastic variance models --but I have not seen the layering of uncertainties taken into account.

Note: Counterfactuals, Estimation of the Future v/s Sampling Problem

Note that it is hard to escape higher order uncertainties, even outside of the use of counterfactual: even when sampling from a conventional population, an error rate can come from the production of information (such as: is the information about the sample size correct? is the information correct and reliable?), etc. These higher order errors exist and could be severe in the event of convexity to parameters, but they are qualitatively different with forecasts concerning events that have not taken place yet.

This discussion is about an epistemic situation that is markedly different from a sampling problem as treated conventionally by the statistical community, particularly the Bayesian one. In the classical case of sampling by Gosset ("Student", 1908) from a normal distribution with an unknown variance (Fisher, 1925), the Student T Distribution (itself a power law) arises for the estimated mean since the square of the variations (deemed Gaussian) will be Chi-square distributed. The initial situation is one of completely unknown variance, but that is progressively discovered through sampling; and the degrees of freedom (from an increase in sample size) rapidly shrink the tails involved in the underlying distribution.

The case here is the exact opposite, as we have an a priori approach with no data: *we start with a known priorly estimated or "guessed" standard deviation, but with an unknown error on it expressed as a spread of branching outcomes*, and, given the a priori aspect of the exercise, we have no sample increase helping us to add to the information and shrink the tails. We just deal with nested counterfactuals.

Note that given that, unlike the Gosset's situation, we have a finite mean (since we don't hold it to be stochastic and know it a priori) hence we necessarily end in a situation of finite first moment (hence escape the Cauchy distribution), but, as we will see, a more

complicated second moment.

See the discussion of the Gosset and Fisher approach in Chapter 1 of Mosteller and Tukey (1977). [I thank Andrew Gelman and Aaron Brown for the discussion].

Main Results

Note that unless one stops the branching at an early stage, all the results raise small probabilities (in relation to their remoteness; the more remote the event, the worse the relative effect).

1. Under the regime of proportional constant (or increasing) recursive layers of uncertainty about rates of uncertainty, the distribution has infinite variance, even when one starts with a standard Gaussian.
2. Under the other regime, where the errors are decreasing (proportionally) for higher order errors, the ending distribution becomes fat-tailed but in a benign way as it retains its finite variance attribute (as well as all higher moments), allowing convergence to Gaussian under Central Limit.
3. We manage to set a boundary between these two regimes.
4. In both regimes the use of a thin-tailed distribution is not warranted unless higher order errors can be completely eliminated *a priori*.

Epistemic not statistical re-derivation of power laws: Note that previous derivations of power laws have been statistical (cumulative advantage, preferential attachment, winner-take-all effects, criticality), and the properties derived by Yule, Mandelbrot, Zipf, Simon, Bak, and others result from structural conditions or breaking the independence assumptions in the sums of random variables allowing for the application of the central limit theorem. This work is entirely epistemic, based on standard philosophical doubts and regress arguments.

Methods and Derivations

Layering Uncertainties

The idea is to hunt for convexity effects from the layering of higher order uncertainties (Taleb, 1997).

Take a standard probability distribution, say the Gaussian. The measure of dispersion, here σ, is estimated, and we need to attach some measure of dispersion around it. The uncertainty about the rate of uncertainty, so to speak, or higher order parameter, similar to what called the "volatility of volatility" in the lingo of option operators (see Taleb, 1997, Derman, 1994, Dupire, 1994, Hull and White, 1997) --here it would be "uncertainty rate about the uncertainty rate". And there is no reason to stop there: we can keep nesting these uncertainties into higher orders, with the uncertainty rate of the uncertainty rate of the uncertainty rate, and so forth. There is no reason to have certainty anywhere in the process.

Higher order integrals in the Standard Gaussian Case

We start with the case of a Gaussian and focus the uncertainty on the assumed standard deviation. Define $\phi(\mu,\sigma,x)$ as the Gaussian density function for value x with mean μ and standard deviation σ.

A 2^{nd} order stochastic standard deviation is the integral of ϕ across values of $\sigma \in]0,\infty[$, under the measure $f(\overline{\sigma}, \sigma_1, \sigma)$, with σ_1 its scale parameter (our approach to trach the error of the error), not necessarily its standard deviation; the expected value of σ_1 is $\overline{\sigma_1}$.

$$f(x)_1 = \int_0^\infty \phi(\mu, \sigma, x)\, f(\overline{\sigma}, \sigma_1, \sigma)\, d\sigma \tag{1}$$

Generalizing to the N^{th} order, the density function *f(x)* becomes

$$f(x)_N = \int_0^\infty \dots \int_0^\infty \phi(\mu, \sigma, x)\, f(\overline{\sigma}, \sigma_1, \sigma)\; f(\overline{\sigma_1}, \sigma_2, \sigma_1)\; \dots f(\overline{\sigma_{N-1}}, \sigma_N, \sigma_{N-1})\, d\sigma\, d\sigma_1\, d\sigma_2 \dots d\sigma_N \tag{2}$$

The problem is that this approach is parameter-heavy and requires the specifications of the subordinated distributions (in finance, the lognormal has been traditionally used for σ^2 (or Gaussian for the ratio $\text{Log}[\frac{\sigma_t^2}{\sigma^2}]$ since the direct use of a Gaussian allows for negative values). We would need to specify a measure f for each layer of error rate. Instead this can be approximated by using the mean deviation for σ, as we will see next.

Note that branching variance does not always result in higher Kurtosis (4th moment) compared to the Gaussian; in the case of N-2, using the Gaussian and stochasticzing both μ and σ will lead to bimodality the lowering of the 4th moment.

Discretization using nested series of two-states for σ- a simple multiplicative process

A quite effective simplification to capture the convexity, the ratio of (or difference between) $\phi(\mu,\sigma,x)$ and $\int_0^\infty \phi(\mu, \sigma, x)\, f(\overline{\sigma}, \sigma_1, \sigma)\, d\sigma$ (the first order standard deviation) would be to use a weighted average of values of σ, say, for a simple case of one-order stochastic volatility:

$$\sigma\,(1\,\pm\,a\,(1)),\;\;0\leq\,a\,(1)\,<1$$

where $a(1)$ *is* the proportional mean absolute deviation for σ, in other word the measure of the absolute error rate for σ. We use $\frac{1}{2}$ as the probability of each state.

Thus the distribution using the first order stochastic standard deviation can be expressed as:

$$f(x)_1 = \frac{1}{2}\,\{\phi\,(\mu,\;\sigma\,(1\,+\,a(1)),\;x)\,+\,\phi\,(\mu,\,\sigma(1-a(1)),\,x)\} \tag{3}$$

Illustration of the Convexity Effect: Figure 1 shows the convexity effect of a(1) for a probability of exceeding the deviation of x=6. With a[1]=$\frac{1}{5}$, we can see the effect of multiplying the probability by 7.

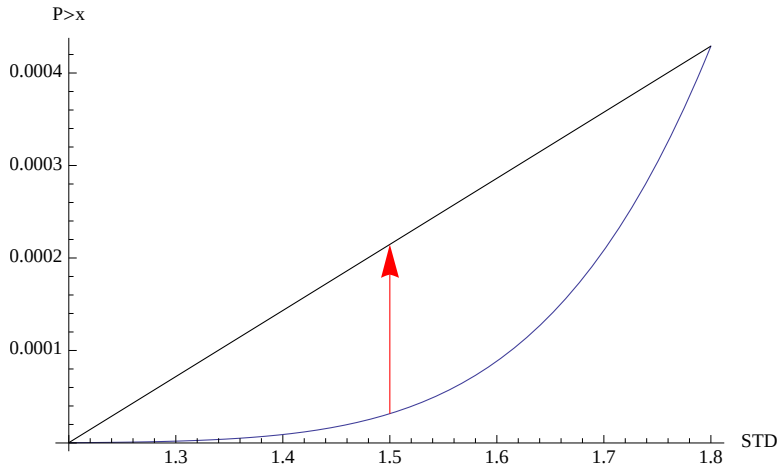


Figure 1 Illustration of he convexity bias for a Gaussian raising small probabilities: The plot shows the STD effect on P>x, and compares P>6 with a STD of 1.5 compared to P> 6 assuming a linear combination of 1.2 and 1.8 (here a(1)=1/5).

Now assume uncertainty about the error rate a(1), expressed by a(2), in the same manner as before. Thus in place of a(1) we have $\frac{1}{2}$ a(1)(1± a(2)).

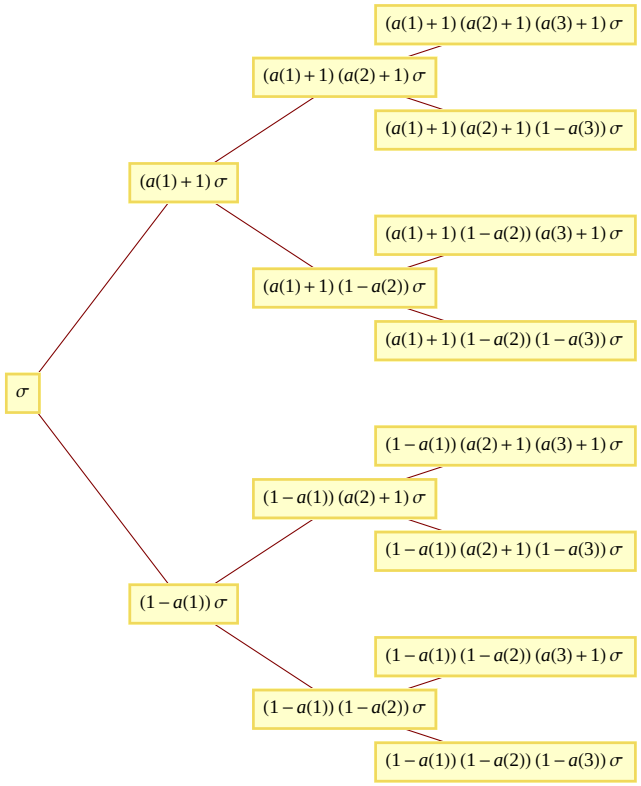


Figure 2- Three levels of error rates for σ following a multiplicative process

The second order stochastic standard deviation:

$$f(x)_2 = \frac{1}{4}\{\phi(\mu,\ \sigma(1+a(1)(1+a(2))),\ x) + \phi(\mu,\ \sigma(1-a(1)(1+a(2))),\ x) + \phi(\mu,\ \sigma(1+a(1)(1-a(2))),\ x) + \phi(\mu,\ \sigma(1-a(1)(1-a(2))),\ x)\} \tag{4}$$

and the $\mathtt{N}^{\mathtt{th}}$ order:

$$f(x)_N = \frac{1}{2^N}\sum_{i=1}^{2^N}\phi\left(\mu,\ \sigma M_i^N,\ x\right) \tag{5}$$

where M_i^N is the $\mathtt{i}^{\mathtt{th}}$ scalar (line) of the matrix $\mathtt{M}^{\mathtt{N}}$ $\left(2^{\mathtt{N}}\times 1\right)$

$$\mathtt{M}^{\mathtt{N}} = \left\{\prod_{\mathtt{j}=1}^{\mathtt{N}}\left(\mathtt{a}\left(\mathtt{j}\right)\mathtt{T}[\![\mathtt{i},\ \mathtt{j}]\!]+1\right)\right\}_{\mathtt{i}=1}^{2^{\mathtt{N}}} \tag{6}$$

and T[[i,j]] the element of $\mathtt{i}^{\mathtt{th}}$line and $\mathtt{j}^{\mathtt{th}}$column of the matrix of the exhaustive combination of N-Tuples of (-1,1),that is the N-dimentionalvector {1,1,1,...} representing all combinations of 1 and -1.

for N=3

$$T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{pmatrix} \quad \text{and } M^3 = \begin{pmatrix} (1-a(1))\,(1-a(2))\,(1-a(3)) \\ (1-a(1))\,(1-a(2))\,(a(3)+1) \\ (1-a(1))\,(a(2)+1)\,(1-a(3)) \\ (1-a(1))\,(a(2)+1)\,(a(3)+1) \\ (a(1)+1)\,(1-a(2))\,(1-a(3)) \\ (a(1)+1)\,(1-a(2))\,(a(3)+1) \\ (a(1)+1)\,(a(2)+1)\,(1-a(3)) \\ (a(1)+1)\,(a(2)+1)\,(a(3)+1) \end{pmatrix}$$

so $M_1^3 = \{(1-a(1))\,(1-a(2))\,(1-a(3))\}$, etc.

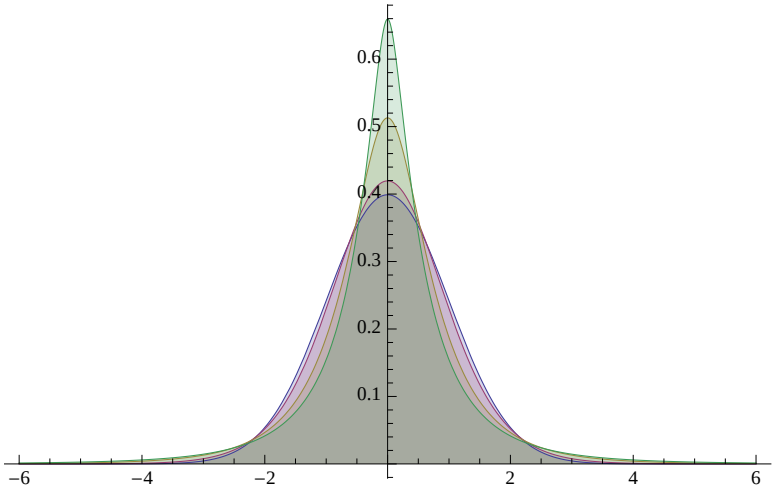


Figure 3, Thicker tails (higher peaks) for higher values of N; here N=0,5,10,25,50, all values of a=$\frac{1}{10}$

A remark seems necessary at this point: the various error rates $a(i)$ are not similar to sampling errors, but rather projection of error rates into the future.

Note: we are assuming here, that σ is stochastic with steps $(1\pm a(n))$, not σ^2. An alternative method would be the mixture with a "low" variance $(\sigma\,(1-v))^2$ and a "high" one $\left(\sigma\sqrt{-v^2+2\,v+1}\right)^2$ selecting a single v so that σ^2 remains the same in expectation. With $1 > v \geq 0$, the total standard deviation.

The Final Mixture Distribution

The mixture weighted average distribution (recall that ϕ is the ordinary Gaussian with mean μ, std σ and the random variable x).

$$f\,(x \mid \mu,\ \sigma,\ M,\ N) = 2^{-N} \sum_{i=1}^{2^N} \phi\left(\mu,\ \sigma\,M_i^N,\ x\right) \tag{7}$$

Regime 1 (Explosive): Case of a Constant parameter *a*

Special case of constant *a*: Assume that a(1)=a(2)=...a(N)=a, i.e. the case of flat proportional error rate *a*. The Matrix *M* collapses into a conventional binomial tree for the dispersion at the level *N*.

$$f\,(x \mid \mu,\ \sigma,\ M,\ N) = 2^{-N} \sum_{j=0}^{N} \binom{N}{j} \phi\left(\mu,\ \sigma\,(a+1)^j\,(1-a)^{N-j},\ x\right) \tag{8}$$

Because of the linearity of the sums, when a is constant, we can use the binomial distribution as weights for the moments (note again the artificial effect of constraining the first moment μ in the analysis to a set, certain, and known *a priori*).

$$\begin{array}{c}\text{Moment}\\ \mu\\ \sigma^2 (a^2+1)^N + \mu^2\\ 3\,\mu\,\sigma^2 (a^2+1)^N + \mu^3\\ 6\,\mu^2\,\sigma^2 (a^2+1)^N + \mu^4 + 3\,(a^4+6\,a^2+1)^N\,\sigma^4\\ 10\,\mu^3\,\sigma^2 (a^2+1)^N + \mu^5 + 15\,(a^4+6\,a^2+1)^N\,\mu\,\sigma^4\\ 15\,\mu^4\,\sigma^2 (a^2+1)^N + \mu^6 + 15\,((a^2+1)\,(a^4+14\,a^2+1))^N\,\sigma^6 + 45\,(a^4+6\,a^2+1)^N\,\mu^2\,\sigma^4\\ 21\,\mu^5\,\sigma^2 (a^2+1)^N + \mu^7 + 105\,((a^2+1)\,(a^4+14\,a^2+1))^N\,\mu\,\sigma^6 + 105\,(a^4+6\,a^2+1)^N\,\mu^3\,\sigma^4\\ \iota^6\,\sigma^2 (a^2+1)^N + \mu^8 + 105\,(a^8+28\,a^6+70\,a^4+28\,a^2+1)^N\,\sigma^8 + 420\,((a^2+1)\,(a^4+14\,a^2+1))^N\,\mu^2\,\sigma^6 + 210\,(a^4+6\,a^2+\end{array}$$

For clarity, we simplify the table of moments, with μ=0

Order	Moment
1	0
2	$(a^2+1)^N\,\sigma^2$
3	0
4	$3\,(a^4+6\,a^2+1)^N\,\sigma^4$
5	0
6	$15\,(a^6+15\,a^4+15\,a^2+1)^N\,\sigma^6$
7	0
8	$105\,(a^8+28\,a^6+70\,a^4+28\,a^2+1)^N\,\sigma^8$

Note again the oddity that in spite of the explosive nature of higher moments, the expectation of the absolute value of x is both independent of *a* and *N*, since the perturbations of σ do not affect the first absolute moment $\int |x|\,f(x)\,dx = \sqrt{\frac{2}{\pi}}\,\sigma$ (that is, the initial assumed σ). The situation would be different under addition of *x*.

Every recursion multiplies the variance of the process by $(1+a^2)$. The process is similar to a stochastic volatility model, with the standard deviation (not the variance) following a lognormal distribution, the volatility of which grows with M, hence will reach infinite variance at the limit.

Consequences

For a constant $a > 0$, and in the more general case with variable a where a(n) ≥ a(n-1), the moments explode.

A- Even the smallest value of $a > 0$, since $(1+a^2)^N$ is unbounded, leads to the second moment going to infinity (though not the first) when N→∞. So something as small as a .001% error rate will still lead to explosion of moments and invalidation of the use of the class of $\mathcal{L}^2$ distributions.

B- In these conditions, we need to use power laws for epistemic reasons, or, at least, distributions outside the $\mathcal{L}^2$ norm, regardless of observations of past data.

Note that we need an *a priori* reason (in the philosophical sense) to cutoff the N somewhere, hence bound the expansion of the second moment.

Convergence to Properties Similar to Power Laws

We can see on the example next Log-Log plot (Figure 1) how, at higher orders of stochastic volatility, with equally proportional stochastic coefficient, (where a(1)=a(2)=...=a(N)= $\frac{1}{10}$) how the density approaches that of a power law (just like the Lognormal distribution at higher variance), as shown in flatter density on the LogLog plot. The probabilities keep rising in the tails as we add layers of uncertainty until they seem to reach the boundary of the power law, while ironically the first moment remains invariant -- because of uncertainty only addressed it.

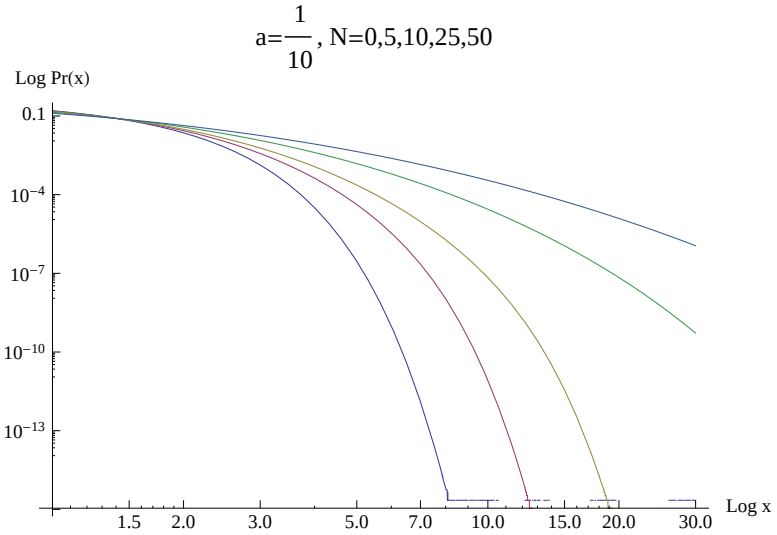


Figure x - LogLog Plot of the probability of exceeding x showing power law-style flattening as N rises. Here all values of a= 1/10

The same effect takes place as a increases towards 1, as at the limit the tail exponent P>x approaches 1 but remains >1.

Effect on Small Probabilities

Next we measure the effect on the thickness of the tails. The obvious effect is the rise of small probabilities.

Take the exceedant probability, that is, the probability of exceeding K, given N, for parameter a constant :

$$P > K \mid N = \sum_{j=0}^{N} 2^{-N-1} \binom{N}{j} \operatorname{erfc}\left(\frac{K}{\sqrt{2}\ \sigma\ (a+1)^{j}\ (1-a)^{N-j}} \right) \tag{9}$$

where erfc(.) is the complementary of the error function, 1-erf(.), $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$

Convexity effect:The next Table shows the ratio of exceedant probability under different values of N divided by the probability assuming a standard Gaussian.

$a = \frac{1}{100}$

N	$\frac{P>3,N}{P>3,N=0}$	$\frac{P>5,N}{P>5,N=0}$	$\frac{P>10,N}{P>10,N=0}$
5	1.01724	1.155	7
10	1.0345	1.326	45
15	1.05178	1.514	221
20	1.06908	1.720	922
25	1.0864	1.943	3347

$a = \frac{1}{10}$

N	$\frac{P>3,N}{P>3,N=0}$	$\frac{P>5,N}{P>5,N=0}$	$\frac{P>10,N}{P>10,N=0}$
5	2.74	146	1.09×10^{12}

10	4.43	805	8.99×10^{15}
15	5.98	1980	2.21×10^{17}
20	7.38	3529	1.20×10^{18}
25	8.64	5321	3.62×10^{18}

Regime 2: Cases of decaying parameters *a(n)*

As we said, we may have (actually we need to have) *a priori* reasons to decrease the parameter *a* or stop *N* somewhere. When the higher order of *a*(i) decline, then the moments tend to be capped (the inherited tails will come from the lognormality of σ).

Regime 2-a; First Method: "bleed" of higher order error

Take a "bleed" of higher order errors at the rate λ, $0\leq\lambda<1$, such as $a(N)=\lambda$ a(N-1), hence $a(N)=\lambda^N a(1)$, with $a(1)$ the conventional intensity of stochastic standard deviation. Assume μ=0.

With N=2, the second moment becomes:

$$M2(2)=\left(a(1)^2+1\right)\sigma^2\left(a(1)^2\lambda^2+1\right) \tag{10}$$

With N=3,

$$M2(3)=\sigma^2\left(1+a(1)^2\right)\left(1+\lambda^2 a(1)^2\right)\left(1+\lambda^4 a(1)^2\right) \tag{11}$$

finally, for the general N:

$$M3(N)=\left(a(1)^2+1\right)\sigma^2\prod_{i=1}^{N-1}\left(a(1)^2\lambda^{2i}+1\right) \tag{12}$$

We can reexpress (12) using the $Q-$ Pochhammer symbol $(a;q)_N=\prod_{i=1}^{N-1}\left(1-aq^i\right)$

$$M2(N)=\sigma^2\left(-a(1)^2;\lambda^2\right)_N \tag{13}$$

Which allows us to get to the limit

$$\text{Limit } M2(N)_{N\to\infty}=\sigma^2\frac{(\lambda^2;\lambda^2)_2\,(a(1)^2;\lambda^2)_\infty}{(\lambda^2-1)^2\,(\lambda^2+1)} \tag{14}$$

As to the fourth moment:

By recursion:

$$M4(N)=3\,\sigma^4\prod_{i=0}^{N-1}\left(6\,a(1)^2\lambda^{2i}+a(1)^4\lambda^{4i}+1\right) \tag{15}$$

$$M4(N)=3\,\sigma^4\left(\left(2\sqrt{2}-3\right)a(1)^2;\lambda^2\right)_N\left(-\left(3+2\sqrt{2}\right)a(1)^2;\lambda^2\right)_N \tag{16}$$

$$\text{Limit } M4(N)_{N\to\infty}=3\,\sigma^4\left(\left(2\sqrt{2}-3\right)a(1)^2;\lambda^2\right)_\infty\left(-\left(3+2\sqrt{2}\right)a(1)^2;\lambda^2\right)_\infty \tag{17}$$

So the limiting second moment for λ=.9 and a(1)=.2 is just 1.28 σ^2, a significant but relatively benign convexity bias. The limiting fourth moment is just `9.88` σ^4, more than 3 times the Gaussian's (3 σ^4), but still finite fourth moment. For small values of a and values of λ close to 1, the fourth moment collapses to that of a Gaussian.

Regime 2-b; Second Method, a Non Multiplicative Error Rate

For N recursions

$$\sigma\,(1\ \pm\ (a(1)\,(1\pm\ (\,a\,(2)\,(1\pm\ a(3)\,(\ \ ...\,)))$$

$$P(x,\ \mu,\ \sigma,\ N) = \frac{\sum_{i=1}^{L} f(x,\ \mu,\ \sigma\,(1 + (\,T^N.\,A^N\,)_i)}{L} \tag{18}$$

$\left(M^N.T + 1\right)_i$ is the i^{th} component of the $(N \times 1)$ dot product of T^N the matrix of Tuples in (6), L the length of the matrix, and A is the vector of parameters

$$A^N = \left\{a^j\right\}_{j=1,\ldots N}$$

So for instance, for N=3, T= $\{1, a, a^2, a^3\}$

$$T^3\,.\,A^3\ =\ \begin{pmatrix} a + a^2 + a^3 \\ a + a^2 - a^3 \\ a - a^2 + a^3 \\ a - a^2 - a^3 \\ -a + a^2 + a^3 \\ -a + a^2 - a^3 \\ -a - a^2 + a^3 \\ -a - a^2 - a^3 \end{pmatrix}$$

The moments are as follows:

$$\mathrm{M1}(N) = \mu \tag{19}$$

$$\mathrm{M2}(N) = \mu^2 + 2\,\sigma \tag{20}$$

$$\mathrm{M4}(N) = \mu^4 + 12\,\mu^2\,\sigma + 12\,\sigma^2 \sum_{i=0}^{N} a^{2\,i} \tag{21}$$

at the limit of $N \to \infty$

$$\mathrm{Lim}_{N\to\infty}\,\mathrm{M4}(N) = \mu^4 + 12\,\mu^2\,\sigma + 12\,\sigma^2\,\frac{1}{1 - a^2} \tag{22}$$

which is very mild.

Conclusions and Open Questions

So far we examined two regimes, one in which the higher order errors are proportionally constant, the other one in which we can allow them to decline (in two different methods and parametrizations). The difference between the two is easy to spot: the second category corresponds to naturally thin-tailed domains (higher errors decline rapidly), which can determine to be so *a priori*, something very rare on mother earth. Outside of these very special situations (say in some strict applications or clear cut sampling problems from a homogeneous population, or similar matters stripped of higher order uncertainties), the Gaussian and its siblings (along with the measures such as STD, correlation, etc.) should be completely abandoned in forecasting, along with any attempt to measure small probabilities. So thicker-tailed distributions are to be used more prevalently than initially thought.

- Can we separate the two domains along the rules of tangibility/subjectivity of the probabilistic measurement? Daniel Kahneman had a saying about measuring future states: how can one "measure" something that does not exist? So we could use:
  - Regime 1: elements entailing forecasting and "measuring" future risks. So should we use time as a dividing criterion: Anything that has time in it (meaning involves a forecast of future states) needs to fall into the first regime of non-declining proportional uncertainty parameters a(i).
  - Regime 2: conventional statistical measurements of matters patently thin-tailed, say as in conventional sampling theory, with a strong *a priori* acceptance of the methods without any form of skepticism.
- We can even work backwards, using the behavior of the estimation errors a(n) < a(1) or a(n) ≥ a(1) as a way to separate uncertainties.

Note 1

Infinite variance is not a problem at all -- yet economists have been historically scared of it. All we have to do is avoid using variance and measures in the $\mathcal{L}^2$ norm. For instance we can do much of what we currently do (even price financial derivatives) by using mean absolute deviation of the random variable, E[|x|] in place of σ, so long as the tail exponent of the power law exceeds 1 (Taleb, 2008).

Note 2

There is most certainly a cognitive dimension, rarely (or, I believe, never) addressed or investigated, in the following mental shortcomings that, from the research, appears to be common among probability modelers:

- Inability (or, perhaps, as the cognitive science literature seems to now hold, lack of motivation) to perform higher order recursions among people with Asperger (*I know that he knows that I know that he knows...*). See the second edition of *The Black Swan*, Taleb (2010).

- Inability (or lack of motivation) to transfer from one situation to another (similar to the problem of weakness of central coherence). For instance, a researcher can accept power laws in one domain yet not recognize them in another, not integrating the ideas (lack of central coherence). I have observed this total lack of central coherence with someone who can do stochastic volatility models but is unable to understand them outside the exact same conditions when doing other papers.

Note that this author is currently working on the association between models of uncertainty and mental biases and defects on the part of the operators.

Acknowledgments

Adam Elga, Jean-Philippe Bouchaud, Raphael Douady, Charles Tapiero, Aaron Brown, Dana Meyer, Andrew Gelman, Felix Salmon.

References

Abramovich and Stegun (1972) *Handbook of Mathematical Functions*, Dover Publications

David Lewis (1973) *Counterfactuals*, Harvard U. Press

Derman, E., Kani, I. (1994). Riding on a smile. *Risk* 7, 32–39.

Dupire, Bruno (1994) Pricing with a smile, *Risk*, 7, 18–20.

Feguson, Niall,(1997) [ed]. *Virtual History: Alternatives and Counterfactuals*. New York: Basic Books

Fisher, R.A. (1925), Applications of "Student's" distribution, *Metron* 5 90-104

Hull, J., White, A. (1997) The pricing of options on assets with stochastic volatilities, *Journal of Finance* , 42

Mandelbot, B. (1997) *Fractals and Scaling in Finance*, Springer.

Markowitz, H. (1952), *Portfolio Selection*, Journal of Finance.

Mosteller, Frederick & John W Tukey (1977). *Data Analysis and Regression : a Second Course in Statistics*. Addison-Wesley.

"Student" (1908) The probable error of a mean, *Biometrica* VI, 1-25

Taleb, N.N. (1997) *Dynamic Hedging: Managing Vanilla and Exotic Options*, Wiley

Taleb, N.N. (2008) Finite variance is not necessary for the practice of quantitative finance. *Complexity* 14(2)

Taleb, N.N. (2009) Errors, robustness and the fourth quadrant, *International Journal of Forecasting*, 25-4, 744--759

Parker, G., Tetlock, P.E. (2006). Counterfactual thought experiments: Why we can't live with them and how we must learn to live with them. In P.E. Tetlock, R.N. Lebow & G. Parker (eds) *Unmaking the West: What-If Scenarios that Rewrite World History*. Ann Arbor, MI: University of Michigan Press. 2006.

Why Did The Crisis of 2008 Happen?

Nassim Nicholas Taleb

DRAFT

3rd Version, August 2010

This paper —while a standalone invited essay for *New Political Economy* — synthesizes the various technical documents by the author as related to the financial crisis. It can also be used as a technical companion to *The Black Swan.*

Summary of Causes: The interplay of the following five forces, all linked to the misperception, misunderstanding, and hiding of the risks of consequential low probability events (Black Swans).

I-CAUSES

1) **Increase in hidden risks of low probability events (tail risks)** across all aspects of economic life, not just banking; *while tail risks are not possible to price*, neither mathematically nor empirically. The same nonlinearity came from the increase in debt, operational leverage, and the use of complex derivatives.

a- The author has shown that it is *impossible* to measure the risks in the tails of the distribution[i]. The errors swell in proportion to the remoteness of the event. Small variations in input, smaller than any uncertainty we have in estimation of parameters, assuming generously one has the right model, can underestimate the probability of events called of "10 sigma" (that is, 10 standard deviations) by close to *a trillion* times —a fact that has been (so far) strangely ignored by the finance and economics establishment.

b- Exposures have been built in the "fourth quadrant" [ii] , where errors are both consequential and impossible to price and vulnerability to these errors is large.

c- Fragility in the Fourth Quadrant can be re-expressed as concavity to errors, where losses from uncertain events vastly exceed possible profits from it, the equivalent of "short volatility". These exposures have been increasing geometrically.

2) **Asymmetric and flawed incentives that favor risk hiding in the tails,** two flaws in the compensation methods, based on cosmetic earnings not truly risk-adjusted ones a) *asymmetric payoff*: upside, never downside (free option); b) *flawed frequency*: annual compensation for risks that blow-up every few years, with absence of claw-back provisions.

a- *Misunderstanding of elementary notion of probabilistic payoffs across economic life*. The general public fails to notice that a manager "paid on profits" is not really "paid on profits" in the way it is presented and not compensated in the same way as the owner of a business given the absence of negative payment on losses (the *fooled by randomness* argument). States of the world in which there can be failure are ignored —"probabilistic blindness". This asymmetry is called the "manager option", or the "free option", as it behaves exactly like a call option on the company granted by the shareholders, for free or close to little compensation. Thanks to the bailout of 2008-2009 (TARP), banks used public funds to generate profits, and compensated themselves generously in the process, yet managed to convince the public and government that this compensation was justified since they brought profits to the public purse—hiding the fact that the public would have been the sole payer in the event of losses.

b- *Mismatch of bonus frequency.* Less misunderstood by policymakers, a manager paid on an annual frequency does not have an incentive to maximize profits; his incentive is to extend the time to losses so he can accumulate bonuses before eventual "blowup" for which he does not have to repay previous compensation. This provides the incentive to

make a series of asymmetric bets (high probability of small profits, small probability of large losses) *below* their probabilistic fair value[iii].

c- The agency problem is far more vicious in the tails, as it can explain the growing left-skewness (fragility) of corporations as they get larger (left-skewness is shown in Zeckhauser & Patel, 1999, rediscussed in argument on convexity).

3) **Increased promotion of methods helping to hide tail risks** VaR and similar methods promoted tail risks. See my argument that information has harmful side effects as it does increase overconfidence and risk taking.

a- I said that knowledge degrades very quickly in the tails of the distributions, making tail risks non-measurable (or, rather, impossible to *estimate* —"measure" conveys the wrong impression). Yet vendors have been promoting method of risk management called "Value at Risk", VaR, that just measures the risks in the tail! it is supposed to project the expected extreme loss in an institution's portfolio that can occur over a specific time frame at a specified level of confidence (Jorion,1997). Example: a standard daily VaR of $1 million at a 1% probability tells you that you have less than a 1% chance of losing $1 million or more on a given day. There are many modifications around VaR, "conditional VaR" [1], equally exposed to errors in the tails. Although such definition of VaR is often presented as a "maximum" loss, it is technically not so in an open-ended exposure: since, conditional on losing more than $1 million, you may lose a lot more, say $5 million. So simply, VaR encourages risk-taking in the tails and the appearance of "low volatility".

Note here that regulators made banks shift from hard heuristics (robust to model error) to such "scientific" measurements.

Criticism has been countered with the argument that "we have nothing better"; ignoring of iatrogenic effects and mere phronetic common sense.

b- *Iatrogenics of measuments (harm done by the healer)*: these estimations presented as "measures" are known to increase risk taking. Numerous experiments provide evidence that professionals are significantly influenced by numbers that they know to be irrelevant to their decision, like writing down the last 4 digits of one's social security number before making a numerical estimate of potential market moves. German judges rolling dice before sentencing showed an increase of 50% in the length of the sentence when the dice show a high number, without being conscious of it.[2]

c- *Linguistic conflation*: Calling these risk estimation "measures" create confusion in the mind of people, making them think that something in current existence (not yet to exist in the future) is being measured —these metrics are never presented as mere predictions with an abnormally huge error (as we saw, several orders of magnitude).

4) **Increased role of tail events** in economic life thanks to "complexification" by the internet and globalization, in addition to optimization of the systems.

a- *The logic of winner take all effects*: *The Black Swan* provides a review of "fat tail effects" coming from the organization of systems; consider the island effect, how a continent will have more acute concentration effects as species concentration drop in larger areas. The increase in "winner-take-all" effects is evident across economic variables (which includes blowups).

b- Optimization makes systems left-slewed, more prone to extreme losses —which can be seen in concavity effects under the perturbation of parameters.

5) **Growing misunderstanding of tail risks** Ironically while tail risks have increased, financial and economic theories that discount tail risks have been more vigorously promoted (while operators understood risks heuristically in the past[3]), particularly after the crash of 1987, after the "Nobel" for makers of "portfolio theory". Note the outrageous fact that the *entire* economics establishment missed the rise in these risks, without incurring subsequent problems in credibility.

[1] Data shows that methods meant to improve the standard VaR, like "expected shortfall" or "conditional VaR" are equally defective with economic variables --past losses do not predict future losses. Stress testing is also suspicious because of the subjective nature of "reasonable stress" number --we tend to underestimate the magnitude of outliers. "Jumps" are not predictable from past jumps.

[2] See English and Mussweiler, English Mussweiler and Strack, 2006, LeBoeuf and Shafir, 2006.

[3] The key problem with finance theory has been supplanting embedded and time-derived heuristics, such as the interdicts against debt and forecasting, with models akin to "replacing a real hand with an artificial one".

Principal errors by the economics establishment that contribute to increasing fragility:

a- *Ignorance of "true" fat tail effects*; or misunderstanding that fat tails lead to massive imprecision in the measurement of low probability events (such as the use of Poisson jumps by Merton, 1976 or the more general versions of subordinated processes —these models fit the past with precision on paper but are impossible to calibrate in practice and induce a false sense of confidence). Misunderstanding that true-fat-tails cancels the core of financial theory and econometric methods used in practice.

b- *Lack of awareness of the effect of parameter estimation on a model.* Some models —actually almost all models — take parameters for granted when the process of parameter discovery in real-life leads to massive degradation of their results owing to convexity effects from such layer of uncertainty.

c- *Interpolation v/s Extrapolation*. Misunderstanding of the "atypicality of events" —looking for past disturbances for guidance when we have obvious evidence of lack of precedence of such events. For instance, Rogoff and Reinhart (2010) look at past data without realizing that in fat tailed domains, one should extrapolate from history, instead of interpolating or looking for naive similarities (Lucian's largest mountain).

d- *Optimization*. It can be shown that optimization causes fragility when concave under perturbation errors, i.e., most cases.

e- *Economies of scale*. There are fragilities coming from size, both for the institutions and causing externalities[iv].

II-Responsible Parties

1) Government Officials of Both Administrations promoting blindness to tail risks and nonlinearities (e.g. Bernanke's pronouncement of "great moderation") and flawed tools in the hands of policymakers not making the distinction between different classes of randomness.

2) Bankers/Company executives: The individuals had an incentive to hide tail risks as a safe strategy to collect bonuses.

3) Risk vendors and professional associations: CFA, IAFE promotion of portfolio theory and Value-at-risk methods.

4) Business schools and the economics establishment: They kept promoting and teaching portfolio theory and inadequate risk measurement methods on grounds that "we need to give students something" (arguments used by medieval medicine). They still do[4].

5) Regulators: Promoted quantitative risk methods (VaR) over heuristics, use of flawed risk metrics (AAA), and encouraged a certain class of risk taking.

6) Bank of Sweden Prize, a.k.a. "Nobel" in Economics: gave the Nobel stamp to empirically, mathematically, and scientifically invalid theories, such as portfolio theory, Engle's GARCH, and many more. In general their scientific invalidity comes from the use of wrong models of uncertainty that provide exactly the opposite results to what an empirically and mathematically more rigorous model of uncertainty would do.

Ethical considerations. Surprisingly the economics establishment should have been aware of the use the wrong tools and complete fiasco in the theories, but they kept pushing the warnings under the rug, or hiding their responses. There has been some diffusion of responsibility that is at the core of the system. This author has debated: Robert Engle, Myron Scholes, Robert Merton, and Stephen Ross, among others, without any hint of their willing to accept the very notion of the risks they were creating with their Procrustean bed methods of approximation —prompting the following metaphor by this author: "they are cutting part of someone's brains and claiming that we have a human with 99% accuracy". The only favorable reaction this author encountered was even more outrageous, from those, like Robert Shiller, with the half-way "you may have a point but you go too far" that can be vastly more damaging to society that just regular attacks.

III- Suggested Remedies

As we saw with banks, Toyota's problem, the BP oil spill, an economic system with a severe agency

[4] In early 2009 a Forbes journalist in the process of writing my profile spoke to NYU's Robert Engle who got the Bank of Sweden Prize ("Nobel") for methods that patently have never worked outside papers. He reported to me that Engle response was that *academia was not responsible for tail risks*, since it is the government job to cover the losses beyond a certain point. This is the worst moral hazard argument that played into the hands of the *Too Big to Fail* problem.

problem builds a natural tendency to push and hide risks in the tails, even without help from the economics establishment. Risks keep growing *where they can be seen the least*; there is a need to break the moral hazard by making everyone accountable both chronologically and statistically.

Hence the principle: ***The captain goes down with the ship; all captains and all ships*****:** making everyone involved in risk-bearing accountable, no exception, not a single one. Morally, legally, whatever can be done. That includes the "Nobel" committee (Bank of Sweden), the academic establishment, the rating agencies, forecasters, bank managers, etc[5].

Time to realize that capitalism is not about free options[6].

Note that organizations such as the CFA and American Finance Association, RiskMetrics and such vendors, and finance departments in business schools, those that promoted tools that blew up society do not seem concerned at all into changing their methods or accepting their role. And they are *currently*, at the time of writing, still in the process of blowing up society.

REFERENCES

Birte Englich and Thomas Mussweiler,2001, "Sentencing under Uncertainty: Anchoring Effects in the Courtroom," Journal of Applied Social Psychology, vol. 31, no. 7 92001), pp. 1535-1551

Birte Englich, Thomas Mussweiler, and Fritz Strack, 2006, "Playing Dice with Criminal Sentences: the Influence of Irrelevant Anchors on Experts' Judicial Decision Making," Personality and Social Psychology Bulletin, vol. 32, no 2 (Feb. 2006), pp. 188-200.

Degeorge, François, Jayendu Patel, and Richard Zeckhauser, 1999, "Earnings Management to Exceed Thresholds." Journal of Business 72(1): 1–33.

Derman, E. and Taleb, N.N. (2005) The Illusion of Dynamic Replication, Quantitative Finance, vol. 5, 4

Haug, E.G. and Taleb, N.N. ,2010, Option Traders Use Sophisticated Heuristics, Never the Black-Scholes-Merton Equation, forthcoming, Journal of Economic Behavior and Organizations.

[5] Conversations of the author with the King of Sweden and members of the Swedish Academy resulted in the following astonishing observation: they do feel concerned, nor act as if they are in any way responsible for the destruction since, for them, "this is not the Nobel", just a bank of Sweden price.

[6] Speculators using their own funds have been reviled, but unlike professors, *New York Times* journalists, and others, they (particularly those without the free option of society's bailout) bear directly the costs of their mistakes.

Le Boeuf. R.A., and Shafir E. The Long and Short of It: Physical Anchoring Effects. *J. Behav. Dec. Making*, 19: 393–406 (2006)

Makridakis, S., & Taleb, N., 2009, "Decision making and planning under low levels of predictability", *International Journal of Forecasting*

Merton R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Merton, R. C. (1992). *Continuous-time finance* (revised edition). Blackwell.

Mussweiler, T., & Strack, F. (2000). Numeric judgments under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology, 39*, 495-518.

Rogoff, K. and Reinhart C. 2010, *This Time is Different, Eight Centuries of Financial Folly*, Princeton University Press.

Taleb N.N.and Pilpel, A. (2007)Epistemology and Risk Management, "Risk and Regulation", 13, Summer 2007

Taleb, N.N. (1997). Dynamic Hedging: Managing Vanilla and Exotic Options. New York: John Wiley & Sons. ISBN 0-471-15280-3.

Taleb, N. N. (2004) "I problemi epistemologici del risk management " in: Daniele Pace (a cura di) "Economia del rischio. Antologia di scritti su rischio e decisione economica", Giuffrè, Milano

Taleb, N. N. (2004) "On Skewness in Investment Choices." Greenwich Rountable Quarterly 2.

Taleb, N. N. (2004) "Roots of Unfairness." Literary Research/Recherche littéraire. 21(41-42): 241-254.[57]

Taleb, N. N. (2004) "These Extreme Exceptions of Commodity Derivatives." in Helyette German, Commodities and Commodity Derivatives. New York: Wiley.

Taleb, N. N. (2004) Bleed or Blowup: What Does Empirical Psychology Tell Us About the Preference For Negative Skewness? , Journal of Behavioral Finance, 5

Taleb, N. N. (2005) "Fat Tails, Asymmetric Knowledge, and Decision making: Essay in Honor of Benoit Mandelbrot's 80th Birthday." Technical paper series, Willmott (March): 56-59.

Taleb, N. N. (2008) Infinite Variance and the Problems of Practice, Complexity, 14(2).

Taleb, N., and Tapiero, C. Too Big to Fail and the Fallacy of Large Institutions (forthcoming)

Taleb, N., and Tapiero, C.,2010, The Risk Externalities of Too Big to Fail in press,Physica A

Taleb, N.N. (2007) "Black Swan and Domains of Statistics", The American Statistician, August 2007, Vol. 61, No. 3

Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 25 (2009) 744–759
international journal of forecasting
www.elsevier.com/locate/ijforecast

Errors, robustness, and the fourth quadrant

Nassim Nicholas Taleb

New York University–Polytechnic Institute and Universa Investments, United States

Abstract

The paper presents evidence that econometric techniques based on variance – L^2 norm – are flawed and do not replicate. The result is un-computability of the role of tail events. The paper proposes a methodology to calibrate decisions to the degree (and computability) of forecast error. It classifies decision payoffs in two types: simple (true/false or binary) and complex (higher moments); and randomness into type-1 (thin tails) and type-2 (true fat tails), and shows the errors for the estimation of small probability payoffs for type 2 randomness. The fourth quadrant is where payoffs are complex with type-2 randomness. We propose solutions to mitigate the effect of the fourth quadrant, based on the nature of complex systems.

Keywords: Complexity; Decision theory; Fat tails; Risk management

1. Background and purpose

It appears scandalous that, of the hundreds of thousands of professionals involved, including prime public institutions such as the World Bank, the International Monetary Fund, different governmental agencies and central banks, private institutions such as banks, insurance companies, and large corporations, and, finally, academic departments, only a few individuals considered the possibility of the total collapse of the banking system that started in 2007 (and is still worsening at the time of writing), let alone the economic consequences of such breakdown. Not a single official forecast turned out to be close to the outcome experienced—even those issuing "warnings" did not come close to the true gravity of the situation. A few warnings about the risks, such as Taleb (2007a) or the works of the economist Nouriel Roubini,[1] went unheeded, often ridiculed.[2] Where did such sophistication go? In the face of miscalculations of such proportion, it would seem fitting to start an examination of the conventional forecasting methods for risky outcomes and assess their fragility—indeed, the size of the damage comes from confidence in forecasting and the mis-estimation of potential forecast errors for a certain classes of variables and a certain type of exposures. However, this was not

E-mail address: nnt@fooledbyrandomness.com.

[1] "Dr. Doom", *New York Times*, August 15, 2008.

[2] Note the irony that the ridicule of the warnings in Taleb (2007a) and other ideas came from the academic establishment, not from the popular press.

0169-2070/$ - see front matter © 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2009.05.027

the first time such events have happened—nor was it a "Black Swan" (when capitalized, an unpredictable outcome of high impact) to the observer who took a close look at the robustness and empirical validity of the methods used in economic forecasting and risk measurement.

This examination, while grounded in economic data, generalizes to all decision-making under uncertainty in which there is a potential miscalculation of the risk of a consequential rare event. The problem of concern is the rare event, and the exposure to it, of the kind that can fool a decision maker into taking a certain course of action based on a misunderstanding of the risks involved.

2. Introduction

Forecasting is a serious professional and scientific endeavor with a certain purpose, namely to provide predictions to be used in formulating decisions, and taking actions. The forecast translates into a decision, and, accordingly, the uncertainty attached to the forecast, i.e., the error, needs to be endogenous to the decision itself. This holds particularly true of risk decisions. In other words, the use of the forecast needs to be determined — or modified — based on the estimated accuracy of the forecast. This in turn creates an interdependency about what we should or should not forecast—as some forecasts can be harmful to decision makers.

Fig. 1 gives an example of harm coming from building risk management on the basis of extrapolative (usually highly technical) econometric methods, providing decision-makers with false confidence about the risks, and leaving society exposed to several trillions in losses that put capitalism on the verge of collapse.

A key word here, *fat tails*, implies the outsized role in the total statistical properties played by one single observation—such as one massive loss coming after years of stable profits or one massive variation unseen in past data.

- "Thin-tails" lead to ease in forecasting and tractability of the errors;
- "Thick-tails" imply more difficulties in getting a handle on the forecast errors and the fragility of the forecast.

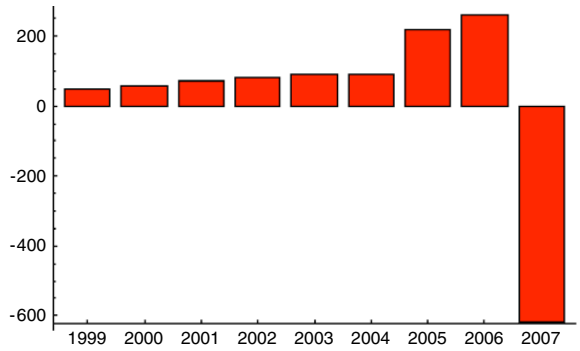


Fig. 1. Indy Mac's annual income (in millions) between 1998 and 2007. We can see fat tails at work. Tragic errors come from underestimating potential losses, with the best known cases being FNMA, Freddie Mac, Bear Stearns, Northern Rock, and Lehman Brothers, in addition to numerous hedge funds.

Close to 1000 financial institutions have shut down in 2007 and 2008 from the underestimation of outsized market moves, with losses up to 3.6 trillion.[3] Had their managers been aware of the unreliability of the forecasting methods (which were already apparent in the data), they would have requested a different risk profile, with more robustness in risk management and smaller dependence on complex derivatives.

2.1. The smoking gun

We conducted a simple scientific examination of economic data, using a near-exhaustive set that includes 38 "tradable" variables[4] that allow for daily prices: major equity indices across the globe (US, Europe, Asia, Latin America), most metals (gold, silver), major interest rate securities, and main currencies — what we believe represents around 98% of tradable volume.

[3] Bloomberg, Feb 5, 2009.

[4] We selected a near-exhaustive set of economic data that includes "tradable" securities that allow for a future or a forward market: most equity indices across the globe, most metals, most interest rate securities, and most currencies. We collected all available traded futures data—what we believe represents around 98% of tradable volume. The reason we selected tradable data is because of the certainty of the practical aspect of a price on which one can transact: a nontradable currency price can lend itself to all manner of manipulation. More precisely we selected "continuously rolled" futures in which the returns from holding a security are built-in. For instance, analyses of Dow Jones that fail to account for dividend payments or analyses of currencies that do not include interest rates provide a bias in the measurement of the mean and higher moments.

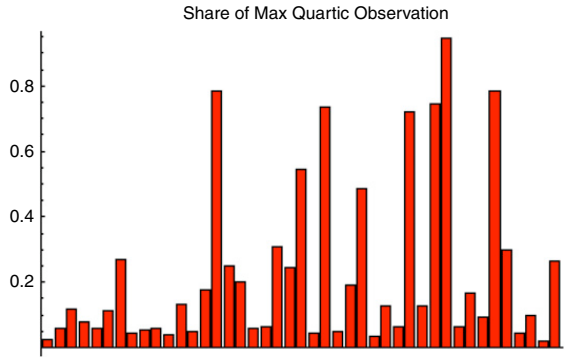


Fig. 2. The smoking gun: Maximum contribution to the fourth moment kurtosis coming from the largest observation in ∼10,000 (29–40 years of daily observations) for 43 economic variables. For the Gaussian the number is expected to be ∼0.006 for $n = 10{,}000$.

We analyzed the properties of the logarithmic returns $r_{t,\Delta t} = \mathrm{Log}\left(\frac{X_t}{X_{t-\Delta t}}\right)$, where Δt can be 1 day, 10 days, or 66 days (non-overlapping intervals).[5]

A conventional test of nonnormality used in the literature is the excess kurtosis over the normal distribution. Thus, we measured the fourth noncentral moment $k(\Delta t) = \frac{\sum r_{t,\Delta t}^4}{n}$ of the distributions and focused on the stability of the measurements.

By examining Table 1 and Figs. 2 and 3, it appears that:

(1) Economic variables (currency rates, financial assets, interest rates, commodities) are patently fat tailed—with no known exception. The literature (Bundt & Murphy, 2006) shows that this also applies to data not considered here, owing to a lack of daily changes, such as GDP, or inflation.

(2) Conventional methods, not just those relying on a Gaussian distribution, but those based on least-square methods, or using variance as a measure of dispersion, are, according to the data, incapable of tracking the kind of "fat-tails" we see (more technically, in the L^2 norm, as will be discussed in Section 5). The reason is that most of the kurtosis is concentrated in a few observations, making it practically unknowable using conventional methods—see Fig. 2. Other tests in Section 5 (the conditional expectation above a threshold) show further instability. This incapacitates least-square methods, linear regression, and similar tools, including risk management methods such as "Gaussian Copulas" that rely on correlations or any form of the product of random variables.[6, 7, 8]

[5] By convention we use $t = 1$ as one business day.

[6] This should predict, for instance, the total failure in practice of the ARCH/GARCH methods (Engle, 1982), in spite of their successes in-sample, and in academic citations, as they are based on the behavior of squares.

[7] One counterintuive result is that sophisticated operators do not seem to be aware of the norm they are using, thus mis-estimating volatility, see Goldstein and Taleb (2007).

[8] Practitioners have blamed the naive L^2 reliance on the risk management of credit risk for the blowup of banks in the crisis that started in 2007. See Felix Salmon's "Recipe For Disaster: The Formula That Killed Wall Street" in Wired. 02/23/2009.

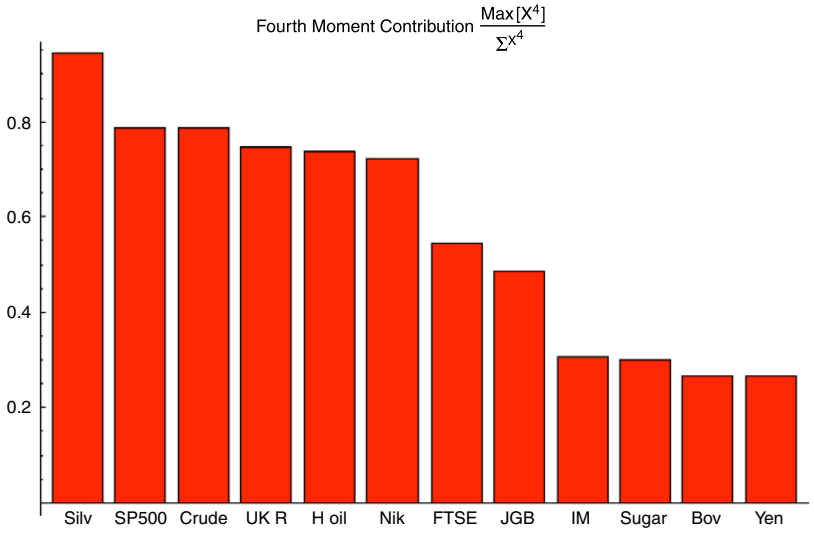


Fig. 3. A selection of the 12 most acute cases among the 43 economic variables.

Table 1
Fourth Noncentral Moment at daily, 10 day, and 66 day windows for the random variables.

	K (1)	K (10)	K (66)	Max quartic	Years
Australian Dollar/USD	6.3	3.8	2.9	0.12	22
Australia TB 10y	7.5	6.2	3.5	0.08	25
Australia TB 3y	7.5	5.4	4.2	0.06	21
BeanOil	5.5	7.0	4.9	0.11	47
Bonds 30Y	5.6	4.7	3.9	0.02	32
Bovespa	24.9	5.0	2.3	0.27	16
British Pound/USD	6.9	7.4	5.3	0.05	38
CAC40	6.5	4.7	3.6	0.05	20
Canadian Dollar	7.4	4.1	3.9	0.06	38
Cocoa NY	4.9	4.0	5.2	0.04	47
Coffee NY	10.7	5.2	5.3	0.13	37
Copper	6.4	5.5	4.5	0.05	48
Corn	9.4	8.0	5.0	0.18	49
Crude Oil	29.0	4.7	5.1	0.79	26
CT	7.8	4.8	3.7	0.25	48
DAX	8.0	6.5	3.7	0.2	18
Euro Bund	4.9	3.2	3.3	0.06	18
Euro Currency/DEM previously	5.5	3.8	2.8	0.06	38
Eurodollar Depo 1M	41.5	28.0	6.0	0.31	19
Eurodollar Depo 3M	21.1	8.1	7.0	0.25	28
FTSE	15.2	27.4	6.5	0.54	25
Gold	11.9	14.5	16.6	0.04	35
Heating Oil	20.0	4.1	4.4	0.74	31
Hogs	4.5	4.6	4.8	0.05	43
Jakarta Stock Index	40.5	6.2	4.2	0.19	16
Japanese Gov Bonds	17.2	16.9	4.3	0.48	24
Live Cattle	4.2	4.9	5.6	0.04	44
Nasdaq Index	11.4	9.3	5.0	0.13	21
Natural Gas	6.0	3.9	3.8	0.06	19
Nikkei	52.6	4.0	2.9	0.72	23
Notes 5Y	5.1	3.2	2.5	0.06	21
Russia RTSI	13.3	6.0	7.3	0.13	17
Short Sterling	851.8	93.0	3.0	0.75	17
Silver	160.3	22.6	10.2	0.94	46
Smallcap	6.1	5.7	6.8	0.06	17
SoyBeans	7.1	8.8	6.7	0.17	47
SoyMeal	8.9	9.8	8.5	0.09	48
Sp500	38.2	7.7	5.1	0.79	56
Sugar # 11	9.4	6.4	3.8	0.3	48
SwissFranc	5.1	3.8	2.6	0.05	38
TY10Y Notes	5.9	5.5	4.9	0.1	27
Wheat	5.6	6.0	6.9	0.02	49
Yen/USD	9.7	6.1	2.5	0.27	38

(3) There is no evidence of "convergence to normality" by aggregation, i.e., looking at the kurtosis of weekly or monthly changes. The "fatness" of the tails seems to be conserved under aggregation.

Clearly, had decision-makers been aware of such facts, and such unreliability of conventional methods in tracking large deviations, fewer losses would have been incurred, as they would have reduced exposures in some areas rather than rely on more "sophisticated" methods. The financial system has been fragile, as this simple test shows, with the evidence staring at us all along.

2.2. The problem of large deviations

2.2.1. The empirical problem of small probabilities

The central problem addressed in this paper is that small probabilities are difficult to estimate empirically (since the sample set for these is small), with a greater error rate than that for more frequent events. But since, in some domains, their effects can be consequential, the error concerning the contribution of small probabilities to the total moments of the distribution becomes disproportionately large. The problem has been dealt with by assuming a probability distribution and extrapolating into the tails—which brings model error into play. Yet, as we will discuss, model error plays a larger role with large deviations.

2.2.2. Links to decision theory

It is not necessary here to argue that a decision maker needs to use a full tableau of payoffs (rather than the simple one-dimensional average forecast) and that payoffs from decisions vary in their sensitivity to forecast errors. For instance, while it is acceptable to take a medicine that might be effective with a 5% error rate, but offers no side effects otherwise, it is foolish to play Russian roulette with the knowledge that one should win with a 5% error rate—indeed, standard theory of choice under uncertainty requires the use of full probability distributions, or at least a probability associated with every payoff. But so far this simple truism has not been integrated into the forecasting activity itself—as no classification has been made concerning the tractability and consequences of the errors. To put it simply, the mere separation between forecasting and decisions is lacking in both rigor and practicality, as it ruptures the link between forecast error and the quality of the decision.

The extensive literature on decision theory and choices under uncertainty so far has limited itself to (1) assuming *known* probability distributions (except for a few exceptions in which this type of uncertainty has been called "ambiguity"[9]), and (2) ignoring fat tails. This paper introduces a new structure of fat tails and classification of classes of randomness into the analysis, and focuses on the interrelation between errors and decisions. To establish a link between decision and quality of forecast, this analysis operates along two qualitative lines: qualitative differences between decisions along their vulnerability to error rates on one hand, and qualitative differences between two types of distributions of error rates. So there are two *distinct* types of decisions, and two *distinct* classes of randomness.

This classification allows us to isolate situations in which forecasting needs to be suspended—or a revision of the decision or exposure may be necessary. What we call the "fourth quadrant" is the area in which both the magnitude of forecast errors is large and the sensitivityt to these errors is consequential. What we recommend is either changes in the payoff itself (clipping exposure) or the shifting of exposures away from that part. For that we will provide precise rules.

The paper is organized as follows. First, we classify decisions according to targeted payoffs. Second, we discuss the problem of rare events, as these are the ones that are both consequential and hard to predict. Third, we present the classification of the two categories of probability distributions. Finally, we present the "fourth quadrant" and what we need to do to escape it, thus answering the call for how to handle "decision making under low predictability".

3. The different types of decisions

The first type of decisions is simple, it aims at "binary" payoffs, i.e. you just care whether something is true or false. Very true or very false does not matter. Someone is either pregnant or not pregnant. A biological experiment in the laboratory or a bet about the outcome of an election belong to this category. A scientific statement is traditionally considered "true" or "false" with some confidence interval. More technically, they depend on the zeroth moment, namely just on the probability of events, and not their magnitude —for these one just cares about "raw" probability.[10]

[9] Ellsberg's paradox, Ellsberg (1961); see also Gardenfors and Sahlin (1982) and Levi (1986).

[10] The difference can be best illustrated as follows. One of the most erroneous comparisons encountered in economics is the one between the "wine rating" and "credit rating" of complex securities. Errors in wine rating are hardly consequential for the buyer (the "payoff" is binary); errors in credit ratings have bankrupted banks, as these carry massive payoffs.

Clearly these are not very prevalent in life—they mostly exist in laboratory experiments and in research papers.

The second type of decisions depends on more complex payoffs. The decision maker does not just care about the frequency, but about the impact as well, or, even more complex, some function of the impact. So there is another layer of uncertainty of impact. These depend on higher moments of the distribution. When one invests one does not care about the frequency, how many times he makes or loses, he cares about the expectation: how many times money is made or lost *times* the amount made or lost. We will see that there are even more complex decisions.

More formally, where $p[x]$ is the probability distribution of the random variable x, and D the domain on which the distribution is defined, the payoff $\lambda(x)$ is defined by integrating on D as:

$$\lambda(x) = \int f(x)p(x)\mathrm{d}x.$$

Note that we can incorporate utility or nonlinearities of the payoff in the function $f(x)$. But let us ignore utility for the sake of simplification.

For a simple payoff, $f(x) = 1$. So $L(x)$ becomes the simple probability of exceeding x, since the final outcome is either 1 or 0 (or 1 and -1).

For more complicated payoffs, $f(x)$ can be complex. If the payoff depends on a simple expectation, i.e., $\lambda(x) = E[x]$, the corresponding function $f(x) = x$, and we need to ignore frequencies since it is the payoff that matters. One can be right 99% of the time, but this does not matter at all, since with some skewed distributions, the consequences of the expectation of the 1% error can be too large. Forecasting typically has $f(x) = x$, a linear function of x, while measures such as least squares depend on the higher moments $f(x) = x^2$.

Note that some financial products can even depend on the fourth moment (see Table 2).[11]

Next we turn to a discussion of the problem of rare events.

[11] More formally, a linear function with respect to the variable x has no second derivative; a convex function is one with a positive second derivative. By expanding the expectation of $f(x)$ we end up with $E[f(x)] = f(x)e[\Delta x] + 1/2f''(x)E[\Delta x2] + \cdots$, and hence higher orders matter to the extent of the importance of higher derivatives.

4. The problem of rare events

The passage from theory to the real world presents two distinct difficulties: "inverse problems" and "pre-asymptotics".

4.1. Inverse problems

It is the greatest difficulty one can encounter in deriving properties. In real life we do not observe probability distributions, we just observe events. So we do not know the statistical properties — until, of course, after the fact — as we can see in Fig. 1. Given a set of observations, plenty of statistical distributions can correspond to the exact same realizations—each would extrapolate differently outside the set of events on which it was derived. The inverse problem is more acute when more theories, more distributions can fit a set of data—particularly in the presence of nonlinearities or nonparsimonious distributions.[12]

So this inverse problem is compounded of two problems:

+ *The small sample properties of rare events*, as these will be naturally rare in a past sample. This is also acute in the presence of nonlinearities, as the families of possible models/parametrization explode in numbers.
+ *The survivorship bias effect of high impact rare events*. For negatively skewed distributions (with a thicker left tail), the problem is worse. Clearly, catastrophic events will be necessarily absent from the data, since the survivorship of the variable itself will depend on such effect. Thus, left tailed distributions will (1) overestimate the mean; (2) underestimate the variance and the risk.

Fig. 4 shows how we normally lack data in the tails; Fig. 5 shows the empirical effect (see Fig. 6).

4.2. Pre-asymptotics

Theories can be extremely dangerous when they were derived in idealized situations, the asymptote, but are used outside the asymptote (at its limit, say infinity

[12] A Gaussian distribution is parsimonious (with only two parameters to fit). But the problem of adding layers of possible jumps, each with a different probabilities, opens up endless possibilities of combinations of parameters.

Table 2
Tableau of decisions.

Mo	M1	M2+
"True/False"	Expectations	
	LINEAR PAYOFF	NONLINEAR PAYOFF
$f(x) = 0$	$f(x) = 1$	$f(x)$ nonlinear$(= x^2, x^3,$ etc.)
Medicine (health not epidemics)	Finance: nonleveraged investment	Derivative payoffs
Psychology experiments	Insurance, measures of expected shortfall	Dynamically hedged portfolios
Bets (prediction markets)	General risk management	Leveraged portfolios (around the loss point)
Binary/Digital derivatives	Climate	Cubic payoffs (strips of out of the money options)
Life/Death	Economics (Policy)	Errors in analyses of volatility
	Security: Terrorism, Natural catastrophes	Calibration of nonlinear models
	Epidemics	Expectation weighted by nonlinear utility
	Casinos	Kurtosis-based positioning ("volatility trading")

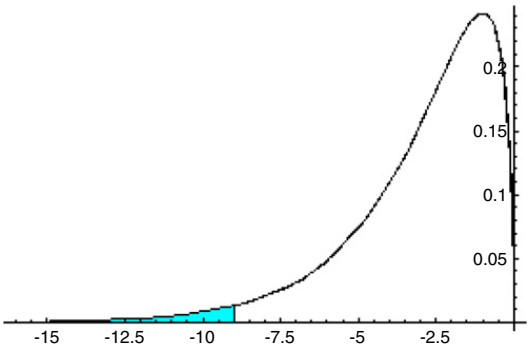


Fig. 4. The confirmation bias at work. The shaded area shows what tend to be missing from the observations. For negatively-skewed, fat-tailed distributions, we do not see much of negative outcomes for surviving entities AND we have a small sample in the left tail. This illustrates why we tend to see a better past for a certain class of time series than is warranted.

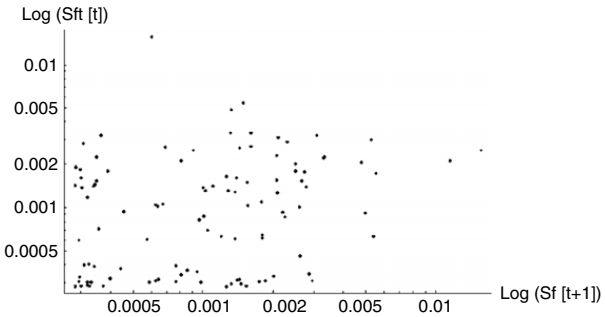


Fig. 5. Outliers don't predict outliers. The plot shows (on a logarithmic scale) a shortfall in one given year against the shortfall the following one, repeated throughout for the 43 variables. A shortfall here is defined as the sum of deviations in excess of 7%. Past large deviations do not appear to predict future large deviations, at different lags.

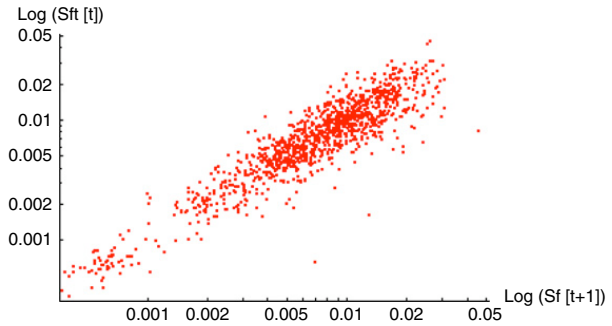


Fig. 6. Regular events predict regular events. This plot shows, by comparison with Fig. 5, how, for the same variables, the mean deviation in one period predicts the one in the subsequent period.

or the infinitesimal). Some asymptotic properties do work well pre-asymptotically (as we'll see, with type-1 distributions), which is why casinos do well, but others do not, particularly when it comes to the class of fat-tailed distributions.

Most statistical education is based on these asymptotic, laboratory-style Platonic properties—yet we take economic decisions in the real world that very rarely resembles the asymptote. Most of what students of statistics do is assume a structure, typically with a known probability. Yet the problem we have is not so much making computations once you know the probabilities as finding the true distribution.

5. The two probabilistic structures

There are two classes of probability domains — very distinct qualitatively and quantitatively — according to precise mathematical properties. The first, Type-1, we call "benign" thin-tailed nonscalable,

the second, Type 2, "wild" thick tailed scalable, or fractal (the attribution "wild" comes from the classification of Mandelbrot, 1963, 2001).

Taleb (2009) makes the distinction along the lines of convergence to the Central Limit Theorem. Type-1 converges in an acceptable form, while Type-2 either does not converge (infinite variance), or converges only in a remote asymptote and needs to be treated pre-asymptotically. Taleb (2009) also shows that one of the mistakes in the economics literature that "fattens the tails", with two main classes of nonparsimonious models and processes (the jump-diffusion processes of Merton, 1976,[13] or stochastic volatility models such as Engels' ARCH[14]) is to believe that the second type of distribution is amenable to analyses like the first—except with fatter tails. In reality, a fact commonly encountered by practitioners is that fat-tailed distributions are very unwieldy—as we can see in Fig. 2. Furthermore, we often face a problem of mistaking one for the other: a process that is extremely well behaved, but, on occasions, delivers a very large deviation, can easily be mistaken for a thin-tailed one—a problem known as the "problem of confirmation" (Taleb, 2007a,b). So we need to be suspicious of the mistake of taking Type-2 for Type-1, as it is more severe (and more readily made) than the one in the other direction.[15]

As we saw from the data presented, this classification of "fat tails" does not just mean having a fourth moment worse than the Gaussian. The Poisson distribution, or a mixed distribution with a known Poisson jump, would have tails thicker than the Gaussian; but this mild form of fat tails can be dealt with rather easily—the distribution has all its moments finite. The problem comes from the structure of the decline in probabilities for larger deviations and the ease with which the tools at our disposal can be tripped into producing erroneous results from observations of data in a finite sample and jumping to wrong decisions.

[13] See the general decomposition into diffusion and jump (non-scalable) in Duffie, Pan, and Singleton (2000) and Merton (1976); and the discussion in Baz and Chacko (2004) and Haug (2007).

[14] Engle (1982).

[15] Makridakis et al. (1993) and Makridakis and Hibon (2000) present evidence that more complicated methods of forecasting do not deliver superior results to simple ones (already bad). The obvious reason is that the errors in calibration swell with the complexity of the model.

5.1. The scalable property of type-2 distributions

Take a random variable x. With scalable distributions, asymptotically, for x large enough (i.e. "in the tails"), $\frac{P[X>nx]}{P[X>x]}$ depends on n, not on x (the same property can hold for $P[X < nx]$ for negative values). This induces statistical self-similarities. Note that owing to the finiteness of the realizations of random variables, and the lack of samples in the tails, we might not be able to observe such a property, yet not be able to rule out.

For economic variables, there is no fundamental reason for the ratio of "exceedances" (i.e., the cumulative probability of exceeding a certain threshold) to decline, as both the numerator and the denominators are multiplied by 2.

This self-similarity at all scales generates power-law, or Paretian, tails, i.e., above a crossover point, $P[X > x] = Kx^{-\alpha}$.[16, 17]

Let us now draw the implications of type-2 distributions.

5.1.1. Finiteness of moments and higher order effects

For thick tailed distributions, moments higher than α are not "finite", i.e., they cannot be computed. They can certainly be measured in finite samples—thus giving the illusion of finiteness. But they typically show a great degree of instability. For instance, a distribution with an infinite variance will always provide, in a sample, the illusion of finiteness of variance.

In other words, while errors converge for type-1 distributions, the expectations of higher orders of x, say of order n, such as $1/n!E[x^n]$, where x is the error, do not decline; in fact, they become explosive (see Fig. 7).

[16] Scalable discussions: introduced by Mandelbrot (1963), Mandelbrot (1997) and Mandelbrot and Taleb (in press).

[17] Complexity and power laws: Amaral et al. (1997), Sornette (2004), and Stanley, Amaral, Gopikrishnan, and Plerou (2000); for scalability in different aspects of financial data, Gabaix, Gopikrishnan, Plerou, and Stanley (2003a,b), Gabaix, Ramalho, and Reuter (2003c), Gopikrishnan, Meyer, Amaral, and Stanley (1998), Gopikrishnan, Plerou, Amaral, Meyer, and Stanley (1999), and Gopikrishnan, Plerou, Gabaix, and Stanley (2000). For the statistical mechanics of scale-free networks see Albert, Jeong, and Barabási (2000), Albert and Barabasi (2002) and Barabási and Albert (1999). The "sandpile effect" (i.e., avalanches and cascades) is discussed by Bak (1996) and Bak, Tang, and Wiesenfeld (1987, 1988), as power laws arise from conditions of self-organized criticality.

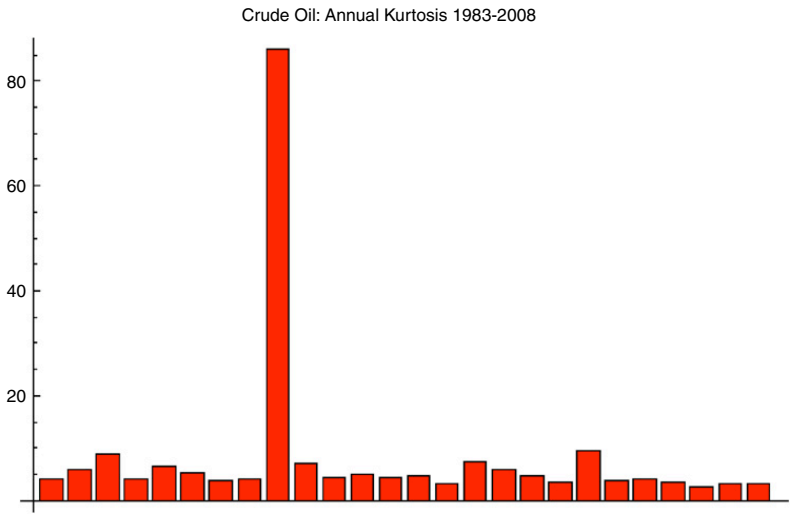


Fig. 7. Kurtosis over time: example of an "infinite moment". The graph shows the fourth moment for crude oil in annual nonoverlapping observations between 1982 and 2008. The instability shows in the dependence of the measurement on the observation window.

5.1.2. "Atypicality" of moves

For thin tailed domains, the conditional expectation of a random variable X, conditional on its exceeding a number K, converges to K for larger values of K.

$$\lim_{K\to\infty} E[X|_{X>K}] = K.$$

For instance, the conditional expectation for a Gaussian variable (assuming a mean of 0) conditional on the variable exceeding 0 is approximately 0.8 standard deviations. But with K equals 6 standard deviations, the conditional expectation converges to 6 standard deviations. The same applies to all of the random variables that do not have a Paretian tail. This induces some "typicality" of large moves.

For fat tailed variables, such a limit does not seem to hold:

$$\lim_{K\to\infty} E[X|_{X>K}] = Kc,$$

where c is a constant. For instance, the conditional expectation of a market move, given that it is in excess of 3 mean deviations, will be around 5 mean deviations. The expectation of a move conditional on it being higher than 10 mean deviations will be around 18. This property is quite crucial.

The atypicality of moves has the following significance.

– One may correctly predict a given event, say, a war, a market crash, or a credit crisis. But the amplitude of the damage will be unpredicted. The open-endedness of the outcomes can cause a severe miscalculation of the expected payoff function. For instance, the investment bank Morgan Stanley predicted a credit crisis but was severely hurt (and needed to be rescued) because it did not anticipate the extent of the damage.
– Methods like Value-at-Risk[18] that may correctly compute, say, a 99% probability of not losing no more than a given sum, called "value-at-risk", will nevertheless miscompute the conditional expectation should such a threshold be exceeded. For instance, one has 99% probability of not exceeding a $1 million loss, but should such a loss occur, it can be $10 million or $100 million.

This lack of typicality is of some significance. Stress testing and scenario generation are based on assuming a "crisis" scenario and checking robustness to it. Unfortunately such luxury is not available for fat tails, as "crises" do not have a typical magnitude.

Tables 3 and 4 show the evidence of a lack of convergence to thin tails, and hence a lack of "typicality" of the moves. We stopped for segments for which the number of observations becomes small, since a lack of observations in the tails can provide the illusion of "thin" tails.

[18] For the definition of Value at Risk see, Jorion (2001); for a critique, see Joe Nocera, "Risk Mismanagement: What led to the Financial Meltdown", *New York Time Magazine*, Jan 2, 2009.

Table 3
Conditional expectation for moves $> K$, 43 economic variables.

K, Mean deviations	Mean move (in MAD) in excess of K	n
1	2.01443	65,958
2	3.0814	23,450
3	4.19842	8,355
4	5.33587	3,202
5	6.52524	1,360
6	7.74405	660
7	9.10917	340
8	10.3649	192
9	11.6737	120
10	13.8726	84
11	15.3832	65
12	19.3987	47
13	21.0189	36
14	21.7426	29
15	24.1414	21
16	25.1188	18
17	27.8408	13
18	31.2309	11
19	35.6161	7
20	35.9036	6

Table 4
Conditional expectation for moves $< K$, 43 economic variables.

K, Mean deviations	Average move (in MAD) below K	n
−1	−2.06689	62,803
−2	−3.13423	23,258
−3	−4.24303	8,676
−4	−5.40792	3,346
−5	−6.66288	1,415
−6	−7.95766	689
−7	−9.43672	392
−8	−11.0048	226
−9	−13.158	133
−10	−14.6851	95
−11	−17.02	66
−12	−19.5828	46
−13	−21.353	38
−14	−25.0956	27
−15	−25.7004	22
−16	−27.5269	20
−17	−33.6529	16
−18	−35.0807	14
−19	−35.5523	13
−20	−38.7657	11

5.1.3. Preasymptotics

Even if we eventually converge to a probability distribution of the kind well known and tractable, it is central that the time to convergence plays a large role.

For instance, much of the literature invokes the Central Limit Theorem to assume that fat-tailed distributions with a finite variance converge to a Gaussian under summation. If daily errors are fat-tailed, cumulative monthly errors will become Gaussian. In practice, this does not appear to hold. The data, as we saw earlier, show that economic variables do not remotely converge to the Gaussian under aggregation.

Furthermore, finiteness of variance is a necessary but highly insufficient condition. Bouchaud and Potters (2003) showed that the tails remain heavy while the body of the distribution becomes Gaussian (see Fig. 8).

5.1.4. Metrics

Much of time series work seems to be based on metrics which are in the square domain, and hence patently intractable. Define the norm L^p:

$$\left(\frac{1}{n}\sum |x|^p\right)^{\frac{1}{p}};$$

it will increase along with p. The numbers can become explosive, with rare events taking a disproportionately larger share of the metric at higher orders of p. Thus the variance/standard deviation ($p = 2$), as a measure of dispersion, will be far more unstable than mean deviation ($p = 1$). The ratio of mean-deviation to variance (Taleb, 2009) is highly unstable for economic variables. Thus, modelizations based on variance become incapacitated. More practically, this means that for distributions with a finite mean (tail exponent greater than 1), the mean deviation is more "robust".[19]

[19] A note on the weaknesses of nonparametric statistics: the mean deviation is often used as a robust, nonparametric or distribution-free statistic. It does work better than the variance, as we saw, but does not contain information on rare events, by the argument seen before. Likewise, nonparametric statistical methods (relying on the empirical frequency) will be extremely fragile to the "black swan problem", since the absence of large deviations in the past leave us in a near-total opacity about their occurrence in the future—as we saw in Fig. 4, these are confirmatory. In other words, nonparametric statistics that consist of fitting a kernel to empirical frequencies, assume, even more than other methods, that a large deviation will have a predecessor.

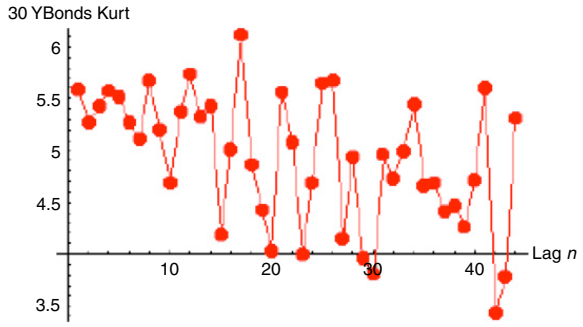


Fig. 8. Behavior of kurtosis under aggregation: we lengthen the window of changes from 1 day to 50 days. Even for variables with an infinite fourth moment, the kurtosis tends to drop under aggregation in small samples, then rise abruptly after a large observation.

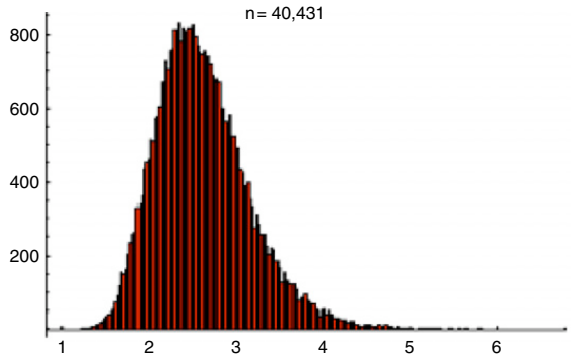


Fig. 9. Estimation error from 40 thousand economic variables.

5.1.5. *Incidence of rare events*

One common error is to believe that thickening the tails leads to an *increase* of the probability of rare events. In fact, it usually leads to a decrease of the incidence of such events, but the magnitude of the event, should it happen, will be much larger.

Take, for instance, a normally distributed random variable. The probability of exceeding 1 standard deviation is about 16%. Observed returns in the markets, with a higher kurtosis, present a lower probability of exceeding the same threshold, around 7%–10%, but the depth of the excursions is greater.

5.1.6. *Calibration errors and fat tails*

One does not need to accept power laws to use them. A convincing argument is that if we don't know what a "typical" event is, fractal power laws are the most effective way to *discuss* the extremes mathematically. It does not mean that the real world generator is actually a power law—it means that we don't understand the structure of the external events it delivers and need a tool of analysis. Also, fractals simplify the mathematical discussions because all you need to do is to perturbate one parameter, here the α, and it increases or decreases the role of the rare event in the total properties.

Say, for instance, that, in an analysis, you move α from 2.3 to 2 for data in the publishing business; the sales of books in excess of 1 million copies would triple! This method is akin to generating combinations of scenarios with series of probabilities and series of payoffs, fattening the tail at each time.

The following argument will help illustrate the general problem with forecasting under fat tails. Now the problem: *Parametrizing a power law lends itself to extremely large estimation errors* (since heavy tails have inverse problems). Small changes in the α main parameter used by power laws lead to extremely large effects in the tails.

And we don't observe the α—an uncertainty that comes from the measurement error. Fig. 9 shows more than 40 thousand computations of the tail exponent α from different samples of different economic variables (data for which it is impossible to refute fractal power laws). We clearly have problems figuring out what the α is: our results are marred by errors. The mean absolute error in the measurement of the tail exponent is in excess of 1 (i.e. between $\alpha = 2$ and $\alpha = 3$). Numerous papers in econophysics found an "average" alpha between 2 and 3—but if you process the >20 million pieces of data analyzed in the literature, you find that the variations between single variables are extremely significant.[20]

Now this mean error has massive consequences. Fig. 10 shows the effect: the expected value of your losses in excess of a certain amount (called the "shortfall") is multiplied by >10 from a small change in the α that is less than its mean error.[21]

[20] One aspect of this inverse problem is even pervasive in Monte Carlo experiments (much better behaved than the real world), see Weron (2001).

[21] Note that the literature on extreme value theory (Embrechts, Klüppelberg, & Mikosch, 1997) does not solve much of the problem, as the calibration errors stay the same. The argument about calibration we saw earlier makes the values depend on the unknowable tail exponent. This calibration problem explains how Extreme Value Theory works better on computers than in the real world (and has failed completely in the economic crisis of 2008–2009).

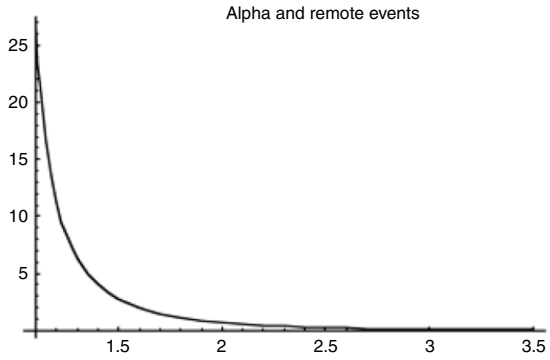


Fig. 10. The value of the expected shortfall (expected losses in excess of a certain threshold) in response to changes in the tail exponent α. We can see it explode by an order of magnitude.

6. The map

First quadrant: Simple binary decisions, under type-1 distributions: forecasting is safe. These situations are, unfortunately, more common in laboratories and games than in real life. We rarely observe these in payoffs in economic decision making. Examples: some medical decisions, casino bets, prediction markets.

Second quadrant: Complex decisions under type-1 distributions: Statistical methods may work satisfactorily, though there are some risks. True, thin-tails may not be a panacea, owing to preasymptotics, lack of independence, and model error. There are clearly problems there, but these have been addressed extensively in the literature (see Freedman, 2007).

Third quadrant: Simple decisions, under type-2 distributions: there is little harm in being wrong—the tails do not impact the payoffs.

Fourth quadrant: Complex decisions under type-2 distributions: this is where the problem resides. We need to avoid the prediction of remote payoffs—though not necessarily ordinary ones. Payoffs from remote parts of the distribution are more difficult to predict than closer parts.

A general principle is that, while in the first three quadrants you can use *the best* model you can find, this is dangerous in the fourth quadrant: no model should be better than just any model. So the idea is to exit the fourth quadrant.

The recommendation is to move into the third quadrant—it is not possible to change the distribution; but it is possible to change the payoff, as will be discussed in Section 7 (see Table 5).

The subtlety is that, while we have a poor idea about the expectation in the 4th quadrant, exposures to rare events are not symmetric.

7. Decision-making and forecasting in the fourth quadrant

7.1. Solutions by changing the payoff

Finally, the main idea proposed in this paper is to endogenize decisions, i.e., escape the 4th quadrant whenever possible by changing the payoff in reaction to the high degree of unpredictability and the harm it causes. How?

Just consider that the property of "atypicality" of the moves can be compensated by truncating the payoffs, thus creating an organic "worst case" scenario that is resistant to forecast errors. Recall that a binary payoff is insensitive to fat tails precisely because above a certain level, the domain of integration, changes in probabilities do not impact the payoff. So making the payoff no longer open-ended mitigates the problems, thus making it more tractable mathematically.

A way to express it using moments: all moments of the distribution become finite in the absence of open-ended payoffs, by putting a floor L below which $f(x) = 0$, as well a ceiling H. Just consider that if you are integrating payoffs in a finite, rather than an open-ended domain, i.e. between L and H, respectively, *the tails of the distributions outside that domain no longer matter*. Thus the domain of integration becomes the domain of payoff.

$$\lambda(x) = \int_L^H f(x)\, p(x) \mathrm{d}x.$$

With an investment portfolio, for instance, it is possible to "put a floor" on the payoff using insurance, or, even better, by changing the allocation. Insurance products are tailored with a maximum payoff; catastrophe insurance products are also set with a "cap", though the cap might be high enough to allow for a dependence on the error of the distribution.[22]

[22] Insurance companies might cap the payoff of a single claim, but a collection of capped claims might represent some problems, as the maximum loss becomes so large as to be almost undistinguishable from that with an uncapped payoff.

Table 5
The four quadrants.

	Simple payoffs	Complex payoffs
Distribution 1 ("thin tailed")	First quadrant: Extremely safe	Second quadrant: Safe
Distribution 2 (no or unknown characteristic scale)	Third quadrant: Safe	Fourth quadrant: Dangers[a]

[a] The dangers are limited to exposures in the negative domain (i.e., adverse payoffs). Some exposures, we will see, can only be "positive".

7.1.1. The effect of skewness

We omitted earlier to discuss asymmetry in either the payoff or the distribution. Clearly, the fourth quadrant can present either left or right skewness. If we suspect right-skewness, the true mean is more likely to be underestimated by the measurement of past realizations, and the total potential is likewise poorly gauged. A biotech company (usually) faces positive uncertainty, a bank faces almost exclusively negative shocks.

More significantly, by raising the L (the lower bound), one can easily produce positive skewness, with a set floor for potential adverse outcomes and open upside. For instance, what Taleb (2007a) calls a "barbell" investment strategy consists of allocating a high portion of a portfolio to T-Bills (or equivalent), say α, with $0 < \alpha < 1$, and a small portion $(1 - \alpha)$ to high-variance securities. While the total portfolio has medium variance, $L = (1 - \alpha)$ times the face value invested, another portfolio of the same variance might lose 100%.

7.1.2. Convex and concave to error

If a source of uncertainty can offer more benefits than a potential harm, then there may be gains from it—which we label "convex" or "concave".

More generally, we can be concave to model error if the payoff from the error (obtained by changing the tails of the distribution) has a negative second derivative with respect to the change in the tails, or is negatively skewed (like the payoff of a short option). It will be convex if the payoff is positively skewed (like the payoff of a long option).

7.1.3. The effect of leverage in operations and investment

Leveraging in finance has the effect of increasing concavity to model error. As we will see, it is exactly the opposite of redundancy—it causes payoffs to increase, but at the costs of an absorbing barrier should there be an extreme event that exceeds the allowance made in the risk measurement. Redundancy, on the other hand, is the equivalent of de-leveraging, i.e. by having more idle "inefficient" capital on the side. But a a second look at such funds can reveal that there may be a direct expected value from being able to benefit from opportunities in the event of asset deflation, and hence "idle" capital needs to be analyzed as an option.

7.2. Solutions by mitigating forecasting errors

7.2.1. Optimization vs. redundancy

The optimization paradigm of the economics literature meets some problems in the fourth quadrant: what if we have a consequential forecasting error? Aside from the issue that the economic agent is optimizing on the future states of the world, with a given probability distribution, nowhere[23] have the equations taken into account the possibility of a large deviation that would allow *not* optimizing consumption and having idle capital. Also, the psychological literature on well-being (Kahneman, 1999) shows an extremely concave utility function of income—if one spends such income. But if one hides it under the mattress, one will be less vulnerable to an extreme event. So there is an enhanced survival probability for those who have additional margin.

While economics have been mired in conventional linear analysis, stochastic optimization with Bellman-style equations that fall into the category Type-1, a different point of view is provided by complex systems analysis. One of the central attributes of complex systems is redundancy (May, Levin, & Sugihara, 2008).

[23] See Merton (1992) for a discussion of the general consumption Capital Asset Pricing Market.

Biological systems — those that have survived millions of years — include a large share of redundancies.[24, 25] Just consider the number of double organs (lungs, kidneys, ears). This may suggest an option-theoretic analysis: redundancy is like an option. One certainly pays for it, but it may be necessary for survival. And while redundancy means similar functions used by identical organs or resources, biological systems have, in addition, recourse to "degeneracy", the possibility of one organ to perform more than one function, which is the analog of redundancy at a functional level (Edelman & Gally, 2001).

When institutions such as banks optimize, they often do not realize that a simple model error can blow through their capital (as it just did) (see Fig. 11).

Examples: In one day in August 2007, Goldman Sachs experienced 24 times the average daily transaction volume[26]—would 29 times have blown up the clearing system? Another severe instance of an extreme "spike" lies in an event of September 18, 2008, in the aftermath of the Lehman Bothers Bankruptcy. According to congress documents, only made public in February 2009:

> On Thursday (Sept 18), at 11 am the Federal Reserve noticed a tremendous draw-down of money market accounts in the US, to the tune of $550 billion[27] was being drawn out in the matter of an hour or two.
>
> If they had not done that [add liquidity], their estimation is that by 2 pm that afternoon, $5.5 trillion would have been drawn out of the money market system of the U.S., which would have collapsed the entire economy of the U.S., and within 24 h the world economy would have collapsed. It would have been the end of our economic system and our political system as we know it.[28]

For naive economics, the best way to effectively reduce costs is to minimize redundancy, and hence avoiding the option premium of insurance. Indeed,

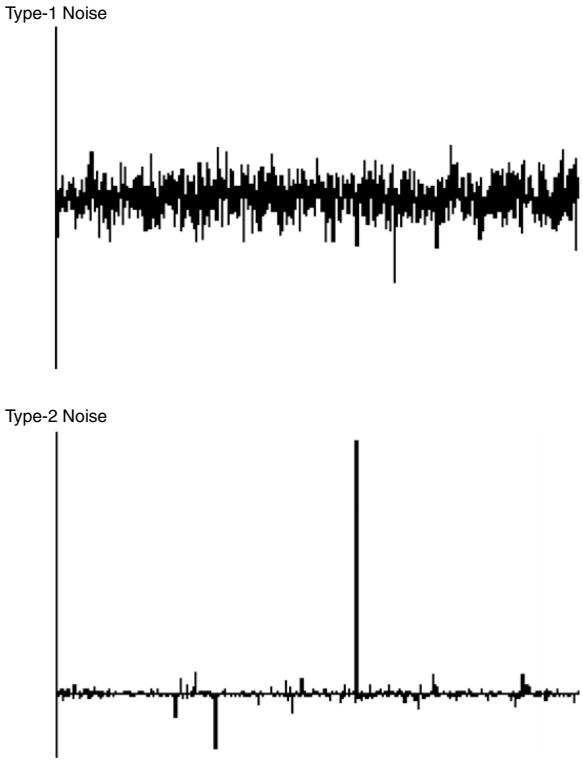


Fig. 11. Comparison between Gaussian-style noise and Type-2 noise with extreme spikes—which necessitates more redundancy (or insurance) than normally required. Policymakers and forecasters were not aware that complex systems tend to produce the second type of noise.

some systems tend to optimize and therefore become more fragile. Albert and Barabasi (2002) and Barabási and Albert (1999) warned (ahead of the North Eastern power outage of August 2003) how electricity grids, for example, optimize to the point of not coping with unexpected surges—which predicted the possibility of a blackout of the magnitude of the one that took place in the North Eastern U.S. in August 2003. We cannot discuss "flat earth" globalization without realizing that it is overoptimized to the point of maximal vulnerability.

7.2.2. *Time and sample size*

It takes much, much longer for a fat-tailed time series to reveal its properties—in fact, many can, in short episodes, masquerade as thin-tailed. At the worst, we don't know how long it would take to know. But we can have a pretty clear idea whether *organically*, because of the nature of the payoff, the "Black Swan" can hit on the left (losses) or on the

[24] May et al. (2008).

[25] For the scalability of biological systems, see Burlando (1993), Enquist and Niklas (2001), Harte, Kinzig, and Green (1999), Ritchie and Olff (1999) and Solé, Manrubia, Benton, Kauffman, and Bak (1999).

[26] Personal communication, Pentagon Highland Forum, April meeting, 2008.

[27] Even if the number, as is possible, is off by one order of magnitude, the consequences remain extremely severe.

[28] http://www.liveleak.com/view?i=ca2_1234032281.

right (profits). This point can be used in climatic analysis. Things that have worked for a long time are preferable—they are more likely to have reached their ergodic states.

Likewise, portfolio diversification needs to be larger, much larger than anticipated. A mean variance Markowitz-style portfolio construction fails in the real world on several accounts. Taleb (2009) shows that, even if we assume finite variance, but fat tails and an unknown variance, the process of discovery of the variance itself makes portfolio theory totally unusable. DeMiguel, Garlappi, and Uppal (2007) show that a naive $1/n$ allocation outperforms out-of-sample any form of "optimal" portfolio—compatible with the notion that fat tails (and unknown future properties from past samples) require much broader diversification than is required by modern portfolio theory.

7.2.3. The problem of moral hazard

It is optimal (both economically and psychologically) to make a series of annual bonuses betting on hidden risks in the fourth quadrant, then "blow up" (Taleb, 2004). The problem is that bonus payments are made with a higher frequency (i.e. annually) than is warranted from the statistical properties (when it takes longer to capture the statistical properties).

7.2.4. Metrics

Conventional metrics based on type 1 randomness fail to produce reliable results—while the economics literature is grounded in them. Concepts like "standard deviation" are not stable and do not measure anything in the fourth quadrant. This is also true for "linear regression" (the errors are in the fourth quadrant), "Sharpe ratio", the Markowitz optimal portfolio,[29] ANOVA, Least squares, etc. "Variance" and "standard deviation" are terms invented years ago when we had no computers. Note that from the data shown and the instability of the kurtosis, no sample will ever deliver the true variance in a reasonable time. However, note that truncating payoffs blunts the effects of the inadequacy of the metrics.

[29] The framework of Markowitz (1952), as it is built on the L^2 norm, does not stand any form of empirical or even theoretical validity, owing to the dominance of higher moment effects, even in the presence of "finite" variance, see Taleb (2009).

8. Conclusion

To conclude, we offered a method of robustifying payoffs from large deviations and making forecasts possible to perform. The extensions can be generalized to a larger notion of society's safety—for instance how we should build systems (internet, banking structure, etc.) to be impervious to random effects.

Acknowledgements

A longer literary version of the ideas of this paper was posted on the web on the EDGE website at www.edge.org, eliciting close to 600 comments and letters, which helped in the elaboration of this version. The author thanks the commentators and various reviewers, and Yossi Vardi for the material on the events of Sept 18, 2008.

References

Albert, R., & Barabasi, A.-L. (2002). Statistical mechanics of complex network. *Review of Modern Physics*, *74*, 47–97.

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*, 378–382.

Amaral, L. A. N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., et al. (1997). Scaling behavior in economics: I. Empirical results for company growth. *Journal de Physique I (France)*, *7*, 621.

Bak, P. (1996). *How nature works*. New York: Copernicus.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of 1/f noise. *Physical Review Letters*, *59*, 381–384.

Bak, P., Tang, C., & Wiesenfeld, K. (1988). Self-organized criticality. *Physical Review A*, *38*, 364–374.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–511.

Baz, J., & Chacko, G. (2004). *Financial derivatives: Pricing, applications, and mathematics*. Cambridge University Press.

Bouchaud, J.-P., & Potters, M. (2003). *Theory of financial risks and derivatives pricing, from statistical physics to risk management* (2nd ed.). Cambridge University Press.

Bundt, T., & Murphy, R. P. (2006). *Are changes in macroeconomic variables normally distributed? Testing an assumption of neoclassical economics*. Preprint, NYU Economics Department.

Burlando, B. (1993). The fractal geometry of evolution. *Journal of Theoretical Biology*, *163*(2), 161–172.

DeMiguel, Garlappi, & Uppal, (2007). *Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?* Working Paper.

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, *68*, 1343–1376.

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the Nationall Academy of Sciences of the United States of America*, *98*(24), 13763–13768.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics*, *75*, 643–669.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Berlin: Spring Verlag.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom. *Econometrica*, *50*(4), 987–1007.

Enquist, B. J., & Niklas, K. J. (2001). Invariant scaling relations across tree-dominated communities. *Nature*, *410*(6829), 655–660.

Freedman, D. (2007). *Statistics* (4th ed.). W.W. Norton & Company.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003a). A theory of power law distributions in financial market fluctuations. *Nature*, *423*, 267–230.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003b). *Are stock market crashes outliers?* Mimeo 43.

Gabaix, X., Ramalho, R., & Reuter, J. (2003c). *Power laws and mutual fund dynamics*. MIT mimeo.

Gardenfors, P., & Sahlin, N. E. (1982). Unreliable probabilities, risk taking, and decision making. *Syn-these*, *53*, 361–386.

Goldstein, D. G., & Taleb, N. N. (2007). We don't quite know what we are talking about when we talk about volatility. *Journal of Portfolio Management*, *33*(4), 84–86.

Gopikrishnan, P., Meyer, M., Amaral, L., & Stanley, H. E. (1998). Inverse cubic law for the distribution of stock price variations. *European Physical Journal B*, *3*, 139–140.

Gopikrishnan, P., Plerou, V., Amaral, L., Meyer, M., & Stanley, H. E. (1999). Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, *60*, 5305–5316.

Gopikrishnan, P., Plerou, V., Gabaix, X., & Stanley, H. E. (2000). Statistical properties of share volume traded in financial markets. *Physical Review E*, *62*, R4493–R4496.

Harte, J., Kinzig, A., & Green, J. (1999). Self-similarity in the distribution and abundance of species. *Science*, *284*(5412), 334–336.

Haug, E. G. (2007). *Derivatives models on models*. New York: John Wiley & Sons.

Jorion, P. (2001). *Value-at-risk: The new benchmark for managing financial risk*. McGraw Hill.

Kahneman, D. (1999). Objective happiness. In D. Kahneman, E. Diener, & N. Schwartz (Eds.), *Well being: Foundations of hedonic psychology*. New York: Russell Sage Foundation.

Levi, I. (1986). The paradoxes of Allais and Ellsberg. *Economics and Philosophy*, *2*, 23–53.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451–476.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2-competition: A real-time judgmentally based forecasting study (with commentary). *International Journal of Forecasting*, *9*, 5–29.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mandelbrot, B. (1997). *Fractals and scaling in finance*. Springer-Verlag.

Mandelbrot, B. (2001). *Quantitative Finance*, *1*, 113–123.

Mandelbrot, B., & Taleb, N. N. (2009). Mild vs. wild randomness: Focusing on risks that matter. In F. Diebold, N. Doherty, R. Herring (Eds.), *The known, the unknown and the unknowable in financial institutions*. Princeton, NJ: Princeton University Press (in press).

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, *7*, 77–91.

May, R. M., Levin, S. A., & Sugihara, G. (2008). Complex systems: Ecology for bankers. *Nature*, *451*, 893–895.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, *3*, 125–144.

Merton, R. C. (1992). *Continuous-time finance* (revised edition). Blackwell.

Ritchie, M. E., & Olff, H. (1999). Spatial scaling laws yield a synthetic theory of biodiversity. *Nature*, *400*(6744), 557–560.

Solé, R. V., Manrubia, S. C., Benton, M., Kauffman, S., & Bak, P. (1999). Criticality and scaling in evolutionary ecology. *Trends in Ecological Evolution*, *14*(4), 156–160.

Sornette, D. (2004). *Critical phenomena in natural sciences: Chaos, fractals, self-organization and disorder: Concepts and tools* (2nd ed.). Berlin, Heidelberg: Springer.

Stanley, H. E., Amaral, L. A. N., Gopikrishnan, P., & Plerou, V. (2000). Scale invariance and universality of economic fluctuations. *Physica A*, *283*, 31–41.

Taleb, N. N. (2004). The practitioner's perspective: Bleed or blowup? Why do we prefer asymmetric payoffs? *Journal of Behavioral Finance*, *5*(1), 2–7.

Taleb, N. (2007a). *The black swan: The impact of the highly improbable*. US: Random House, Penguin (UK).

Taleb, N. N. (2007b). Black swans and the domains of statistics. *The American Statistician*, *61*(3), 198–200.

Taleb, N. N. (2009). Finiteness of variance is irrelevant in the practice of quantitative finance. *Complexity*, *14*(3), 66–76.

Weron, R. (2001). Levy-stable distributions revisited: Tail index >2 does not exclude the Levy-stable regime. *International Journal of Modern Physics*, *12*(2), 209–223.

Convexity, Robustness, and Model Error inside the "Black Swan Domain"

Nassim N Taleb

DRAFT VERSION, September 2010

This is a second technical companion to the essay *On Robustness and Fragility* in the second edition of The Black Swan (a follow up for *the Fourth Quadrant*). It makes the distinction inside the Fourth Quadrant "Black Swan Domain" between fragile an robust to model (or representational) error *on the basis of convexity*.

In addition, it introduces a simple practical heuristic to measure (as an indicator of fragility) the sensitivity of a portfolio (or balance sheet) to model error. And it sets an explicit path to conduct policy.

I. Background[1]

The central idea in *The Black Swan* is about the limits in the knowledge about of small probabilities, both empirically (interpolation) and mathematically (extrapolation)[2], and its consequence. This discussion starts from the basis of the isolation of the "Black Swan domain", called the "Fourth Quadrant"[3], a domain in which 1) there is dependence on small probability events, and 2) the incidence of these events is incomputable. The Fourth Quadrant paper cursorily mentioned that there were two types of exposures, *convex* and *concave* and that we need to "robustify"[4] though convexification. This discusses revolves around convexity biases as explaining the one-way failure of quantitative methods in social science (one-way in the sense that quantitative models in social science are worst than random: their errors go in one direction as they tends to fragilize)[5].

This note discusses the following matters not present in the literature:

- The notion of model error as a convex or concave stochastic variable.
- Why deficit forecasting errors are biased in one direction.
- Why large is fragile to errors.
- Why banks are fragile.
- Why economics as a discipline made the monstrously consequential mistake of treating estimated parameters as nonstochastic variables and why this leads to fat-tails even while using Gaussian models.
- The notion of epistemic uncertainty as embedded in model errors.
- Simple tricks to compute model error.

II. Introduction: Don't cross a river that is *on average* four feet deep

How many times have you crossed the Atlantic —with a nominal flying time of 7 hours— and arrived 1, 2, 3, or 6 hours late? Or even a couple of days late, perhaps owing to the irritability of some volcano. Now, how many times have you landed 1, 2, 3, 6 hours early? Clearly we can see that in some environments uncertainty has a one way effect: extend expected arrival time[6].

[1] This paper is slightly more technical than what I presented at the July 14, 2010 Oxford BT Lecture. I thank Bent Flyvbjerg for help. I also thank my former student and teaching assistant Asim Samiuddin (my best student ever) for his remarkable work in formatting my improvised lectures and integrating the student questions into them. Most effective have been the conversation spanning 16 years with my collaborator and advisor Raphael Douady with whom I am writing more formal mathematical papers on similar issues.

Note that academic economists and others who want to provide a critical comment on my technical work should use this paper and the Fourth Quadrant, not my writing style in *The Black Swan* unless they just want to do literary criticisms.

[2] It took a long time but it looks like I finally managed to convince people that the Black Swan is not about Fat Tails (that's the Grey Swan), but incomputability of small probability events.

[3] Taleb(2009).

[4] At the "Hard Problems in Social Science" symposium, Harvard, April 2010, I presented "what to do in the 4th Q as the hard problem".

[5] Finance professors involved in investment strategies tend to blow up from underestimation of risks in an patently nonrandom way (Taleb, 2010). This explains why.

[6] I adjust for a technicality for hair slicing probabilists, that the true expected arrival time is infinite, simply because of the very small probability of never getting there owing to a plane crash. So to be more rigorous, my expectation operator is slightly modified, we would be talking of expected delay time conditional on eventually arriving to destination.

Simply, this comes from a convexity effect. In this discussion I will integrate explicitly the results of my lifetime of work as a derivatives trader, someone who works with nonlinear payoffs, and only with second and third order (or even higher) terms, and reframe the notion of robustness proposed in the postscript essay to *The Black Swan*[7], in terms of optionality and convexity of payoffs.

Missing Effects: The study of model error is not to question whether a model is precise or not, whether or not it tracks reality; it is to ascertain that the errors from the model don't have missing higher order terms that cause severe biases *in one direction*. Here we can see that uncertainty about the world will, in expectation lead to a *longer* arrival time.

Small Probabilities: Another application explains why I spent my life making bets on unlikely events, on grounds of incompleteness of models. Assume someone tells you that the probability of an event is *0*. But you don't trust his computation. Because a probability cannot be lower than 0, even in Oxford, your expected probability should be higher, at least higher than the expected error rate in the computation of such probability. Model error increases small probabilities in a disproportionate way (and accordingly decreases large probabilities). The effect is only neutral for probabilities in the neighborhood of .5.

Convex function and Jensen's Inequality: I define a convex function as one with a positive second derivative, but this is a mathematical construct that does not translate well into practice. So, more practically, convexity over an interval Δx satisfies the following inequality:

$$\frac{1}{2}\left[f(x+\Delta x)+f(x-\Delta x)\right] > f(x)$$

or more generally a linear combination of functions of points on the horizontal axis (the *x*) is higher than the function of linear combinations[8]. A concave function is the opposite. By Jensen's inequality, if we use for function the expectation operator, then the expectation of an average will be higher than the average of expectations.

$$E\sum f(\omega_i X_i)) > \sum \omega_i E\, f(X_i))$$

For example, take a conventional die (six sides) and consider a payoff equal to the number it lands on. The expected (average) payoff is $\frac{1}{6}\sum_1^6 i = 3\frac{1}{2}$. Now consider that we get the squared payoff, $\frac{1}{6}\sum_1^6 i^2 = 91/6 = 15.1666$, while $\left(\frac{1}{6}\sum_1^6 i\right)^2 = 12\frac{1}{4}$, so, since squaring is a convex function, *the average of a square payoff* is higher than *the square of the average payoff*.[9]

III. TWO TYPES OF VARIATIONS (OR PAYOFFS)

Define the two types of payoffs for now, with a deeper mathematical discussion to come later.

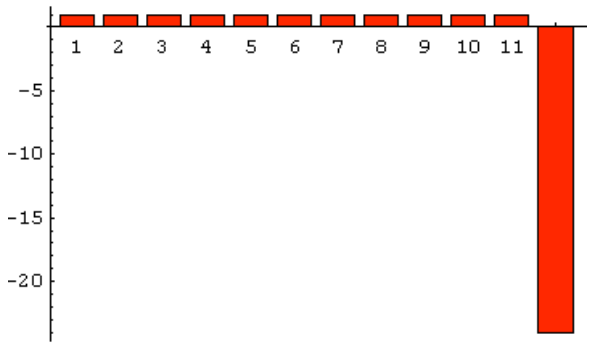


Figure 1 Concave payoff through time, with respect to a source of variation; or concave errors from left-skewed distributions.

Concave to variations and model error: when payoff is negatively skewed with respect to a given source of variations; the shocks and errors can affect a random variable in a negative way more than a positive way, as in Figure 1. It is the equivalent of being short an option somewhere, with respect of a possible parameter. As we will see, even in situations of short an option, there may be an additional source of concavity. A concave payoff (with respect to a source of variation) would have an asymmetric distribution with thicker left-tail.

A conventional measure of skewness is by taking the expectation of the third moment of the variable, x^3,which necessitates finite moments E[x^m], *m*>2, or adequacy of the L^2 norm which is not the case with economic variables. I prefer to use the symmetry of measures of shortfall, i.e. expectation below a certain threshold K, $\int_{-\infty}^{-K} x\, f(x)\, dx$ compared to $\int_K^{\infty} x\, f(x)\, dx$, *K* being a remote threshold for *x* the source of variation.

[7] Taleb (2010).

[8] We will see further down that convexity can be just local.

[9] An interesting application, according to Art de Vany who applies complexity theory to many aspects of human life, is in diet: researchers in nutrician are only concerned with "average" calories consumed, not distribution; a random and volatile feeding (feast or famine style) will less fattening than a steady one owing to concavity effects.

Convex to variations and model error: the opposite, as shown in Figure 2.

Note that whatever is convex to variations is therefore convex to model error –given the mathematical equivalence between variations and epistemic uncertainty.

Thus as an illustration of the payoffs in Figure 1, take the distribution of financial payoffs through time; a portfolio that has a floor set at K would have the downside shortfall $S = \int^{-K} x\, f(x)\, dx$ equaling 0. I have been calling such operation the "robustification" or "convexification" of the portfolio, making it immune to any parameter used in the computation of *f(x)*.

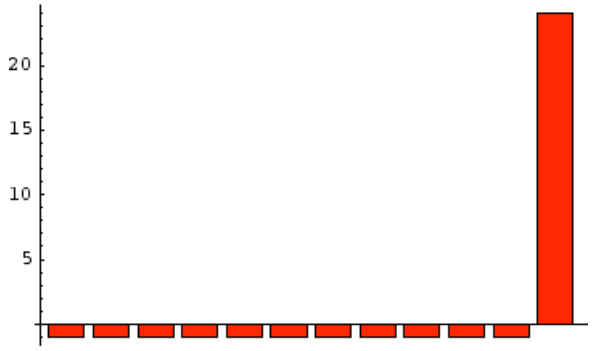


Figure 2 Convex payoff through time, or convex errors.

The typical concavity, and the one that I spent my life immersed in, is the one with respect to small probabilities, as will be discussed a bit later.

Mixed Payoffs:

As shown in Figures 3 and 4, Convexity can be local, that is, only present for a Δx of a certain size; it may become concave for larger Δx, or vice versa. So more technically convexity always needs to be attached to a certain size Δx; an infinitesimal Δx would not work in practice. Many financial institutions had the illusion of convexity, as they were so for small variations, when in fact they were not for large disturbances, "tail events".

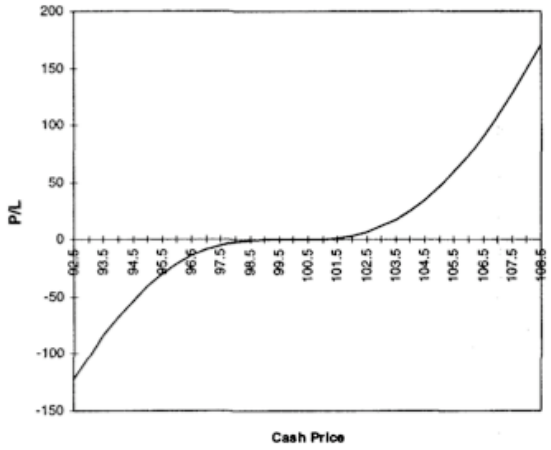


Figure 3 From Dynamic Hedging, Taleb(1997), most payoffs are mixed.

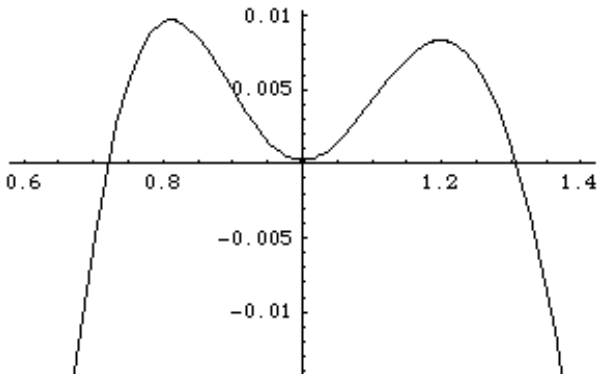


Figure 4 Convexity can be just local for a small or medium size variation, which is why measures need to be broad and fill the tails. Most banks have fallen for this trap: the banking system accumulated concavity where it was invisible.

Robustness and Convexity: As we can see from Figure 2, convex systems will mostly take small insults, for massively large gains, concave ones will appear more stable.

While linear payoffs may appear to be robust, two points:1) linear payoffs are rare, 2) we are never sure that the payoff is truly linear, particularly when it comes to hidden parameters or incompleteness of models — many nonlinear payoffs have been mistaken for linear ones.

IV. COMPARATIVE TABLEAU OF ROBUSTNESS

This document will generalize to cover fragility across all these domains using the same notion of fragility to perturbations or representational errors. This is a rapid

presentation; every entry will be explained in later sections.

FRAGILE	ROBUST
Optimized	Includes Redundancies
Short options	Long options
Model	Heuristic
Rationalism (economics modeling)	Empiricism/Reliance on time tested heuristics
Directed search	Tinkering (convex bricolage)
Nation state --centralized	City State -- decentralized
Debt	Equity
Public Debt	Private Debt
Large	Small
Agent managed	Principal managed
Monomodal	Barbell
Derivative	Primitive
Banks	Hedge funds
Kindle/Electronic files	Book
Man-designed (Craig Venter-style intervention)	Evolution
Positive heuristics	Negative heuristics
Dr John	Fat Tony

V. WHERE ERRORS ARE SIGNIFICANT

Projects: This convexity explains why model error and increased uncertainty *lengthens* rather than reduce expected projects costs and duration. Prof Bent Flyvbjerg, has shown ample empirical evidence of that effect.

Deficits: Convexity effects explain why uncertainty lengthens, doesn't shorten expected deficits. Deficits are convex to model error; you can easily see it in governments chronic underestimation of future deficits. If you run into anyone in the Obama administration, particularly Larry Summers, make them aware of it — they don't get the point.

Economic Models: Something the economics establishment has been missing is that having the right model (which is a very generous assumption), but being uncertain about the parameters will invariably lead to an increase in model error in the presence of convexity and nonlinearities.

As an illustration, say we are using a simple function $f\left(x,\overline{\alpha}\right)$, where α is supposed to be the *average expected rate* $\overline{\alpha} = \int \alpha\, \phi(\alpha)\, d\alpha$. The mere fact that α is uncertain might lead to a bias if we perturbate *from the outside* (of the integral). Accordingly, the convexity bias is easily measured as

$$\int f(\alpha,x)\, \phi(\alpha)\, d\alpha - f(\int \alpha\ \ \phi(\alpha)\, d\alpha, x)$$

As an example let us take the Bachelier-Thorp option equation (often called in the literature the Black-Scholes-Merton formula[10]), an equation I spent 90% of my adult life fiddling with. I use it in my class on model error at NYU-Poly as an ideal platform to explain the effect of assuming a parameter is deterministic when in fact it can be stochastic[11].

A call option (simplifying for absence of interest rate[12]) is the expected payoff:

$$C(S_0,K,\sigma,t) = \int_K^\infty (S-K)\, \Phi(S_0,\mu,\sigma\sqrt{t})\, dS$$

Where , where Φ is the Lognormal distribution, S_o is the initial asset price, K the strike, σ the standard deviation, and t the time to expiration. Only S is stochastic within the formula, all other parameters are considered as descending from some higher deity, or estimated *without* estimation error.

The easy way to see the bias is by producing a nested distribution for the standard deviation σ, say a Lognormal with standard deviation V then the true option price becomes, from the integration from the outside:

[10] See Haug and Taleb (2010).

[11] I am using *deterministic* here only in the sense that it is not assumed to obey a probability distribution; Paul Boghossian has signaled a different philosophical meaning to the notion of *deterministic*.

[12] The technique (which I will use in the rest of the discussions) is called a change of probability measure, to cancel the effect of the interest rate variable, by assuming it is integrated as a *numeraire*, not ignore its existence --Geman et al.

$$\int C(S_0, K, \sigma, t)\, f(\sigma)\, d\sigma$$

The convexity bias is of course well known by option operators who price out-of-the-money options, the most convex, at some premium to the initial Bachelier-Thorpe model, a relative premium that increases with the convexity of the payoff to variations in σ.

Corporate Finance: In short, corporate finance seems to be based on point projections, not distributional projections; thus if one perturbates cash flow projections, say, in the Gordon valuation model, replacing the fixed —and known— growth by continuously varying jumps (particularly under fat tails distributions), companies deemed "expensive", or those with high growth, but low earnings, would markedly increase in expected value, something the market prices heuristically but without explicit reason.

Portfolio Theory: The first defect of portfolio theory and every single theory based on "optimization" is absence of uncertainty about the source of parameters --while these theorists leave it to the econometricians to ferret out the data, not realizing the inconsistency that an unknown parameter has a stochastic character. Of course the second defect is the use of thin-tailed idealized probability distributions.

VI. Distributional Fat Tails and Convexity

I've had all my life much difficulty explaining the following two points connecting dots:

1) that Kurtosis or the fourth moment was equivalent to the variance of the variance; that the square variations around $E[x^2]$ are similar to $E[x^4]$.

2) that the variance (or any measure of dispersion) for a probability distribution maps to a measure of *ignorance*, an epistemological concept. So uncertainty of future parameters increases the variance of it; hence uncertainty about the variance raises the kurtosis, hence fat tails. Not knowing the parameter is a central problem.

The central point behind *Dynamic Hedging* (1997) is the percolation of uncertainty across all higher moments; so if one has uncertainty about the variance, with a rate of uncertainty called, say V(V) (I dubbed it "volatility of volatility"); the higher V(V), the higher the kurtosis, and the fatter the tails. Further, if V(V) had a variance called V(V(V)), the third order variance, which in turn had uncertainty, all the way down to all orders, then, simply, one ends with Paretan tails. I had never heard of Mandelbrot, or his link of Paretan tails with self-similarity, and I needed no fractal argument for that. The interesting point is that mere uncertainty about models leads immediately to the necessity to use power laws *for epistemic reasons*[13].

Another approach is through the notion of epistemic infinity. As explained in *The Black Swan*, Taleb (2010), a finite upper bound for a variable may exist, but since we do not know where it is, "how high (low)", it needs to be accordingly treated as infinite. So there may be a point where distributions become thin-tailed, and cease to be scalable, but in the absence of the knowledge about them, we need to consider them as fat tailed to infinity, hence power laws.

We already saw from the point that options increase in value, with an effect called the "volatility smile"[14].

VII. Model Errors Are Fat-Tailed Even in The Gaussian (Thin-Tailed) World[15]

First, let me show how tail exposures are extremely sensitive to model error regardless of the distribution used —something completely missed in the literature.

Let us start with the mild case of the Gaussian distribution (without even fattening the tails). Take a measure ζ of shortfall, here:

$$\zeta(\sigma, K, \mu) = \int_{-\infty}^{-K} x\, f\, x\, dx$$

where f(x) is Gaussian with mean μ and standard deviation σ.

We are not using the measure to estimate, but for higher order effects to gauge fragility —a procedure that is not affected by the reliability of the estimate.

Difference with the ordinary VAR: This measure deviates from the less rigorous ordinary Value-at-Risk (VAR) since VAR sets the K for which the probability $\int_{-\infty}^{-K} f(x)\, dx$ corresponds to a fixed percentage, say 1%. Aside from the difficulty in computation, and the limitation of the estimation of small probabilities, it

[13] Typical derivations of power laws are: hierarchies (Cantor sets) multiplicative processes, including preferential attachment /cumulative advantage (Zipf, Simon), entropy (Mandelbrot), dimentional constraints, critical points, etc. But I have never seen the epistemic issue ever presented in spite of his dominance of an operator's day to day activity.

[14] By the Breeden-Litzenberger argument, we can see that option prices produce risk-neutral probability distributions for the underlying assets, so we can look at the problem in the inverse direction.

[15] This method was proposed to the staff of the Bank of England on Jan 19 2007 as an indicator of robustness for a portfolio. I do not believe that anything was done on that.

severely ignores fat-tail effects of the expected loss below the threshold K. Furthermore it cannot be used for the estimation of model fragility.

Now take the function γ showing the relative convexity multiplier from changes in σ for a total uncertainty δ (a δ=. 25 means σ can be 25 % lower or 25% higher; a γ=1 is no effect, a γ=2 is the doubling the shortfall). With δ in [0,1[, and assuming for simplicity μ=0,

$$\gamma(K\_, \delta\_, \sigma\_) = \frac{\zeta((1-\delta)\,\sigma, K, 0) + \zeta((\delta+1)\,\sigma, K, 0)}{2\,\zeta(\sigma, K, 0)}$$

which yields to a closed form solution

$$\frac{1}{2}\, e^{\frac{K^2\left(\delta^4-4\,\delta^2-1\right)}{2\,(\delta^2-1)^2\,\sigma^2}} \left(e^{\frac{K^2}{2\,(\delta-1)^2\,\sigma^2}\,(\delta+1)} - e^{\frac{K^2}{2\,(\delta+1)^2\,\sigma^2}\,(\delta-1)} \right)$$

The shocking result is that for 10 standard deviations (that is, routine events), a 25% uncertainty about σ leads to a multiplication of the mass in the tail, causing the underestimation of the risk by a factor of 10^7. I wonder why those using methods such as Value at Risk (VAR) can be so irresponsibly blind!

Table 1: Underestimation of shortfall in excess of K from relative perturbations of 25% up or down with the parameter σ in a simple Gaussian world

K, in Standard deviations	Underestimation of shortfall
0	0
1	0
2	0.36
3	2.16
4	10.13
5	55.26
6	406
7	4,230
8	62,942
9	10^6
10	$4\ 10^7$

The worrisome fact is that a perturbation in σ extends well into the tail of the distribution in a convex way; a portfolio that is sensitive to the tails would explode. That is, we are still here in the Gaussian world! Such explosive uncertainty isn't the result of fat tails in the distribution, merely small imprecision about a future parameter. It is just epistemic! So those who use these models while admitting parameters uncertainty are necessarily committing a severe inconsistency[16] [17].

Of course, uncertainty explodes even more when we replicate conditions of the nonGausian real world upon perturbating tail exponents, see Taleb (2009).

VIII. Less is More: A Heuristic to Measure Model Errors with Simple Perturbations

Because of the sensitivity of models to tail errors, one can detect fragility by perturbating the tails. In general, most of the sensitivity to model error in a portfolio can be captured with the following procedure I've been using for a long time on portfolios containing nonlinear securities.

First step, calculate the expected Shortfall ζ at one σ (which is usually done by bank risk management using the same tools to compute the VAR[18]). Then perturbate a $\Delta\sigma$ at different levels (10%, 25%, 50%) to capture the higher moment effects; a portfolio that experiences variations will be sensitive to model error; but we will not know whether it is robust or fragile.

Second step, compare the performance at $+\Delta\sigma$ and $-\Delta\sigma$ for detection of convexity effects: if profits exceed losses for equivalent $\Delta\sigma$, then the portfolio is convex and robust; otherwise it is deemed fragile.

One limitation is that this only reveals the sensitivity up to the 4th moment; not higher ones, so a portfolio containing very remote payoffs might not react for small $\Delta\sigma$, only larger ones (as we said, convexity is local). For that, the remedy is to redo it for larger and larger $\Delta\sigma$, or, more difficult, have recourse to power laws by varying the α exponent (this would fill the tail all the way to the asymptote).

This method is for dimension 1; it can be generalized for larger dimensions as one needs to perturbate the covariance matrix Σ, without violating the positive-definite character (there are many techniques from

[16] A conversation with Paul Boghossian convinced me that philosophers need to figure out *a priori* what others need empiricism for, merely by reasoning. This argument just outlined is entirely an armchair one, does not even question the mismatch of the formula to the real world or the choice of probability; it just establishes an inconsistency from within the use of such models if the operator does not consider that the parameters descended from some unquestionable deity.

[17] This, along with the other arguments in Taleb (2010) further shows the defects of the notion of "Knightian uncertainty", since *all tails* are uncertain under the slightest perturbation.

[18] The problem of the raw VAR is probabilistic: it does not fill-in the tail.

decomposition techniques in which one can perturbate the principal components or the factors).

IX. Why Large is Concave, hence Fragile, The Case of Squeezes

The Notion of Squeeze: Squeezes are situations in which an operator is obligated to perform an action regardless of price, or with little sensitivity to price. It can cause a concave payoff since price sensitivity is low given the necessity of the action. Say a person needs water or some irreplaceable substance; there are no choices and no substitutes. He will drive the price upwards as a "price for immediacy"[19] [20].

There have been many theories of why size is ugly (or small is beautiful), but these theories are not based on statistical notions and squeezes, the distributions of shocks from the environment, rather on qualitative matters or organizational theories in management characteristically lacking in scientific firmness. Even in biology, the problem has been missed completely. For instance one can argue the absence of land mammals larger than the elephant, but on some theory of ratios and physical limitations; but they don't explain absence of much larger animals; these biological limits are above the actual size we witness. My point here is that the environment delivers resources stochastically, with fragility to squeezes —an elephant needs more water than a mouse, and would, figuratively, pay up for it.

Naive optimization may lead us to believe in economies of scale –since it ignores the stochastic structure that results from aggregation of entities, and the associated vulnerabilities and their costs. However, under a nonlinear loss function, increased exposure to rare events may have the effect of raising the costs of aggregation while giving the impression of benefits – since the costs will be borne during rare, but large-impact events. This result is general; it holds not just for economic systems, but for biological and mechanical ones as well.

Hidden Risks: Define hidden risks as an unanticipated or unknown exposure to a certain stochastic variable that elicits immediate mitigation. These stochastic shocks can be called "Black Swan" effects, as they are not part of the common risks foreseen by the institution or the entity involved. These can be hidden risks by rogue traders, miscalculation of risk positions discovered , or booking errors. An "unintended position" is a hidden risk from the activities of, say, a rogue trader that escapes the detection by the bank officials, and needs to be liquidated as it makes the total risk larger than allowed by the capital of the institution. This can be later generalized to any form of unintentional risk –errors commonly known in the business as "long v/s long" or "short v/s short" – positions that were carried on the books with a wrong sign and constitute the nightmare for operational risk. The vicious aspect of these "unintended positions" is that the sign (long or short) does not matter; it is necessary to reduce that risk unconditionally. Hence a squeeze.

Companies get larger through mergers and industries become concentrated, assuming the notion of "economies of scale", and computing the savings from the cost reductions and such benefits of scale. However, this does not take into account the effect of an increase of risks of blowups –in fact, under any form of loss or error aversion, and *concave execution costs*, the gains from the increase in size should show a steady improvement in performance, punctuated with large and more losses, with a severe increase in negative skewness.

Consider a recent event, known as the Kerviel Affair, which we simplify as follows. Société Génerale lost close to $7 billion, around $6 billion of which came mostly from the liquidation costs of the positions of Jerome Kerviel, a rogue trader, in amounts around $65 billion (mostly in equity indices). The liquidation caused the collapse of world markets by close to 12%. Indeed we stress that the losses of $7 billion did not arise from the risks but from the loss aversion and the fact that execution costs rise *per unit*.

Simple Example –Simplification of The Kerviel Case

Consider the following two idealized situations.

> **Situation 1**: there are 10 banks with a possible rogue trader hiding 6.5 billions, and probability p for such an event for every bank over one year. The liquidation costs for $6.5 billion are negligible. There are expected to be 10 p such events but with total costs of no major consequence.
>
> **Situation 2**: One large bank 10 times the size, similar to the more efficient Société Génerale, with the same probability p, a larger hidden position of $65 billion. It is expected that there will be p such events, but with $6.5 losses per event. Total expected losses are p $6.5 per time unit –lumpier but deeper and with a worse expectation.

We generalize next by assuming that the hidden positions (in absolute value) are power-law distributed and can take any positive value rather than a simple $6.5 or $65 billion. Further we generalize from the idea of hidden position of a rogue trader to hidden excess or deficit in inventory that necessitates action, an "unintended exposure".

[19] Taleb and Tapiero (2010)

[20] For the notion of price for immediacy, Grossman & Miller (1988)

General Mathematical Derivations: Our random variable X is the "unintended exposure". Assume the size of this unintended position is proportional to the capitalization of the institution –for smaller entities engage in smaller transactions than larger ones. So we are considering the splitting of the risk across N companies, with maximal concentration at N=1[21].

Probability Distribution: We use for probability distribution the variable of all unintended risk $\sum X_i$ where X_i are independent random variables, simply scaled as $X_i = X/N$. With k the tail amplitude and α the tail exponent,

$$\pi(k, \alpha, X) = \alpha\, k^{\alpha}\, x^{-1-\alpha}$$

The N-convoluted Pareto distribution for the unintended total position N $\sum X_i$:

$$\pi(k/N, \alpha, X)_N$$

where N is the number of convolutions for the distribution. The mean of the distribution, invariant with respect to N, is $\alpha\, k\, /(\alpha\text{-}1)$.

Losses From Squeeze: For the loss function, take $C[X] = -b\, X^{\beta}$, where squeezing costs is a convex function of X —the larger X, the more one needs to pay up for it.

Assume for simplicity b=1. We take 4 scenarios that should produce various levels of convexity: β= 1 (the linear case, in which we would expect that the total losses would be invariant with N), β= 2,3,4,5 the various levels of concavity.

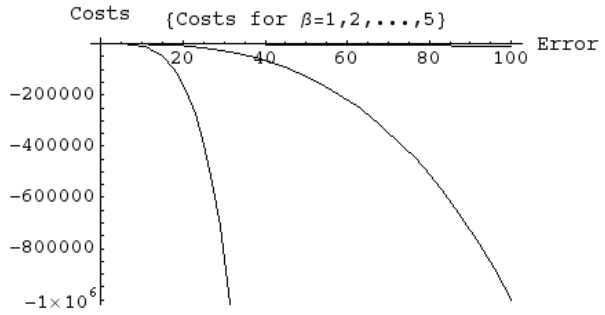


Figure 5- Various convex loss functions of increasing convexity: $-b\, x^{\beta}$ for b=1, a=2,...,5

Resulting distribution of losses:

Change of stochastic variable: the loss y=C[X] has for distribution:

$\pi[C^{-1}[x]]/C'[C^{-1}[x]]$

It follows a Pareto Distribution with tail amplitude k^{β} and tail exponent $\alpha/\boldsymbol{\beta}$

$$L_1(Y) = \frac{\alpha}{\beta} K^{\alpha} Y^{-1-\alpha/\beta}$$

which has for mean

$$\frac{k^{\beta}\alpha}{\alpha-\beta}$$

For the Sum: Under convolution of the probability distribution, in the tails, we end up with asymptotic tail amplitude $N\ (k/N)^{\alpha}$, (Bouchaud and Potters, 2003, section 2.22).

For the convoluted sum of N firms, the asymptotic distribution becomes:

$$L_N(Y) = N \frac{\alpha}{\beta}\left(\frac{K}{N}\right)^{\alpha} Y^{-1-\alpha/\beta}$$

with mean (owing to additivity):

$$M(\alpha,\beta,k,N) = \frac{N\left(\frac{k}{N}\right)^{\beta}\alpha}{\alpha-\beta}$$

Next, we check the ratio of losses in the tails for different values of the ratio of β over α

$$\frac{M(\alpha=3,\beta/\alpha,k,N=1)}{M(\alpha=3,\beta/\alpha,k,N=10)}$$

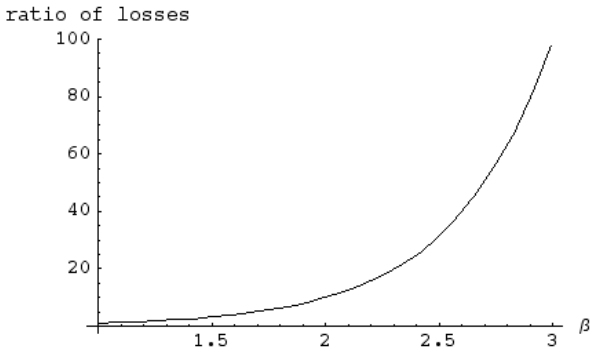


Figure 6 ratio of losses for N=1 entity/ Losses for N=10 entities as β increases. As β reaches α, the expectation of the losses becomes infinite.

[21] The limiting case N=1 corresponds to a mega-large institution commonly known as "government".

Squeezes and Redundancy: We can use the exact same equation for inventory management $C[X]= -b X^{\beta}$ and assume X is the difference between total target inventory, and needed inventory. The convexity of the slope shows how excess inventory, or, in general, whatever lowers squeezability constitutes an insurance.

Price of Convexity: Convexity is priced, in the L^2 norm, from a result of the stochastic differential equation, $\frac{\partial f}{\partial t} = -\frac{1}{2}\frac{\partial^2 f}{\partial X^2}$, where the first derivative is "time decay" or "premium erosion", and the second the convexity effect. But more practically it can priced probabilistically by summing up payoffs.

X. How Do People Sell Left Tails?

1) Outside finance:

- politics
- managing large organizations under an agency problem (steady one-way bonus)
- any job in which performance is cosmetically evaluated with potential hidden tail risks
- people worried about their reputation of "steady earners"
- mismatch between bonus frequency and time to blowup.

2) Examples of directly negatively skewed bets in finance:

Loans and Credit-Related Instruments: You lend to an entity at a rate higher than the risk-free one prevailing in the economy. You have a high probability to earn the entire interest amount, except, of course in the event of default where you may lose (depending on the recovery rate) around half your investment. The lower the risk of default, the more asymmetric the payoff. The same applies to investments in high yielding currencies that are pegged to a more stable one (say the Argentine peso to the dollar) but occasionally experience a sharp devaluation.

Derivative instruments. A trader sells a contingent claim. If the option is out-of-the-money the payoff stream for such strategy is frequent profits, infrequent large losses, in proportion to how far out of the money the option is. It is easy to see in the volumes that most traded options are out-of-the-money[22]. Note that a "delta hedged" such strategy does not significantly mitigate such asymmetry, since the mitigation of such risk of large losses implies continuous adjustment of the position, a matter that fails with discontinuous jumps in the price of the underlying security. A seller of an out-of-the money option can make her profit as frequent as she wishes, possibly 99% of the time by, say selling on a monthly basis options estimated by the market to expire worthless 99% of the time.

Arbitrage. There are classes of arbitrage operations such as: 1) "merger arbitrage" in which the operator engages in betting that the merger will take place at a given probability and loses if the merger is cancelled (the opposite is called a "Chinese"). These trades generally have the long odds against the merger. 2) "Convergence trading" where a high yielding security is owned and an equivalent one is shorted thinking that they converge to each others, which tends to happen except in rare circumstances.

The hedge funds boom caused a proliferation of packaged instruments of some opacity that engage in a variety of the above strategies –ones that do let themselves be revealed through naive statistical observation.

2) Example of comparatively skewed bets:

Covered Calls Writing: Investors have long engaged in the "covered write" strategies in which the operator sells an option against his portfolio which increases the probability of a profit in return for a reduction of the upside. There is an abundant empirical literature on covered writes (see Board, Sutcliffe and Patrinos, 2000, for a review, and Whaley, 2002 for a recent utility-based explanation) in which fund managers find gains in utility from capping payoffs as the marginal utility of gains decreases at a higher asset price. Indeed the fact that individual investors sell options at cheaper than their actuarial value can only be explained by the utility effect. As to a mutual fund manager, doing such "covered writing" against her portfolio increases the probability of beating the index in the short run, but subjects her to long term underperformance as she will give back such outperformance during large rallies.

XI. Moral Hazard & Hidden Left Tail

Why are we suckers for hidden left tails exposures? The combination of moral hazard and psychological confusion about statistical properties from small sample, two effects: crooks of randomness and fools of randomness[23].

Taleb (2004a, 2004b) presented the interplay of psychological issues related to size, to the properties of a Left-skewed Payoff stream[24]:

Property 1: Camouflage of the mean and variance.

[22] See Wilmott(1998) and Taleb (1997) for a discussion of dynamic hedging properties for an option seller.

[23] I owe the metaphor crooks of randomness to Nicolas Tabardel.

[24] John Kay and others call this generously a "Taleb Distribution".

The true mean of the payoff is different from the median, in proportion to the skewness of the bet. A typical return will, say, be higher than the expected return. It is consequently easier for the observer of the process to be fooled by the true mean particularly if he observes the returns without much ideas about the nature of the underlying generator. But things are worse for the variance as most of the time it we be lower than the true one (intuitively if a shock happens 1% of the time then the observed variance over a time window will decrease between realizations then sharply jump after the shock).

Property 2: Sufficiency of sample size.

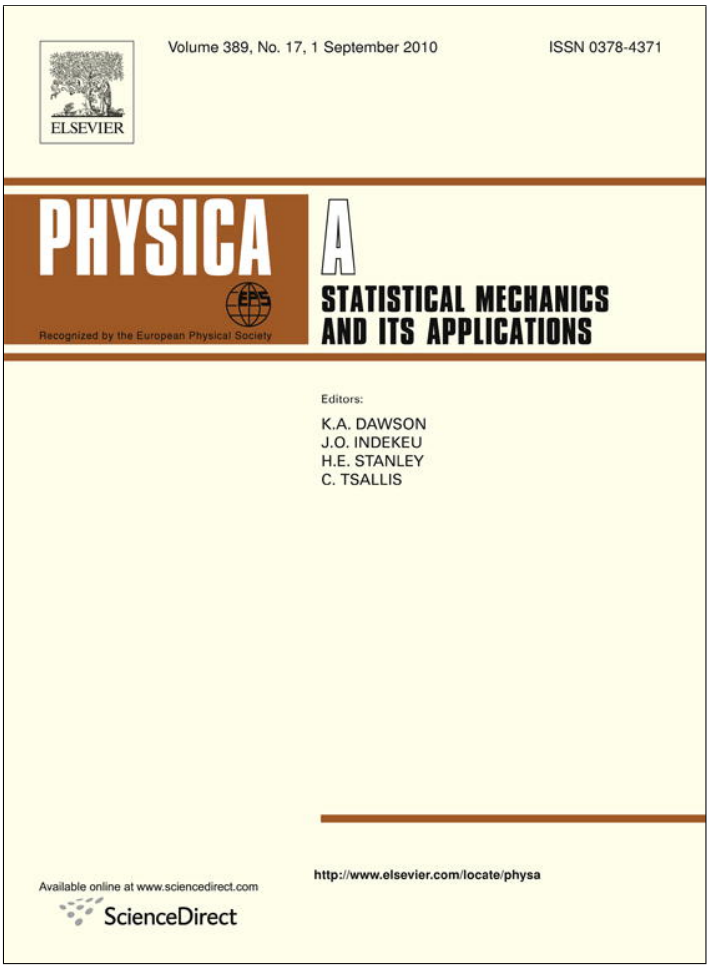
It takes a considerably longer sample to observe the properties under a skewed process than otherwise. Take a bet with 99% probability of making *G* and 1% probability of losing *L*; 99% of the time the properties will not reveal themselves –and when they do it is always a little late as the decision was made before. Contrast that with a symmetric bet where the properties converge rather rapidly.

Property 3: The smooth ride effect.

As we said the observed variance of the process is generally lower than the true variance most of the time. This means, simply, that the more skewness, the more the process will generate steady returns with smooth ride attributes, concentrating the variance in a brief period, the brevity of which is proportional to the variance. In another word, an investor has, without a decrease in risk, a more comfortable ride most of the time, with an occasional crash.

XII. CONCLUSION

This document sets the basic framework for identifying robustness and fragility. Taken to its logical extensions, it should generate measures of comparative fragility for the measurement of tail events.



Physica A 389 (2010) 3503–3507

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Risk externalities and too big to fail

Nassim N. Taleb, Charles S. Tapiero *

Department of Finance and Risk Engineering, The Research Center for Risk Engineering, New York University Polytechnic Institute, 11201 New York, United States

ARTICLE INFO

Article history:
Received 14 May 2009
Received in revised form 1 November 2009
Available online 20 March 2010

Keywords:
Risk management
Tail risks
Corporate finance
Quantitative finance

ABSTRACT

This paper establishes the case for a fallacy of economies of scale in large aggregate institutions and the effects of scale risks. The problem of rogue trading and excessive risk taking is taken as a case example. Assuming (conservatively) that a firm exposure and losses are limited to its capital while external losses are unbounded, we establish a condition for a firm not to be allowed to be too big to fail. In such a case, the expected external losses second derivative with respect to the firm capital at risk is positive. Examples and analytical results are obtained based on simplifying assumptions and focusing exclusively on the risk externalities that firms too big to fail can have.

1. Introduction

"Too Big to Fail" is a dilemma that has plagued economists, policy makers and the public at large. The lure for "size" embedded in "economies of scale" and Adam Smith factories have important risk consequences that have not always been assessed at their proper costs or properly defined. The presumption that the manufacturing sector has convex production functions has fueled the growth of enterprises to sizes that may be both too large to manage, and have losses too large to sustain. This is the case for industrial giants such as GM that have grown into a complex and diversified global enterprise that have accumulated costs too large to maintain. This is also the case for banks that are strategically focused and bear risks that are often ignored. Banks draw their legal rights from a common trust, to manage the supply and the management of money for their own and the common good. The consequences of such failures, overflowing into the commons, far outstrip their direct losses. When banks are perceived too big to fail, they have a greater propensity to assume risks, to "rule the commons", price their services unrelated to their costs or quality and exercise unduly their market power.

Size may lead such firms to assume leverage risks that are unsustainable. This is the case when banks' bonuses are indexed to short term performance, at the expense of hard to quantify risk externalities. These risks arise when all costs and benefits are not incorporated by the market. Externality is therefore an expression of market failure. For banks that are too big to fail, these risk externalities are acute. For example, Frank Rich (The New York Times, Goldman Can Spare You a Dime, October 18, 2009) has called attention to the fact that "Wall Street, not Main Street, still rules Washington". Similarly, Rolfe Winkler (Reuters) pointed out that "Main Street still owns much of the risk while Wall Street gets all the profits". Further, a recent study by the National Academy of Sciences has pointed out to extremely large hidden costs to the energy industry—costs that are not accounted for by the energy industry, but assumed by the public at large.

Banks and Central Banks rather than Governments, are entrusted to manage responsibly the monetary policy—not to be used for their own and selfish needs, not to rule the commons, but to the betterment of society and the supply of the credit needed for a proper functioning of financial markets. A violation of this trust has contributed to a financial meltdown and to the large consequences borne by the public at large. In this case, "too big to fail banks" have contributed to an immense negative externality—costs experienced by the public at large. Thus, banks have been endowed with this trust without being

* Corresponding author. Tel.: +1 212 9881421; fax: +1 717 2603653.
E-mail address: ctapiero@poly.edu (C.S. Tapiero).

0378-4371/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2010.03.014

party to the transactions that have produced such a financial meltdown. If a firm's negative externalities are not compensated by their positive externalities or appropriately regulated, then the social risks can be extremely damaging. In a recent New York Times article (Sunday Business, section, October 4, 2009), Gretchen Morgenson, referring to a research paper of Dean Baker and Travis McArthur, indicated the effects of selective failures, letting selected banks grow larger and "subsidized" at a cost of over 34 Billion dollars yearly over an appreciable amount of time.

A naive optimization to size that does not recognize the nonlinearities of the risks of scale, the risks of dependence they induce and convex their risk externalities, may lead to firms which cannot be economically sustainable [1,2]. Rather, we may experience a risk of blowup. In fact, under any form of loss or error aversion, and concave execution costs, gains from an increase in size should show a steady improvement in performance, punctuated with large and more losses, with a severe increase in negative skewness [3,4].

Under a nonlinear loss function, increased exposure to rare and latent events may have the effect of raising costs of aggregation while giving the impression of benefits — since costs will be borne during rare, but large-impact events. This result is general. It holds not just for economic systems, but for biological, industrial and mechanical ones as well. For example, Fujiwara [5], using an exhaustive list of Japanese bankruptcy data in 1997 (see also Refs. [6–9,4]) pointed out to firms failure regardless of their size. Further, since the growth of firms has been fed by debt, the risk borne by large firms seems to have increased significantly—threatening both the creditor and the borrower. In fact, the growth of size through a growth of indebtedness combined with "too big to fail" risk attitudes has ushered, has contributed to a moral hazard risk, with firms assuming non-sustainable growth strategies on the one hand and important risk externalities on the other. Furthermore, when size is based on intensely networked firm (such as large "supply chains") supply chain risks (see also Refs. [10–12]) may contribute as well to the costs of maintaining such industrial and financial organizations. Saito [13] for example, while examining inter-firm networks noted that larger firms tend to have more inter-firms relationships than smaller ones and are therefore more dependent, augmenting their risks. In particular, they point out that Toyota purchases intermediate products and raw materials from a large number of firms; maintaining close relationships with numerous commercial and investment banks; with a concurrent organization based on a large number of affiliated firms. Such networks have augmented both dependence and supply chains risks. Such dependence is particularly acute in some firms where one supplier may control a critical part needed for the proper function of the whole firm. For example, a small plant in Normandie (France) with no more than a hundred employees could strike out the whole Renault complex. By the same token, a small number of traders at AIG could bring such a "too big to fail" firm to a bankrupt state. This networking growth is thus both a result and a condition for the growth to sizeable firms of scale free characteristic (see also Refs. [9,8]). Simulation experiments to that effect were conducted by Alexsiejuk and Holyst [14] while constructing a simple model of bank bankruptcies using percolation theory on a network of cooperating banks (see also Stauffer on percolation theory [15]). Their simulation have shown that sudden withdrawals from a bank can have dramatic effects on the bank stability and may force a bank into bankruptcy in a short time if it does not receive assistance from other banks. More importantly however, the bankruptcy of a simple bank can start a contagious failure of banks concluded by a systemic financial failure. As a result, too big to fail and its many associated moral hazard and risk externalities is a presumption that while driving current financial policy and protecting some financial and industrial conglomerates (with other entities facing the test of the market on their own and subsidizing such a policy), can be extremely risky for the public at large.

Size for such large entities thus matters as it provides a safety net and a guarantee by public authorities that whatever their policies, their survivability will be ascertained for the greater good and at the expense of public funding. The rationality "too big to fail" is therefore misleading, based on a fallacy that negates the risk of size and does not account for the omnipresent effects of latent, dependent and rare risks as well as their dependent moral hazard and risk externalities.

Scale is neither necessarily robust, in particular with respect to off-model risks. Under loss aversion, the gains from a merger may show a steady improvement in performance, punctuated with large losses, with severe increases in skewness. The essential question is therefore can economies of scale savings compensate their risks. Such an issue has been implicitly recognized by Obama's administration proposal in Congressional committees calling for banks to hold more capital with which to absorb losses. The bigger the bank, the higher the capital requirement should be (New York Times, July, 27, 2009, Editorial). However such regulation does not protect the "commons" from the risk externalities that banks create and the common sustains.

To assess the effects of size and their risk externalities, this paper considers a particular and simple case based on rogue traders' risks and their effects on both a firm's loss and their risk externalities. An example is used to demonstrate that rogue trading or excessive risk taking can have significant impact on a firm risk exposure and on external losses in case of failure — risks that augment significantly, the larger the size of the firm.

2. Too big to fail and hidden risks

Consider the event, known as the Kerviel affair, which we simplify as follows. Societe Generale lost close to $7 Billions dollars, $6 Billions of which came mostly from the liquidations costs of the (hidden) positions of Jerome Kerviel, a rogue trader. In addition, it contributed to external losses that we estimate something around $65 Billions, coming from the liquidation costs of other firms reacting to the meltdown. The former are risks that the bank sustained while the latter is a cost 10 times larger which points out to the systemic risk externalities. These externalities are side effect of the liquidation caused the collapse of world markets by close to 12%!! These extraordinary losses did not put in question the continuity

of Societe Generale but put an important and disproportionate strain on the financial system. This situation has generated consequential externality losses because they were signaling a lack of controls in a bank too big to fail, unable to manage its hidden risks and at the same time created a lack of confidence reverberating in the financial supply chain of Societe Generale. Uncertainty regarding the system as a whole, dominated by banks presumed too powerful and too big to fail was put in question as well. Banks, and in particular large banks, are privy to a trust to maintain the safe operation of the financial system for the betterment of the economy. When such trusts is violated, explicitly through the behavior of their managers or implicitly, by an unreasonable risk taking policy, uncertainty sets in, producing costs commensurate with the size of these firms.

Consider traders' hidden positions defined as risks that are unanticipated or of unknown exposure and resulting in stochastic shocks. These shocks can be called "Black Swan" effects, as they are not part of the common (statistical) risks foreseen by the institution or the entity involved. These shocks are assumed to be both unpredictable in a statistical sense and therefore with large variance or jump processes with important consequences that transcend the bank. These may be hidden risks by rogue traders, miscalculation of risk positions discovered or booking errors, or action taken underscoring an uncontrollable risk taking culture. An "unintended position" is thus a hidden risk from the activities of a rogue trader that escapes the detection by the bank officials, and needs to be liquidated as it makes the total risk larger than allowed by the capital of the institution while at the same time contributes to a risk overflow to the financial system. This risk can be later generalized to any form of unintentional risk – errors commonly known in the business as "long v/s long" or "short v/s short" – positions that were carried on the books with a wrong sign (and constitute the nightmare for operational risk). The vicious aspect of these "unintended positions" is that the sign (long or short) does not matter; it is necessary to reduce that risk unconditionally. Given the multiplying risk factor of large banks, these failings—even if small, assume large consequences for the common financial system they presumably "rule".

Given the nature of a hidden or speculative position, we assume that the positions (in absolute value) has a potential loss probability distribution bounded above by the firm capital (its size) W or $f(x : W) = x \in [0, W)$. In some cases, the risk exposure of such trades may be larger than the firm capital and therefore our assumption may be assumed to be a conservative one. Thus, given a firm loss due to a rogue trader or due to uncontrolled risk of its trading department, we let the total loss, including external losses be given by $g(y|x), y \in [x, \infty), y \geq 0$. As a result, the joint probability distribution of global financial and firm losses is $f(y, x) = g(y|x)f(x : W), y \in [x, \infty), 0 \leq x \leq W$. The external loss of a firm whose capital is W has thus probability and cumulative distributions:

$$g(y) = \int_0^W g(y|x)f(x : W)\mathrm{d}x \quad \text{and} \quad G(Y) = \int_0^W \int_x^Y g(y|x)f(x : W)\mathrm{d}x.$$

The effects of size on the aggregate loss are thus a compounded function of the probabilities of losses of the firm and their external costs. If a firm has a loss whose external consequences (the loss y are extremely large), then they may be deemed to be "too big to fail". "Too big to fail" entails therefore a responsibility by the firms that ought to be regulated and controlled extensively. In this context, "too big to fail" is an issue whose relevance may be measured by its risk externalities. For example, a bank "too big to fail" that assumes risks for the sake of excess short terms profits that are not sustainable is in fact misusing its charter to serve the "commons". Such banks are thus irresponsible "polluters".

3. An example

The example we consider below assumes a mixture Pareto power conditional probability distribution for all losses, including both the firm and external losses. External losses are bounded by the firm losses from below, assumed to be fractional in the hazard rate and bounded by its capital. In particular we have used a truncated Weibull distribution. Such an approach differs of course from the Copula approach that models the co-dependence of losses by the marginal distribution of each distribution. It also differs from a generalization of the Pareto distribution that accounts for a potential correlation between the firm and the external losses. Both approaches are not be applicable as external losses depend necessarily on the firm losses but not vice versa. Further, the use of fractal models, based on modeling a process volatility growth with additional parameters, is for the same reasons, not applicable. Although, a firm loss can be modeled as a truncated Wiener–Levy or fractal model, used to randomizes external losses (in conjunction with other factors, such as market liquidity and other macroeconomic variables). The example we thus consider is of course selected for simplicity and to highlight the effects of a firm size on the external losses.

Explicitly, say that:

$$g(y|x) = \frac{\gamma_x}{(x)^{-\gamma_x}}(y)^{-\gamma_x - 1}, \quad y \geq x, \qquad E(y|x) = x\left(\frac{\gamma_x}{1-\gamma_x}\right), \quad 0 < \gamma_x < 1, \frac{\partial \gamma_x}{\partial x} > 0.$$

In other words, the distribution parameter may be interpreted as the "odds" that a firm loss has on external and global losses. The larger the "odds" the larger the risk externalities. In case of the Kerviel affair, a firm loss of 7 Billion dollars had an external loss of 65 Billion dollars. In this case, the parameter equals $7\frac{\gamma_7}{1-\gamma_7} = 65 + 7$ or $\gamma_7 = 72/79 = 0.911$ and $\frac{\partial E(y|x)}{\partial \gamma_x} > 0$. By the same token since,

$$\frac{\partial E(y|x)}{\partial x} = \left(\frac{\gamma_x}{1-\gamma_x}\right) + x\left(\frac{\partial \gamma_x/\partial x}{(1-\gamma_x)^2}\right) > 0 \quad \text{and} \quad \partial \gamma_x/\partial x > 0.$$

For example, say that γ_x has a logit distribution, or:

$$\gamma_x = F(S_x) = \frac{e^{S_x}}{1+e^{S_x}} = \frac{1}{1+e^{S_x}}, \qquad \frac{\partial \gamma_x}{\partial S_x} = \frac{e^{-S_x}}{\left(1+e^{-S_x}\right)^2} > 0.$$

Then: $\frac{\gamma_x}{1-\gamma_x} = e^{S_x}$ and $E(y|x) = xe^{S_x}$ with S_x a score for a firm "too big to fail". Such a score may be defined as a function of both the loss and economic environmental conditions. A bank whose internal loss is its capital, contribute then to an expected loss of:

$$E(y|W) = We^{S_W}.$$

In other words, in case of the Kerviel affair, assuming a loss of capital of 50 Billion dollars, the external (total loss) is $E(y|50) = 514.28$ Billion dollars.

The loss probability distribution is then:

$$g(y) = \int_0^W \frac{\gamma_x}{y}\left(\frac{y}{x}\right)^{-\gamma_x} f(x:W)\mathrm{d}x \quad \text{and} \quad G(Y) = \int_0^W \left(\frac{Y}{x}\right)^{-\gamma_x} f(x:W)\mathrm{d}x - 1.$$

The probability of a loss greater than Y and its hazard rate are therefore,

$$2 - G(Y) = 2 - \int_0^W \left(\frac{Y}{x}\right)^{-\gamma_x} f(x:W)\mathrm{d}x \quad \text{and} \quad h(Y) = \frac{\int_0^W \frac{\gamma_x}{Y}\left(\frac{Y}{x}\right)^{-\gamma_x} f(x:W)\mathrm{d}x}{2 - \int_0^W \left(\frac{Y}{x}\right)^{-\gamma_x} f(x:W)\mathrm{d}x}.$$

If a firm's expected external loss is $E(y:W)$ then if $\partial E(y:W)/\partial W > 0$ and $\partial^2 E(y:W)/\partial W^2 > 0$, then "size" contributes to a nonlinear and increasing growth in external losses—losses that are risk externalities.

For demonstration purposes, say that the probability distribution $f(x:W)$ is a constrained extreme (Weibull) distribution defined by,

$$f(x:W) = \frac{f(x)}{F(W)} = \frac{c}{\zeta}\frac{\left(\frac{x}{\zeta}\right)^{c-1} e^{-(x/\zeta)^c}}{1-e^{-(W/\zeta)^c}}.$$

The loss probability distribution and its cumulative distribution function are then:

$$g(y) = \frac{c\zeta^{1-c}}{\zeta\left(1-e^{-(W/\zeta)^c}\right)}\int_0^W \gamma_x y^{-\gamma_x-1}x^{\gamma_x+c-1}e^{-(x/\zeta)^c}\mathrm{d}x \quad \text{and}$$

$$G(Y) = \frac{c\zeta^{1-c}}{\zeta\left(1-e^{-(W/\zeta)^c}\right)}\int_0^W Y^{-\gamma_x}x^{\gamma_x+c-1}e^{-(x/\zeta)^c}\mathrm{d}x - 1.$$

With expected losses:

$$\begin{aligned} E(y) &= \frac{c\zeta^{1-c}}{\zeta\left(1-e^{-(W/\zeta)^c}\right)}\int_0^W\int_x^\infty \gamma_x y^{-\gamma_x}x^{\gamma_x+c-1}e^{-(x/\zeta)^c}\mathrm{d}y\mathrm{d}x \\ &= \frac{c\zeta^{1-c}}{\zeta\left(1-e^{-(W/\zeta)^c}\right)}\int_0^W \gamma_x\frac{x^{1-\gamma_x}}{1-\gamma_x}x^{\gamma_x+c-1}e^{-(x/\zeta)^c}\mathrm{d}x = \frac{c\zeta^{1-c}}{\zeta\left(1-e^{-(W/\zeta)^c}\right)}\int_0^W \frac{\gamma_x}{1-\gamma_x}x^c e^{-(x/\zeta)^c}\mathrm{d}x. \end{aligned}$$

The effects of the firm capital size on the expected losses are thus:

$$\frac{\partial E(y)}{\partial W} = \left(c\zeta^{-c}\frac{\gamma_W}{1-\gamma_W}W - E(y)\right)\frac{1}{W^{1-c}\left(e^{(W/\zeta)^c}-1\right)} > 0 \quad \text{since } c\zeta^{-c}\frac{\gamma_W}{1-\gamma_W}W > E(y).$$

The second derivative leads to:

$$\begin{aligned} \frac{\left(1-e^{-(W/\zeta)^c}\right)^2}{W^{c-1}e^{-(W/\zeta)^c}}\frac{\partial^2 E(y)}{\partial W^2} = &\left(\begin{array}{l} c\zeta^{-c}\dfrac{\partial\gamma_W/\partial W}{(1-\gamma_W)^2}W - c\zeta^{-c}W^{c-1}e^{-(W/\zeta)^c} + \dfrac{c\zeta^{-c}\gamma_W}{(1-\gamma_W)} \\ -\left(c\zeta^{-c}\dfrac{\gamma_W}{1-\gamma_W}W - E(y)\right)\dfrac{W^{c-1}e^{-(W/\zeta)^c}}{\left(1-e^{-(W/\zeta)^c}\right)} \end{array}\right) \\ &+ \left(1-e^{-(W/\zeta)^c}\right)\left(\frac{(c-2)}{W} - \zeta^{-c}cW^{c-1}\right) \end{aligned}$$

or

$$\frac{\left(1-e^{-(W/\zeta)^c}\right)^2}{W^{c-1}e^{-(W/\zeta)^c}}\frac{\partial^2 E(y)}{\partial W^2} = c\zeta^{-c}\left(\frac{\gamma_W}{(1-\gamma_W)} + \frac{\partial\gamma_W/\partial W}{(1-\gamma_W)^2}W + \frac{c-2}{W}\left(1-e^{-(W/\zeta)^c}\right) - \frac{1}{W^{1-c}}\right) - \frac{\partial E(y)}{\partial W}.$$

Since

$$\frac{\partial \gamma_W/\partial W}{(1-\gamma_W)^2} W \gg \frac{c-2}{W}\left(1-\mathrm{e}^{-(W/\zeta)^c}\right) - \frac{1}{W^{1-c}}$$

The condition for a positive second derivative is:

$$c\zeta^{-c}\frac{\gamma_W}{(1-\gamma_W)}\left(1-\frac{1}{W^{-c}\left(\mathrm{e}^{(W/\zeta)^c}-1\right)}\right) + c\zeta^{-c}\frac{\partial \gamma_W/\partial W}{(1-\gamma_W)^2} + \frac{E(y)}{W^{1-c}\left(\mathrm{e}^{(W/\zeta)^c}-1\right)} > 0$$

which is guaranteed if $W^c > \zeta^c \ln(1+W^c)$

These conditions establish therefore the conditions for an accelerating loss the larger the firm—a loss that may be far larger than the firm capital loss.

4. Conclusion

The purpose of this paper was to to indicate that size matters in a nonlinear way and that the issues that pertain to managing evaluating firms that are too big to fail require a far greater awareness and a regulation of the risk externalities that these institutions represent. Firms that are "too big to fail" are "polluters" either by design when they over-leverage their financial bets or their speculative positions or when they are struck by a Black Swan. This is the case because their losses have far greater significance than their narrow well being affecting investors that had no part in their actions. In this sense, their costs are a risk externality to be confronted and regulated as such. For this reason, regulation of firms that are too big too fail, require that greater attention be given to their consequential external risks rather than application of VaR techniques to protect their internal losses. The growth of economic units large enough to integrate their external risk is of course not appropriate since the moral hazard risks resulting from their market power will be too great. Similarly, total controls, total regulation, taxation, nationalization etc. are also a poor answer to deal with risk externalities. Such actions may stifle financial innovation and technology and create disincentives to an efficient allocation of money. Coase observed that a key feature of externalities are not simply the result of one CEO or Bank, but the result of combined actions of two or more parties. In case of the financial sector, there are two parties, Banks that are "too big to fail" and the Government as a stand in for the public. Banks are entrusted rights granted by the Government and therefore any violation of the trust (and not only a loss by the bank) would justify either the removal of this trust or the takeover of the bank. A bargaining over externalities would, economically lead to Pareto efficient solutions provided that banking and public rights are fully transparent. However, the non-transparent bonuses that CEOs of large banks apply to themselves while not a factor in banks failure is a violation of the trust signaled by the incentives that banks have created to maintain the payments they distribute to themselves.

References

[1] V. Pareto, Le cours d'Economie Politique, Macmillan, London, 1896.
[2] N.N. Taleb, Errors, robustness, and the fourth quadrant, International Journal of Forecasting 25 (4) (2009) 744–759.
[3] Y. Ijiri, H.A. Simon, Skew Distributions and the Size of Business Firms, North Holland, New York, 1977.
[4] K. Okuyama, M. Takayasu, H. Takayasu, Zipf'ss law in income distribution of companies, Physica A 269 (1999) 125–131.
[5] Y. Fujiwara, Zipf law in firms bankruptcy, Physica A 337 (2004) 219–230.
[6] L.A.N. Amaral, S.V. Bulkdyrev, S.V. Havlin, H. Leschron, P. Mass, M.A. Salinger, H.E. Stanley, M.H.R. Stanley, Journal of Physics I (1997) 621.
[7] M.H.R. Stanley, L.A.N. Amaral, S.V. Bulkdyrev, S.V. Havlin, H. Leschron, P. Mass, M.A. Salinger, H.E. Stanley, Nature 397 (1996) 804.
[8] J.P. Bouchaud, M. Potters, Theory of Financial Risks and Derivatives Pricing, From Statistical Physics to Risk Management, 2nd ed., Cambridge University Press, 2003.
[9] D. Garlaschelli, S. Battiston, M. Castri, V.D.P. Servedio, G. Caldarelli, The scale free nature of market investment network, Physica A 350 (2005) 491–499.
[10] C.S. Tapiero, Consumers risk and quality control in a collaborative supply chain, European Journal of Operations Research 182 (2007) 683–694.
[11] C.S. Tapiero, Risk Finance and Financial Engineering (tentative title), 2 volume, Wiley, 2010 (forthcoming).
[12] Konstantin Kogan, Charles S. Tapiero, Supply Chain Games: Operations Management and Risk Valuation, in: Frederick Hillier (Ed.), Series in Operations Research and Management Science, Springer Verlag, 2007.
[13] Y.U. Saito, T. Watanabe, M. Iwamura, Do larger firms have more interfirm relationships, Physica A 383 (2007) 158–163.
[14] A Aleksiejuk, J.A. Holyst, A simple model of bank bankruptcies, Physica A 299 (2001) 198–204.
[15] D. Stauffer, Introduction to Percolation Theory, Taylor and Francis, London, Philadelphia, A, 1985.

J_ID: ZA5 Customer A_ID: RA-08-758 Cadmus Art: CPLX20263 Date: 7-NOVEMBER-08 Stage: I Page: 1

RESEARCH ARTICLE

Finiteness of Variance Is Irrelevant in the Practice of Quantitative Finance

NASSIM NICHOLAS TALEB
New York University Polytechnic Institute, New York, New York

Received June 16, 2008; accepted June 24, 2008

Outside the Platonic world of financial models, assuming the underlying distribution is a scalable "power law," we are unable to find a consequential difference between finite and infinite variance models—a central distinction emphasized in the econophysics literature and the financial economics tradition. Although distributions with power law tail exponents $\alpha > 2$ are held to be amenable to Gaussian tools, owing to their "finite variance," we fail to understand the difference in the application with other power laws ($1 < \alpha < 2$) held to belong to the Pareto-Lévy-Mandelbrot stable regime. The problem invalidates derivatives theory (dynamic hedging arguments) and portfolio construction based on mean-variance. This article discusses methods to deal with the implications of the point in a real world setting. © *2008 Wiley Periodicals, Inc. Complexity 00: 000–000, 2008*

Key Words: mathematical finance; derivatives; portfolio theory; complexity; power laws

1. THE FOUR PROBLEMS OF PRACTICE

This note outlines problems viewed solely from the vantage point of practitioners of quantitative finance and derivatives hedging, and the uneasy intersection of theories and practice; it aims at asking questions and finding robust and practical methods around the theoretical difficulties. Indeed, practitioners face theoretical problems and distinctions that are not visibly relevant in the course of their activities; furthermore, some central practical problems appear to have been neglected by theory. Models are Platonic: going from theory to practice appears to be a direction that is arduous to travel. In fact, the problem may be even worse: seen from a derivatives practitioner's vantage point, theory may be just fitting practice (albeit with considerable delay), rather than influence it. This article is organized around a class of such problems, those related to the effect of power laws and scalable distributions on practice. We start from the basis that we have no evidence against Mandelbrot's theory that financial and commodity markets returns obey power law distributions [1, 2(a,b,c)], (though of unknown parameters). We do not even have an

Correspondence to: Nassim Nicholas Taleb, New York University Polytechnic Institute, 6 Metro Tech Center, Brooklyn, NY 11201. (e-mail: info@TheBlackSwan.org)
This is an invited article for the Special Issue—Econophysics, Guest Editors: Martin Shubik and Eric Smith.

 Vol. 00, No. 0
DOI 10.1002/cplx.20263
Published online in Wiley InterScience
(www.interscience.wiley.com)

ID: kannanb Date: 7/11/08 Time: 22:44 Path: J:/Production/CPLX/VOL00000/080063/3B2/C2CPLX080063

argument to reject it. We therefore need to find ways to effectively deal with the consequences.

We find ourselves at the intersection of two lines of research from which to find guidance: orthodox financial theory and econophysics. Financial theory has been rather silent on power laws (while accepting some mild forms of "fat tails" though not integrating them or taking them to their logical consequences)—we will see that power laws (even with finite variance) are totally incompatible with the foundations of financial economics, both derivatives pricing and portfolio theory. As to the econophysics literature: by adopting power laws, but with artificial separation parameters, using $\alpha = 2$, it has remedied some of the deficits of financial economics but has not yet offered us help for our problems in practice.[1]

We will first identify four problems confronting practitioners related to the fat tailed distributions under treatment of common finance models: (1) problems arising from the use by financial theory, as a proxy for "fat tails," of milder forms of randomness too dependent on the Gaussian (non-power laws); (2) problems arising from the abstraction of the models and properties that only hold asymptotically; (3) problems related to the temporal independence of processes that lead to assume rapid convergence to the Gaussian basin; and (4) problems related to the calibration of scalable models and to the fitting of parameters.

Note here that we provide the heuristic attribute of a scalable distribution as one where for some "large" value of x, $P > x \sim Kx^{-\alpha}$, where $P > x$ is the "exceedant probability," the probability of exceeding x, and K is a scaling constant.[2] (Note that the same applies to the negative domain). The main property under concern here, which illustrates its scalability, is that, in the tails, $P > nx/P > x$ depends on n, rather than x.

The Criterion of Unboundedness: One critical point is answering the controversial question "is the distribution a power law?" Unlike the econophysics literature, we do not necessarily believe that the scalability holds for x reaching infinity; but, in practice, so long as we do not know where the distribution is eventually truncated, or what the upper bound for x is, we are forced operationally to use a power law. Simply as we said, we cannot safely reject Mandelbrot [1, 2(a,b,c)]. In other words, it is the uncertainty concerning such truncation that is behind our statement of scalability. It is easy to state that the distribution might be log-normal, which mimics a power law for a certain range of values of x. But the uncertainty coming from where the real distribution starts becoming vertical on a log–log plot (i.e., α rising toward infinity) is central—statistical analysis is marred with too high sample errors in the tails to help us. This is a common problem of practice versus theory that we discuss later with the invisibility of the probability distribution[3] (For a typical misunderstanding of the point, see Ref. [11]).

Another problem: The unknowability of the upper bound invites faulty stress testing. Stress testing (say, in finance) is based on a probability-free approach to simulate a single, fixed, large deviation—as if it were the known payoff from a lottery ticket. However the choice of a "maximum" jump or a "maximum likely" jump is itself problematic, as it assumes knowledge of the structure of the distribution in the tails.[4] By assuming that tails are power-law distributed, though of unknown exact parameter, one can project richer sets of possible scenarios.

1.1. First Problem—The Effect of the Reliance on Gaussian Tools and the Dependence on the L^2 Norm

The finance literature uses variance as a measure of dispersion for the probability distributions, even when dealing with fat tails. This creates a severe problem outside the pure Gaussian nonscalable environment.

Financial economics is grounded in general Gaussian tools, or distributions that have all finite moments and correspondingly a characteristic scale, a category that includes the Log-Gaussian as well as subordinated processes with nonscalable jumps such as diffusion-Poisson,

[1] *There has been a family of econophysics papers that derive their principal differentiation from Mandelbrot [1] on the distinction between Levy-Stable basins and other power laws (infinite vs. finite variance) [3–9] and others.*

[2] *In finance, we generally assume x to be the logarithmic return* $\log(P_t/P_{t-\Delta t})$ *where* P *is the price,* t *is the period, and* Δt *is the increment.*

[3] *The inverse problem can be quite severe, leading to the mistake of assuming stochastic volatility (with the convenience of all moments) in place of a scale-free distribution (or, equivalently, one of an unknown scale). Cont and Tankov [10] show how a Student T with 3 degrees of freedom (infinite kurtosis) will mimic a conventional stochastic volatility model.*

[4] *One illustration of how stress testing can be deemed dangerous—as we do not have a typical deviation—is provided by the management of the 2007–2008 subprime crisis. Many firms, such as Morgan Stanley, lost large sums of their capital in the 2007 subprime crisis because their stress test underestimated the outcome, yet was compatible with historical deviations (see Ref. [12]).*

DOI 10.1002/cplx

regime switching models, or stochastic volatility methods (see Refs. [13–16]), outside what Mandelbrot [2(a,b,c)] bundles under the designation scale-invariant or fractal randomness.

All of these distributions can be called "fat-tailed," but not scalable in the above definition, as the finiteness of all moments makes them collapse into thin tails:

1. for some extreme deviation (in excess of some known level), or
2. rapidly under convolution, or temporal aggregation. Weekly or monthly properties are supposed to be closer, in distribution, to the Gaussian than daily ones. Likewise, fat tailed securities are supposed to add up to thin-tailed portfolios, as portfolio properties cause the loss of fat-tailed character rather rapidly, thanks to the increase in the number of securities involved.

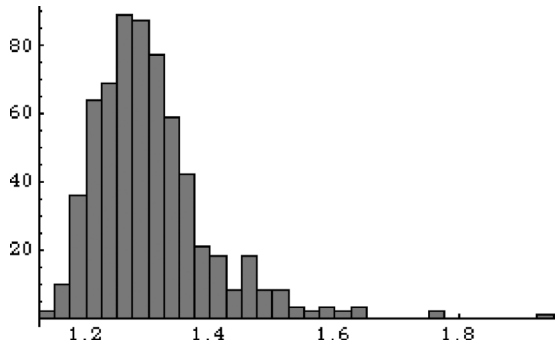
The dependence on these "pseudo-fat tails," or finite moment distributions, led to the building of tools based on the Euclidian norm, such as variance, correlation, beta, and other matters in L^2. It makes finite variance necessary for the modeling, and not because the products and financial markets naturally require such variance. We will see that the scaling of the distribution that affect the pricing of derivatives is the mean expected deviation, in L^1, which does not justify such dependence on the Eucledian metric.

The natural question here is: why do we use variance? Although it may offer some advantages, as a "summary measure" of the dispersion of the random variable, it is often meaningless in an environment in which higher moments do not lose significance.

But the practitioner use of variance can lead to additional pathologies. Goldstein and Taleb [17] show that most professional operators and fund managers use a mental measure of mean deviation as a substitute for variance, without realizing it: because the literature focuses exclusively on L^2 metrics, such as "Sharpe ratio," "portfolio deviations," or "sigmas." Unfortunately, the mental representation of these measures is elusive, causing a substitution. There seems to be a serious disconnect between decision making and projected probabilities. Standard deviation is exceedingly unstable compared with mean deviation in a world of fat tails (see an illustration in Figure 1).

F1

FIGURE 1



Distribution of the monthly STD/MAD ratio for the SP500 between 1955 and 2007. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

1.2. Second Problem—Life Outside the Asymptote: Questions Stemming from Idealization Versus Practice

The second, associated problem comes from the idealization of the models, often in exactly the wrong places for practitioners, leading to the reliance on results that work in the asymptotes, and only in the asymptotes. Furthermore, the properties outside the asymptotes are markedly different from those at the asymptote. Unfortunately, operators live far away from the asymptote, with nontrivial consequences for pricing, hedging, and risk management.

1.2.1. Time Aggregation

Take the example of a distribution for daily returns that has a finite second moment, but infinite kurtosis, say a power-law with exponent <4, of the kind we observe routinely in the markets. It will eventually, under time aggregation, say if we lengthen the period to weekly, monthly, or yearly returns, converge to a Gaussian. But this will only happen at infinity. The distribution will become increasingly Gaussian in the center, but not in the tails. Bouchaud and Potters [18] show how such convergence will be extremely slow, at the rate of $\sqrt{n\log(n)}$ standard deviations, where n is the number of observations. A distribution with a power law exponent $\alpha > 2$, even with a million convolutions, will eventually behave like a Gaussian up until about 3 standard deviations but conserve the power-law attributes outside of such regime. So, at best we are getting a mixed distribution, with the same fat tails as a non-Gaussian and the tails are where the problems reside.

More generally, the time-aggregation of probability distributions with some infinite moment will not obey the Central Limit Theorem in applicable time, thus leaving us with nonasymptotic properties to deal with in an effective manner Indeed, it may not be even a matter of time-window being too short, but for distributions with finite second moment, but with an infinite higher moment, for CLT to apply we need an infinity of convolutions.

DOI 10.1002/cplx

1.2.2. Discreteness

We operate in discrete time while much of the theory concerns mainly continuous time processes [19, 20], or finite time operational or computational approximations to true continuous time processes ([21]; review in Ref. [22]). Accordingly, the results coming from taking the limits of continuous time models, all Gaussian-based (nonscalable) pose difficulties in their applications to reality.

A scalable, unlike the Gaussian, does not easily allow for continuous time properties, because the continuous time limit allowing for the application of Ito's lemma is not reached, as we will see in Section 2.2.

1.3. Third Problem—Stability and Time Dependence

Most of the mathematical treatment of financial processes reposes on the assumption of time-independence of the returns. Whether it is for mathematical convenience (or necessity) it is hard to ascertain; but it remains that most of the distinctions between processes with finite second moment and others become thus artificial as they reposes heavily on such independence.

The consequence of such time dependence is the notion of distributional "stability," in the sense that a distribution loses its properties with the summation of random variables drawn from it. Much of the work discriminating between Levy-fat tails and non-Levy fat tails reposes on the notion that a distribution with tail exponent $\alpha < 2$ is held to converge to a Levy stable basin, while those with $\alpha \geq 2$ are supposed to become Gaussian.

The problem is that such notion of independence is a bit too strong for us to take it at face value. There may be serial independence in returns, but coexisting with some form of serial dependence in absolute returns, and the consequences on the tools of analysis are momentous.[5]

In other words, even in the asymptote, a process with finite variance that is not independently distributed is not guaranteed to become Gaussian.

This point further adds to the artificiality of the distinction between $\alpha < 2$ and $\alpha \geq 2$. Results of derivatives theory in the financial literature exclude path dependence and memory, which causes the aggregation of the process to hold less tractable properties than expected—making the convergence to a Gaussian basin of attraction less granted. Although the returns may be independent, absolute values of these returns may not be, which causes extreme deviations to cumulate in a manner to fatten the tails at longer frequencies (see Sornette [23] for the attributes of the drawdowns and excursions as these are more extreme than regular movements; deviations in the week of the stock market crash of 1987 were more extreme, statistically, than the day of the greatest move]. Naively, if you measure the mean average deviation of returns over a period, then lag them you will find that the measure of deviation is sensitive to the lagging period. (Also see Ref. [24], for a Gaussian-based test].

1.4. Fourth Problem—The Visibility of Statistical Properties in the Data

The final problem is that operators do not observe probability distributions, only realizations of a stochastic process, with a spate of resulting mistakes and systematic biases in the measurement process [25, 26]. Some complicated processes with infinite variance will tend to exhibit finite variance under the conventional calibration methods, such as the Hill estimator or the log-linear regressions [27]. In other words, for some processes, the typical error can be tilted toward the underestimation of the thickness of the tails. A process with an $\alpha = 1.8$ can easily yield $\alpha > 2$ in observations.

An argument in favor of "thin tails," or truncated power laws, is usually made with representations of the exceedant frequencies in log–log space that show the plot line getting vertical at some point, indicating an a pulling toward infinity. The problem is that it is hard to know whether this cutoff is genuine—and not the result of sample insufficiency. Such perceived cutoff can easily be the result of sampling error, given that we should find fewer data points in the far tails [28]. But in fact, assuming truncation is acceptable, we do not know where the distribution is to be truncated. Relying on the past yields in-sample obvious answers, but it does not reveal the true nature of the generator of the series. In the same vein, many researchers suggest the lognormal ([11]; see Ref. [2(a)] for the review), or stretched exponential [23]. On that score, the financial economics literature presents circular arguments, favoring Poisson jumps, and using the same assumed distribution to gauge the sufficiency of the sample, without considering the limitations of the sample in revealing tail events [29, 30].[6] (Simply, rare events are less likely to show in a finite sample; assuming homogeneous past data, 20-year history will not reveal one-in-50-year events.) The best answer, for a practitioner, is to plead ig-

[5] *One candidate process is Mandelbrot's multifractal model in which the tail exponent conserves under convolution. Daily returns can have a $\alpha = 3$, so will monthly returns.*

[6] *In addition, these tests are quite inadequate outside of $\mathbf{L}^2$, because they repose on measurement and forecast of variance.*

DOI 10.1002/cplx

norance: so long as we do not know where the truncation starts, it is safer to stick to the assumption of power laws.

1.5. The Major Consequence of These Four Problems

The cumulation of these four problems results in the following consequence: in "real life," the problems incurred when the tail exponent $\alpha < 2$ effectively prevail just as well when $\alpha > 2$. The literature [5–7, 31, 32] reports "evidence" in the equities markets, of a cubic α, i.e., a tail exponent around 3. Their dataset of around 18 million observations is available to most practitioners (including this author); it is extremely easy to confirm the result in sample. The econonophysics literature thus makes the distinction between Levy-regime and other whereas for us practitioners, because of the time aggregation problem, there is no such distinction. A scalable is a scalable: the tails never become thin enough to allow the use of Gaussian methods.

We will need to consider the consequences of the following two considerations:

1. The infinite moments never allow for derivations based on expansions and Ito's lemma.
2. Processes do not necessarily converge to the Gaussian basin, making conventional tools like standard deviations inapplicable.
3. Parameter discovery is not as obvious as in the Gaussian world.

The main option pricing and hedging consequences of scalability do not arise from the finiteness of the variance but rather from the lack of convergence of higher moments. Infinite kurtosis, which is what empirical data seems to point to in almost every market examined, has the same effect. There are no tangible, or qualitative differences in practice between such earlier models such as Mandelbrot [1], on one hand and later expositions showing finite variance models with a "cubic" tail exponent. Fitting these known processes induces the cost of severe mistracking of empirical reality.

The rest of this article will focus on the application of the above four problems and its consequence to derivatives pricing. We will first present dynamic hedging that is at the center of modern finance's version of derivatives pricing, and its difficulty outside of the Gaussian case owing to incompressible tracking errors. We then show how variance does not appear relevant for an option operator and that distributions with infinite variance are not particularly bothersome outside of dynamic hedging. Then we examine methods of pricing followed with the common difficulties in working with non-Gaussian distributions with financial products. We examine how these results can be extended to portfolio theory.

2. THE APPLICATIONS

2.1. Finite Variance Is Insufficient for Portfolio Theory

First, let us consider portfolio theory. There appears to be an accepted truism (after Markowitz [33]), that mean-variance portfolio allocation requires, but can be satisfied with, only the first two moments of the distribution -and that fatness of tails do not invalidate the arguments presented. I leave aside the requirement for a certain utility structure (a quadratic function) to make the theory work, and assume it to be acceptable in practice.[7]

Where x is the payoff (or wealth), and U the utility function:

$$U(X) = ax - bx^2, \quad a, b > 0,$$

By taking expectations, the utility of x

$$E[U(x)] = aE[x] - bE[x^2]$$

So seemingly higher moments do not matter. Such reasoning may work in the Platonic world of models, but, when turned into an application, even without relaxing any of the assumptions, it reveals a severe defect: where do we get the parameters from? $E[x^2]$, even if finite, is not observable. A distribution with infinite higher moments $E[x^n]$ (with $n > 2$) will not reveal its properties in finite sample. Simply, if $E[x^4]$ is infinite, $E[x^2]$ will not show itself easily. The expected utility will remain stochastic, i.e., unknown. Much of the problems in financial theory come from the dissipation upon application of one of the central hypotheses: that the operator knows the parameters of the distribution (an application of what Taleb [26] calls the "ludic fallacy").

The idea of mean-variance portfolio theory then has no possible practical justification.

2.2. Difficulties with Financial Theory's Approach to Option Pricing

2.2.1 Technë-Epistemë

The idea that operators need theory, rather than the other way around, has been contradicted by historical evidence [35]. They showed how option traders managed in a quite sophisticated manner to deal with option pricing and hedging—there is a long body of literature, from 1902, ignored by the economics literature presenting trading techniques and heuristics. The literature had been shy in considering the hypothesis that option price formation stems from supply and demand, and that traders manage

[7] *This is one of the arguments against the results of Mandelbrot [1] using the "evidence" provided by Officer [34].*

DOI 10.1002/cplx

to develop tricks and methods to eliminate obvious mispricing called "free lunches." So the result is that option theory seems to explain (or simplify) what is being done rather than drive price formation. Gatheral [16] defines the profession of modelers as someone who finds equations that fit prices in the market with minimal errors, rather than the reverse. Accordingly, the equations about the stochastic process do not have much beyond an instrumental use (to eliminate inconsistencies) and they do not correspond to a representation of future states of the world.[8]

2.2.2 Standard Financial Theory

Let us now examine how financial theory "values" financial products (using an engineering/practitioner exposition). The principal difference in paradigm between the one presented by Bachelier [36] and the modern finance one known as Black-Scholes-Merton [19, 37] lies in the following main point. Bachelier's model is based on an actuarial expectation of final payoffs. The same method was later used by a series of researchers, such as Sprenkle [38], Boness [39], Thorp and Kassouf [40], Thorp [41]). They all encountered the following problem: how to produce a risk parameter—a risky asset discount rate—to make it compatible with portfolio theory? The Capital Asset Pricing Model requires that securities command an expected rate of return in proportion to their "riskiness." In the Black-Scholes-Merton approach, an option price is derived from continuous-time dynamic hedging, and only in properties obtained from continuous time dynamic hedging. We will describe dynamic hedging in some details further down. Thanks to such a method, an option collapses into a deterministic payoff and provides returns independent of the market; hence it does not require any risk premium.

The problem we have with the Black-Scholes-Merton approach is that the requirements for dynamic hedging are extremely idealized, requiring the following strict conditions idealization might have gone too far, and dangerously so, of the style "assume the earth was square." The operator is assumed to be able to buy and sell in a frictionless market, incurring no transaction costs. The procedure does not allow for the price impact of the order flow, if an operator sells a quantity of shares, it should not have consequences on the subsequent price. The operator knows the probability distribution, which is the Gaussian, with fixed and constant parameters through time (all parameters do not change). Finally, the most significant restriction: no scalable jumps. In a subsequent revision [20(a)] allows for jumps but these are deemed to be Poisson arrival time, and fixed or, at the worst, Gaussian. The framework does not allow the use of power laws both in practice and mathematically. Let us examine the mathematics behind the stream of dynamic hedges in the Black-Scholes-Merton equation.

Assume the risk-free interest rate $r = 0$ with no loss of generality. The canonical Black-Scholes-Merton model consists in building a dynamic portfolio by selling a call and purchasing shares of stock that provide a hedge against instantaneous moves in the security. Thus the value of the portfolio π locally "hedged" against exposure to the first moment of the distribution is the following:

$$\pi = -C + \frac{\partial C}{\partial S} S$$

where C is the call price, and S the underlying security.

By expanding around the initial values of the underlying security S, we get the changes in the portfolio in discrete time. Conventional option theory applies to the Gaussian in which all orders higher than ΔS^2 and Δt (including the cross product $\Delta S\, \Delta t$) are neglected.

$$\Delta\pi = \frac{\partial C}{\partial t}\Delta t - \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\Delta S^2 + O(\Delta S^3)$$

Taking expectations on both sides, we can see very strict requirements on moment finiteness: all moments need to converge for us to be comfortable with this framework. If we include another term, ΔS^3, it may be of significance in a probability distribution with significant cubic or quartic terms. Indeed, although the nth derivative with respect to S can decline very sharply, for options that have a strike K away from the initial price S, it remains that the moments are rising disproportionately fast, enough to cause potential trouble.

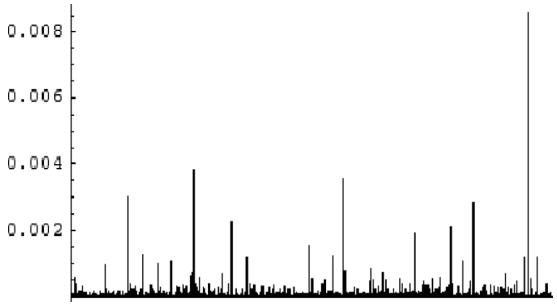
So here we mean all moments need to be finite and losing in impact, no approximation would do. Note here that the jump diffusion model [20(a)] does not cause much trouble for researchers since it has all the moments, which explains its adoption in spite of the inability to fit jumps in a way that tracks them out-of-sample. And the annoyance is that a power law will have every moment higher than a infinite, causing the equation of the Black-Scholes-Merton portfolio to fail.

As we said, the logic of the Black-Scholes-Merton so-called solution is that the portfolio collapses into a deterministic payoff. But let us see how quickly or effectively this works in practice.

[8] *In the same vein, we repeat that the use of power-laws does not necessarily correspond to the belief that the distribution is truly parametrized as a power law, rather selected owing to the absence of knowledge of the properties in the tails.*

DOI 10.1002/cplx

FIGURE 2

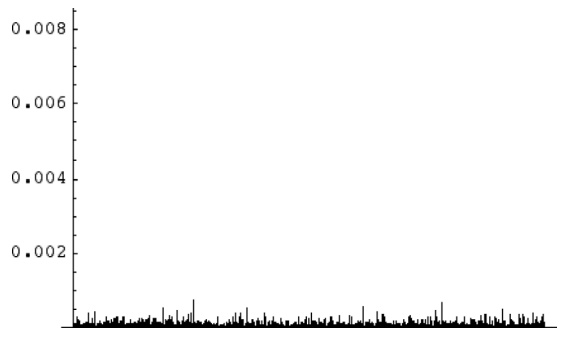


The hedging errors for an option portfolio (under a daily revision regime) over 3000 days, under a constant volatility Student T with tail exponent $\alpha = 3$. Technically the errors should not converge in finite time as their distribution has infinite variance.

2.2.3. The actual replication process

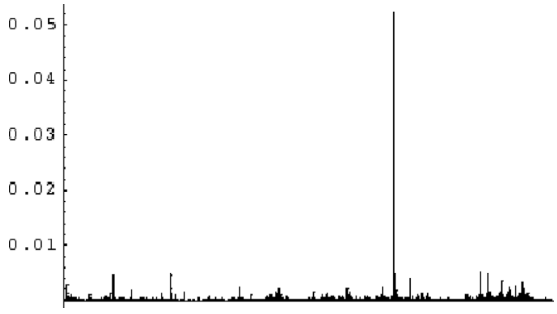
According to standard financial economics [20(b)], the payoff of a call is expected to be replicated in practice with the following stream of dynamic hedges. The procedure is as follows (again, assuming 0 interest rates): Take C, the initial call price; S_t, the underlying security at initial period t; and T, the final expiration of the option. The performance will have three components: (1) C the initial call value as cash earned by the option seller, (2) Max $(S_T - K,0)$, the final call value (intrinsic value) that the option seller needs to disburse, and (3) the stream of dynamic hedges aiming at offsetting the pair $C - Max\ (S_T - K,0)$, in quantities of the underlying held in inventory, revised at different periods.

FIGURE 3



Hedging errors for an option portfolio (equivalent daily revision) under an equivalent "Black-Scholes" world.

FIGURE 4



Portfolio Hedging errors including the stock market crash of 1987.

So we are concerned with the evolution between the two periods t and T and the stream of dynamic hedges. Break up the period $(T - t)$ into n increments Δt. The operator changes the hedge ratio, i.e., the quantities of the underlying security he is supposed to have in inventory, as of time $t + (i - 1)\Delta t$, then gets the difference between the prices of S at periods $t + (i - 1)\ \Delta t$ and $t + i\Delta t$ (called nonanticipating difference). Where P is the final profit/loss:

$$P = -C + (K - S_T) + \sum_{t=1}^{n=\frac{T-t}{\Delta t}} \frac{\partial C}{\partial S}\Big|_{S=S_{t+(t-1)\Delta t},t=1+(i-1)\Delta t} \times (S_{t+i\Delta t} - S_{t+(t-1)\Delta t})$$

Standard option theory considers that the final package P of the three components will become deterministic at the limit of $\Delta t \to 0$, as the stream of dynamic hedges reduces the portfolio variations to match the option. This seems mathematically and operationally impossible.[9]

2.2.4. Failure: How Hedging Errors Can be Prohibitive

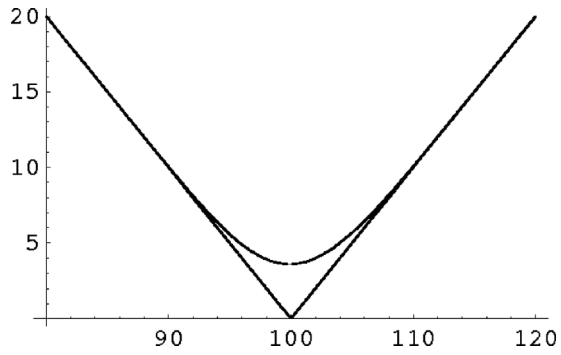
As a consequence of the mathematical property seen above, hedging errors in an cubic α appear to be indistinguishable from those from an infinite variance process. Furthermore, such error has a disproportionately large effect on strikes, as we illustrate in Figures 2–4. Figure 2 illustrate the portfolio variations under a finite variance power law distribution, subjected to the same regime of revision (Δt = 1 business day, 1/252), compared with the Gaussian in Figure 3. Finally, Figure 4 shows the real market (including the crash of 1987). F2-F4

In short, dynamic hedging in a power law world does not remove risk.

[9] *See Bouchaud and Potters [42] for a critique.*

DOI 10.1002/cplx

FIGURE 5



Straddle price: options are piecewise linear, with a hump at the strike, the variance does not enter the natural calculation.

2.2.5. Options Without Variance

On the basis of this dynamic hedging problem, we look at which conditions we need to "price" the option. Most models base the option on variance. Clearly options do not depend on variance, but on mean average deviation, but it is expressed in terms of variance.

Consider the Bachelier expectation framework, the actuarial method of discounting the probabilistic payoffs of the options. Where F is the forward, the price in the market for the delivery of S at period T, and regardless of the probability distribution φ, under the sole restriction that the first moment $\int F\, \varphi(F)\, dF$ exists, the puts and calls can be priced as follows (assuming, to simplify, 0 financing rates):

$$C(K) = \int_K^\infty (F - K)\phi(F)dF$$

$$P(K) = \int_0^K (K - F)\phi(F)dF$$

where $C(K)$ and $P(K)$ are the call and put struck at K, respectively. Thus when the options are exactly at-the-money by the forward, i.e., $K = F$, each delivers half the discounted mean absolute deviation

$$C(K) = P(K) = \frac{1}{2}\int_0^\infty (\Delta F)\phi(F)dK$$

An option's payoff is piecewise linear as can be shown in Figure 5. F5

In a Gaussian world, we have the mean absolute deviation over standard deviation as follows:

$$\frac{\overline{m}}{\sigma} = \frac{2}{\pi}$$

But, with fat tails, the ratio of the dispersion measures drops, as σ reaches infinity when $1 < \alpha < 2$. This means that a simple sample of activity in the market will not reveal much since most of the movements become concentrated in a fewer and fewer number of observations. Intuitively 67% of observations take place in the "corridor" between $+1$ and -1 standard deviations in a Gaussian world. In the real world, we observe between 80% and 99% of observations in that range—so large deviations are rare, yet more consequential. Note that for the conventional results, we get in finance ["cubic"] α , about 90.2% of the time is spent in the $[-1,+1]$ standard deviations corridor. Furthermore, with $\alpha = 3$, the previous ratio becomes

$$\frac{\overline{m}}{\sigma} = \frac{2}{\pi}$$

2.2.6. Case of a Variance Swap

There is an exception to the earlier statement that derivatives do not depend on variance. The only common financial product that depends on L^2 is the "variance swap": a contract between two parties agreeing to exchange the difference between an initially predetermined price and the delivered variance of returns in a security. However, the product is not replicable with single options.

The exact replicating portfolio is constructed [16] in theory with an infinity of options spanning all possible strikes, weighted by K^{-2} , where K is the strike price.

$$\int_0^\infty \frac{1}{K^2} C(K)dK - \Delta S$$

However, to turn this into practice in the real world requires buying an infinite amount of options of a strike K approaching 0, and an inifinitesimal amount of options with strike K approaching infinity. In the real world, strikes are discrete and there is a lower bound K_L and a highest possible one K_H. So what the replication leaves leaves out, $<K_L$ and $>K_H$ leaves us exposed to the large deviations; and would cost an infinite amount to purchase when options have an infinite variance.

The discrete replicating portfolio would be as follows: by separating the options into $n + 1$ strikes between K_L and K_H incrementing with ΔK.

$$\sum\nolimits_{i=0}^{n} \frac{C(i\Delta K + K_L) + P(i\Delta K + K_L)}{(K_L + i\Delta K)^2}$$

Such a portfolio will be extremely exposed to mistracking upon the occurrence of tail events.

DOI 10.1002/cplx

We conclude this section with the remark that, effectively, and without the granularity of the market, the dynamic hedging idea seriously underestimates the effectiveness of hedging errors, from discontinuities and tail episodes. As a matter of fact it plainly does not seem to work both practically and mathematically.

The minimization of daily variance may be effectual in smoothing performance from small moves, but it fails during large variations. In a Gaussian basin very small probability errors do not contribute to too large a share of total variations; in a true fat tails environment, and with nonlinear portfolios, the extreme events dominate the properties. More specifically, an occasional sharp move, such a "22 sigma event" (expressed in Gaussian terms, by using the standard deviation to normalize the market variations), of the kind that took place during the stock market crash of 1987, would cause a severe loss that would cost years to recover. Define the "daily time decay" as the drop in the value of the option over 1/252 years assuming no movement in the underlying security; a crash similar to 1987 would cause a loss of close to hundreds of years of daily time decay for a far out of the money option, and more than a year for the average option.[10]

2.3. How Do We Price Options Outside of the Black-Scholes-Merton Framework?

It is not a matter of "can." We "need" to do so once we lift the idealized conditions, and we need to focus on the properties of the errors.

We just saw that scalability precludes dynamic hedging as a means to reach a deterministic value for the portfolio. There are of course other impediments for us, merely the fact that in practice we cannot reach the level of comfort owing to transaction costs, lending and borrowing restrictions, price impact of actions, and a well-known problem of granularity that prevent us from going to the limit. In fact continuous-time finance [20(b)] is an idea that got plenty of influence in spite of both its mathematical stretching and its practical impossibility.

If we accept that returns are power-law distributed, then finite or infinite variance matter little. We need to use expectations of terminal payoffs, and not dynamic hedging. The Bachelier framework, which is how option theory started, does not require dynamic hedging. Derman and Taleb [46] argue how one simple financing assumption, the equality of the cost-of-carry of both puts and calls, leads to the recovery of the Black-Scholes equation in the Bachelier framework without any dynamic hedging and the use of the Gaussian. Simply a European put hedged with long underlying securities has the same payoff as the call hedged with short underlying securities and we can safely assume that cash flows must be discounted in an equal manner. This is common practice in the trading world; it might disagree with the theories of the Capital Asset Pricing Model, but this is simply because the tenets behind CAPM do not appear to draw much attention on the part of practitioners, or because the bulk of tradable derivatives are in fixed-income and currencies, products unconcerned by CAPM. Futhermore, practitioners, are not concerned by CAPM (see Taleb [47] and Haug [48]). We do not believe that we are modeling a true expectation, rather fitting an equation to work with prices. We do not "value." Finally, thanks to this method, we no longer need to assume continuous trading, absence of discontinuity, absence of price impact, and finite higher moments. In other words, we are using a more sophisticated version of the Bachelier equation; but it remains the Bachelier [36] nevertheless. And, of concern here, we can use it with power laws with or without finite variance.

2.4. What Do We Need? General Difficulties with the Applications of Scaling Laws

This said, while a Gaussian process provides a great measure of analytical convenience, we have difficulties building an elegant, closed-form stochastic process with scalables.

Working with conventional models presents the following difficulties.

2.4.1. First Difficulty: Building a Stochastic Process

For pricing financial instruments, we can work with terminal payoff, except for those options that are path-dependent and need to take account of full-sample path.

Conventional theory prefers to concern itself with the stochastic process $dS/S = m\ dt + \sigma\ dZ$ (S is the asset price, m is the drift, t is time, and σ is the standard deviation) owing to its elegance, as the relative changes result in exponential limit, leading to summation of instantaneous logarithmic returns, and allow the building of models for the distribution of price with the exponentiation of the random variable Z, over a discrete period Δt, $S_{t+\Delta t} = S_t\ e^{a+bz}$. This is convenient: for the expectation of S, we need to integrate an exponentiated variable

$$E[S_{t+\Delta t}] = S_t \int e^{a+bz}\phi(z)dz$$

which means that we need the density of z, in order to avoid finite expectations for S, to present a compensating exponential decline and allow bounding the integral. This explains the prevalence of the Lognormal distribution in

[10] *One can also fatten the tails of the Gaussian and get a power law by changing the standard deviation of the Gaussian: Dupire [43(a)], Derman and Kani [44], Borland [45]. See Gatheral [16] for a review.*

DOI 10.1002/cplx

asset price models—it is convenient, if unrealistic. Further, it allows working with Gaussian returns for which we have abundant mathematical results and well known properties.

Unfortunately, we cannot consider the real world as Gaussian, not even Log-Gaussian, as we have seen earlier.

How do we circumvent the problem? This is typically done at the cost of some elegance. Many operators [49] use, even with a Gaussian, the arithmetic process first used in Bachelier [36],

$$S_{t+\Delta t} = a + bz + S_t$$

which was criticized for delivering negative asset values (though with a minutely small probability). This process is used for interest rate changes, as monetary policy seems to be done by fixed cuts of 25 or 50 basis points regardless of the level of rates (whether they are 1% or 8%).

Alternatively, we can have recourse to the geometric process for a , large enough, Δt (say 1 day)

$$S_{t+\Delta t} = S_t(1 + a + bz)$$

Such a geometric process can still deliver negative prices (a negative, extremely large value for z) but is more in line with the testing done on financial assets, such as the SP500 index, as we track the daily returns $r_t = (P_t - P_{t-\Delta t})/P_{t-\Delta t}$ in place of $\log(P_t/P_{t-\Delta t})$. The distribution can be easily truncated to prevent negative prices (in practice the probabilities are so small that it does not have to be done as it would drown in the precision of the computation).

2.4.2. Second Difficulty: Dealing with Time Dependence

We said earlier that a multifractal process conserves its power law exponent across timescales (the tail exponent α remains the same for the returns between periods t and $t + \Delta t$, independently of Δt). We are not aware of an elegant way to express the process mathematically, even computationally, nor can we do so with any process that does not converge to the Gaussian basin. But for practitioners, theory is not necessary; traders need tricks. So the difficulty can be remedied, in the pricing of securities by, simply, avoiding to work with processes, and limiting ourselves to working with distributions between two discrete periods. Traders call that "slicing" [47], in which we work with different periods, each with its own sets of parameters. We avoid the complication of studying the process between these discrete periods.

3. CONCLUDING REMARKS

This article outlined the following difficulties: working in quantitative finance, portfolio allocation, and derivatives trading while being suspicious of the idealizations and assumptions of financial economics, but avoiding some of the pitfalls of the econophysics literature that separate models across tail exponent $\alpha = 2$, truncate data on the occasion, and produce results that depend on the assumption of time independence in their treatment of processes. We need to find bottom-up patches that keep us going, in place of top-down, consistent but nonrealistic tools and ones that risk getting us in trouble when confronted with large deviations. We do not have many theoretical answers, nor should we expect to have them soon. Meanwhile option trading and quantitative financial practice will continue under the regular tricks that allow practice to survive (and theory to follow).

ACKNOWLEDGMENTS

The author acknowledges Simon Benninga, Benoit Mandelbrot, Jean-Philippe Bouchaud, Jim Gatheral, Espen Haug, Martin Shubik, Eric Smith, Aaron Brown, and an anonymous reviewer.

REFERENCES

1. Mandelbrot, B. The variation of certain speculative prices. J Bus 1963, 36, 394–419.
2. (a) Mandelbrot, B. Fractals and Scaling in Finance; Springer-Verlag, 1997; (b) Mandelbrot, B. Scaling in financial prices, I: Tails and dependence. Quant Finance 2001a, 1, 113–123; (c) Mandelbrot, B. Scaling in financial prices, II: Multifractals and the star equation. Quant Finance 2001b, 1, 124–130.
3. Plerou, V.; Gopikrishnan, P.; Gabaix, X.; Amaral, L.; Stanley, H.E. Price fluctuations, market activity, and trading volume. Quant Finance 2001, 1, 262–269.
4. Stanley, H.E.; Amaral, L.; Gopikrishnan, P.; Plerou, V. Scale invariance and universality of economic fluctuations. Phys A 2000, 283, 31–41.
5. Gabaix, X.; Gopikrishnan, P.; Plerou, V.; Stanley, E. A theory of power law distributions in financial market fluctuations. Nature 2003, 423, 267–230.
6. Gabaix, X.; Gopikrishnan, P.; Plerou, V.; Stanley, E. Are Stock Market Crashes Outliers? MIT Mimeo. Massachusetts Institute of Technology: Cambridge, 2003.
7. Gopikrishnan, P.; Meyer, M.; Amaral, L.; Stanley, H.E. Inverse cubic law for the distribution of stock price variations. Eur Phys J B 1998, 3, 139–140.

DOI 10.1002/cplx

8. Gopikrishnan, P.; Plerou, V.; Amaral, L.; Meyer, M.; Stanley, H.E. Scaling of the distribution of fluctuations of financial market indices. Phys Rev E 1999, 60, 5305–5316.
9. Gopikrishnan, P.; Plerou, V.; Gabaix, X.; Stanley, H.E. Statistical properties of share volume traded in financial markets. Phys Rev E 2000, 62, R4493–R4496.
10. Cont, R.; Tankov, P. Financial modelling with Jump Processes; Chapman & Hall/CRC Press: 2003.
11. Perline, R. Strong, weak, and false inverse power laws. Stat Sci 2005, February, 20–21.
12. The Risk Maverick. Bloomberg, May 2008.
13. Hull, J.; White, A. The pricing of options on assets with stochastic volatilities, Journal of Finance 1987, 42, 281–300.
14. Heston, S.L. A closed-form solution for options with stochastic volatility, with application to bond and currency options. Rev Financ Studies 1993, 6, 327–343.
15. Duffie, D.; Pan, J.; Singleton, K. Transform analysis and asset pricing for affine jump diffusions. Econometrica 2000, 68, 1343–1376.
16. Gatheral, J. Stochastic Volatility Modeling; Wiley: 2006.
17. Goldstein, D.G.; Taleb, N.N. We don't quite know what we are talking about when we talk about volatility. J Portfolio Manag 2007.
18. Bouchaud, J.-P.; Potters, M. Theory of financial risks and derivatives pricing. In: From Statistical Physics to Risk Management, 2nd ed.; Cambridge University Press: 2003.
19. Merton, R.C. Theory of rational option pricing. Bell J Econ Manag Sci 1973, 4, 141–183.
20. (a) Merton, R.C. Option pricing when underlying stock returns are discontinuous. J Financ Econ 1976, 3, 125–144; (b) Merton, R.C. Continuous-Time Finance, revised edition; Blackwell: 1992.
21. Cox, J.C.; Ross, S.A. The valuation of options for alternative stochastic processes. J Financ Econ 1976, 3, 145–166.
22. Baz, J.; Chacko, G. Financial Derivatives: Pricing, Applications, and Mathematics; Cambridge University Press: 2004.
23. Sornette. Critical Phenomena in Natural Sciences: Chaos, Fractals, Self-Organization and Disorder: Concepts and Tools, 2nd ed.; Springer, 2004.
24. Lo, A.; MacKinley, A.C. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1988, 1, 41–66.
25. Taleb, N.N.; Pilpel, A. I problemi epistemologici del risk management In: Daniele Pace (a cura di) Economia del rischio. Antologia di scritti su rischio e decisione economica, Giuffre: Milan, 2004.
26. Taleb, N.N. The Black Swan and the domains of statistics. Am Stat 2007, 61, 3.
27. Weron, R. Levy-stable distributions revisited: Tail index >2 does not exclude the levy-stable regime. Int J Modern Phys 2001, 12, 209–223.
28. Taleb, N. Fat Tails, asymmetric knowledge, and decision making. Wilmott Magazine March 2005, 56–59.
29. Coval, J.; Shumway, T. Expected option returns. Journal of Finance 2001, 56, 983–1009.
30. Bodarenko, Why are Put Options So Expensive? SSRN Preprint. 2004.
31. Gabaix, X.; Ramalho, R.; Reuter, J. Power Laws and Mutual Fund Dynamics. MIT Mimeo. Massachusetts Institute of Technology: Cambridge, 2003c.
32. Plerou, V.; Gopikrishnan, P.; Gabaix, X.; Stanley, H.E. On the origins of power law fluctuations in stock prices. Quantitative Finance 2004, 4, C11–15.
33. Markowitz, H. Portfolio selection. J Finance 1952, 7, 77–91.
34. Officer, R.R. J Am Stat Assoc 1972, 67, 807–812.
35. Haug, E.G.; Taleb, N.N. Why We Never Used the Black Scholes Mertn Option Pricing Formula. Wilmott Magazine, in press.
36. Bachelier, L. Theory of speculation. In: The Random Character of Stock Market Prices; Cootner, P., Ed. (edited in 1964); MIT Press: Cambridge, MA, 1900.
37. Black, F.; Scholes, M. The pricing of options and corporate liabilities. J Pol Econ 1973, 81, 631–659.
38. Sprenkle, C. Warrant prices as indicators of expectations and preferences. Yale Econ Essays 1961, 1, 178–231.
39. Boness, A. Elements of a theory of stock-option value. J Pol Econ 1964, 72, 163–175.
40. Thorp, E.O.; Kassouf, S.T. Beat the Market; Random House: New York, 1967.
41. Thorp, E.O. 1969. Optimal gambling systems for favorable games. Rev Int Stat Inst 1969, 37.
42. Bouchaud, J.-P.; Potters, M. Welcome to a non-Black-Scholes world. Quant Finance 2001, 1, 482–483.
43. (a) Dupire, B. Pricing with a smile. Risk 1994, 7, 18–20; (b) Dupire, B. A new approach for understanding the impact of volatility on option prices. Discussion paper. Nikko Financial Products, 1998.
44. Derman, E.; Kani, I. Riding on a smile. Risk 1994, 7, 32–39.
45. Borland, L. Option pricing formulas based on a non-Gaussian stock price model. Phys Rev Lett 2002, 89, 9.
46. Derman, E.; Taleb, N.N. The illusion of dynamic delta replication. Quant Finance 2005, 5, 323–326.
47. Taleb, N.N. Dynamic Hedging: Managing Vanilla and Exotic Options; Wiley: New York, 1996.
48. Haug, E.G. Derivatives Models on Models; Wiley: New York, 2007.
49. Wilmott, P. Paul Wilmott on Quantitative Finance; Wiley: Chichester, 2000.

DOI 10.1002/cplx

Statistical Undecidability

Raphael Douady, CNRS & RiskData
Nassim N. Taleb, NYU-Poly

October 2010

Presentation of the result:

Using the metadistribution of possible distributions for a given measure, we define a condition under which it is possible to make a decision based on the observation of random variable, which we call "statistical decidability". We provide a sufficient condition on the metadistribution for the decision to be "statistically decidable" and conjecture that decisions based on a metadistribution with non compact support are always "statistically undecidable". There is the need for a strong *undefeasable a priori* without which decisions are not statistically justified — an effect that is very significant for decisions affected by small probabilities.

Decisions are not made on naive measure of True/False in simple cumulative probability space, but on a higher moments (say, expectation or some similar decision measure such as utility) —off some numerical decidability criterion. Unlike the Gödel result, which has not yet shown practical significance, the added dimension of consequence or utility of decision makes enormous consequences, making situations completely undecidable statistically.

Bayesian updating methods do not bring any remedy as they are much more prior-dependent than is thought naively by preselecting prior data and *a priori* (nonrevisable) distribution (i.e, without metadistribution). Maximum likelihood estimations are even worse as, by inverting the question of the distribution of the objective criterion and that of the sample conditionally to a choice of distribution, they provide absolutely no control on the objective criterion. In both cases, two observers can observe the same series, without ever converging.

Introduction

Let Ω be the space of possible eventualities (the *"random space"*) and μ be the (unknown) probability distribution on it. We need to take an "informed" decision, based on a criterion $\Phi(\mu)$ that depends on μ. Therefore Φ is a function defined on $\wp(\Omega)$ with values in a set V depending on the nature of the decision. For example:

- Yes/No decision: $\Phi : \wp(\Omega) \to V = \{0,1\}$
- Quantitative decision: $\Phi : \wp(\Omega) \to V = \mathbf{R}$ or $\mathbf{R}^d$

The decision will be taken with respect to the estimated distribution of $\Phi(\mu)$ knowing all or some of the available information.
Let us assume that Φ is continuous with respect to some norm $\|.\|_{\wp(\Omega)}$ on $\wp(\Omega)$. We shall assume that μ is drawn from an *a priori* distribution π on the σ-algebra spanned by this norm.

Let $\pi_\Phi = \Phi_* \pi$ be the image measure in V, that is, the distribution of $\varphi = \Phi(\mu)$ according to the distribution π. The decision will in fact not be taken with respect to $\Phi(\mu)$, which is unknown, but with respect to a criterion $\Psi(\pi_\Phi) \in V$, where the function $\Psi : \wp(V) \to V$ is assumed to be continuous with respect to a norm $\|.\|_{\wp(V)}$ and such that, for a Dirac mass δ_a on $a \in V$, one has $\Psi(\delta_a) = a$ (in other words, Ψ coincides with Φ when μ is perfectly known).
Let us now assume that the information is given by a sample of values of random variables $X_i(\omega)$, $i \in \{1,\ldots,n\}$, $\omega \in \Omega$, drawn at random from the probability distribution μ. Our decision question can be restated as:

- *What is the distribution of* $\Phi(\mu)$ *knowing* $(X_1,\ldots,X_n)$ *?* (Q_1)

Let us consider the compound random variables $(\xi_1,\ldots,\xi_n)$ defined by picking μ at random with respect to π, then ω at random with respect to μ and compute $X_i(\omega)$. Our question Q_1 can now be restated in questions Q_2 and Q_3 as follows:

- *What is the joint distribution of* $(\varphi, \xi_1,\ldots, \xi_n)$ *in* $V \times \mathbf{R}^{nd}$ *?* (Q_2)
- *What is the conditional distribution of* φ *in* V *knowing* $(\xi_1,\ldots,\xi_n)$ *?* (Q_3)

We can see Q_3 as a function $g_\pi : \mathbf{R}^{nd} \to \wp(V)$, then the decision criterion is the function $\psi = \Psi \circ g_\pi$. For this criterion to be usable, it must be well defined, continuous with respect to input values of $(\xi_1,\ldots,\xi_n)$ – hence g must be continuous when the image space $\wp(V)$ is equipped with the norm $\|.\|_{\wp(V)}$ – and converge to the criterion φ when n tends to $+\infty$.
Now comes the general question that π itself is generally unknown. At best, we assume that μ is picked within a certain class $C \subset \wp(\Omega)$.

Definition

A decision based on criteria Φ and Ψ and distribution π is *statistically decidable* if the following holds:

1. For any fixed n, the function $\psi : \mathbf{R}^{nd} \to V$ is well defined. If it is given as an integral with respect to π, then the integrand must be π-integrable.
2. For any fixed n, the function $\psi : \mathbf{R}^{nd} \to V$ is continuous with respect to the sample $(\xi_1,\ldots,\xi_n)$
3. Let us assume that $(X_1,\ldots,X_n)$ are drawn from a given measure μ and let us consider the sample error $\varepsilon(X_1,\ldots,X_n) = |\psi(X_1,\ldots,X_n) - \Phi(\mu)|$ and its expectation $\mathrm{Err}(\mu) = \mathrm{E}_\mu[\varepsilon(X_1,\ldots,X_n)]$. Then $\mathrm{Err}(\mu)$ must tend to 0 when n tends to $+\infty$ both π-almost surely and in $\mathrm{L}^1(\pi)$.

Otherwise it is said *statistically undecidable.* The latter condition is probably the most important of all: it means that no uncertainty on the distribution is left aside when the sample is large enough, so that the decision criterion corresponds to that originally fixed by the problem.
When π is unknown within a class $\Gamma \subset \wp(\wp(\mathbf{R}^d))$, then for the decision to be *statistically decidable,* functions $\psi = \Psi \circ g_\pi$ must be equi-continuous and the convergence of errors to 0 must be uniform in the class Γ.

Bayesian Statistics

Bayesian statistics are based on a prior distribution μ_0 then, given a sample X, the probability is modified to a posterior distribution μ_1 that depends on the prior probability of the sample:

$$\mu_1(A) = \frac{\mu_0(X \mid A)}{\mu_0(X)} \mu_0(A)$$

Explain why the knowledge of $\Phi(\mu_1)$ doesn't give any info the distribution of $\Phi(\mu)$ knowing X.

Maximum Likelihood

Given a sample $X = (X_1,\ldots,X_n)$, one defines the likelihood of a distribution $L(\mu) = \prod_{i=1}^{n} f_\mu(X_i)$ where f_μ is the pdf of μ. Then assuming $\mu = \mu_\alpha$ depends on a parameter $\alpha \in \mathbf{R}^d$ with $d < n$, one selects the parameter $a_{\max}$ that maximizes the likelihood $L(\mu_{\alpha\max})$.
Explain why the knowledge of $\Phi(\mu_{\alpha\max})$ doesn't give any info the distribution of $\Phi(\mu)$ knowing X.

Fourier Transform

Let us consider question (Q_3). By definition of conditional distributions, for any test functions $h(\mu)$ and $u_i(\xi_i)$, $i = 1\ldots n$, one has:

$$\int h(\mu)u_1(\xi_1)\ldots u_n(\xi_n)d\pi_\xi(\mu) = \int h(\mu)u_1(X_1)\ldots u_n(X_n)d\mu(X_1)\ldots d\mu(X_n)d\pi(\mu)$$

Assume that $\Psi(\pi) = \int U(\varphi(\mu))d\pi(\mu)$ and set $\psi(\xi) = \Psi \circ g_\pi(\xi)$. One has:

$$\begin{aligned}\int \psi(x)u_1(x_1)\ldots u_n(x_n)dx_1\ldots dx_n &= \int U(\Phi(\mu))u_1(X_1)\ldots u_n(X_n)d\mu(X_1)\ldots d\mu(X_n)d\pi(\mu)\\ &= \int U(\mathrm{E}_\mu(f))\mathrm{E}_\mu(u_1)\ldots\mathrm{E}_\mu(u_n)d\pi(\mu)\end{aligned}$$

Where $\Phi(\mu) = \int f d\mu$.
Using functions $u(x) = \exp(itx)$, we get the Fourier transform of ψ:

$$\begin{aligned}\hat{\psi}(t_1,\ldots,t_n) &= \int U(\mathrm{E}_\mu(f))\hat{\mu}(t_1)\ldots\hat{\mu}(t_n)d\pi(\mu)\\ &= \int U\left(\frac{1}{2\pi}\int \hat{f}(s)\hat{\mu}(s)ds\right)\hat{\mu}(t_1)\ldots\hat{\mu}(t_n)d\pi(\mu)\end{aligned}$$

We can therefore deduce the following:

Theorem
The function ψ is continuous – hence the statistical problem is decidable *– if:*

$$\int\left|\int U\left(\frac{1}{2\pi}\int \hat{f}(s)\hat{\mu}(s)ds\right)\hat{\mu}(t_1)\ldots\hat{\mu}(t_n)d\pi(\mu)\right| dt_1\ldots dt_n < +\infty$$

Conversely, if ψ is continuous – i.e. if the problem is decidable *– then:*

$$\lim_{\sum t_i^2 \to +\infty} \int U\left(\frac{1}{2\pi}\int \hat{f}(s)\hat{\mu}(s)ds\right)\hat{\mu}(t_1)\ldots\hat{\mu}(t_n)d\pi(\mu) = 0$$

Would this condition not be satisfied, then the problem would be undecidable.

Conjectures

Here is a list of conjectures that express "generic statistical undecidability":

1. If, for any criterion Ψ of the form $\Psi(\pi) = \int U(\varphi(\mu))d\pi(\mu)$, the problem is statistically decidable, then the metadistribution π has compact support in $\wp(\Omega)$. This result would show that for a problem to be statistically decidable, one needs either to make assumptions on the growth of the criterion at infinity, or strong a priori assumptions, such as a finitely parameterized class, on the acceptable measures.
2. Whatever the norm on $\wp(\wp(\Omega))$, the map $\pi \to \Psi \circ g$ is generically discontinuous. This means that very minor changes in the a priori distribution π lead to completely different decision criteria.
3. If the class C of possible π is not compact (a set with non empty interior in $\wp(\wp(\Omega))$ is not compact, whatever the norm), then the set of corresponding criteria is generically not uniformly continuous. This means that even when assuming that π is close to a given a priori probability measure π_0, one cannot control the sensitivity of the decision to inputs.
4. The more Φ depends on areas where μ has low probability, the less $\Psi \circ g$ is continuous, i.e. very close input samples can lead to very different decisions. This assertion, which needs a precise definition of "depending on where μ has low probability", exactly express the fact that small probabilities are harder to estimate than large ones. More precisely, let us assume that the norm $\|.\|_{\wp(\Omega)}$ is the dual of the standard max norm on $L^\infty(\Omega)$. Then the modulus of continuity of $\Psi \circ g$ is generically no better than that of Φ.

The Illusion of Thin-Tails Under Aggregation

Nassim N. Taleb, NYU-Poly

George A. Martin, Associate Director, CISDM, University of Massachusetts

January 2012

We disagree with the statements in Treynor (2011) on logical, epistemological, empirical, and mathematical grounds.

Let us summarize Treynor's key argument: he accepts that equity market returns on a daily basis are fat-tailed, in agreement with the first author. But he then argues that such returns, by aggregation, become thin tailed —in other words, when we look at yearly, not daily, data. To the question "do annual returns behave as Taleb would predict [sic] correlated, with dispersion that shifts too rapidly to have a finite variance?", he offers a stark "no"—offering as evidence the past 195 years of annual equity market returns which seem to disconfirm such theory. In sum, what he offers is *N*=195 years of annual returns as a test of his theory of aggregation (which is, no doubt, afflicted by the survivorship bias associated with a continuous financial market that has existed while many have failed but we can ignore the point for now).

Alas, we believe that the data that he produces is not just entirely compatible with presence of fat tails under aggregation, but what one should expect, as small sample effects may mask the tail effects, *regardless* of whether or not the data has infinite or finite variance. In particular, Treynor (and others) incorrectly rely on the central limit theorem, which states that the distribution of the sum of scaled random variables independently drawn from a finite-variance distribution will converge to a Gaussian distribution. However, the central limit theorem is only valid *asymptotically*, and in finite samples, the *center* of the distribution converges towards a Gaussian distribution before the *tails.* This is a repeat of the mistake made by officer (1972), debunked in Taleb (2009a) which showed how Kurtosis fails to decline under aggregation, once one is aware of the small sample effect.

Let us define a power-law as, for extreme values, the ratios of exceedance probabilities are scale invariant. So take X a random variable, we have a power-law if for x large enough, $P[X>x] \sim O[x^{\alpha}]$.

For aggregation of random variables drawn from distributions with power law tails (and finite variance) of, say, a power law exponent α of 3, the width of the Gaussian component is proportional to $\sigma N^{1/2}(\ln N)^{1/2}$ (Bouchaud and Potters, 2003; Sornette, 2004). The growth in width can be thought of deriving from two components – that inherent growth of the standard deviation of the sum from addition of independent random variables—which constitutes the body of the distribution--and then a factor of $(\ln N)^{1/2}$ that represents any taming of the tails. Further, using extreme value theory, the maximum of a random variable is not affected much by aggregation as the aggregation effect is very slow, even in the presence of finite variance (Mandelbrot and Taleb, 2010).

Clearly we have had for sometime —since Mandelbrot (1963) —sufficient evidence that daily returns in finance and commodities markets are power-law distributed, or, to say the least, insufficient evidence to *reject* it; this fact has been particularly rejuvenated by research conducted by the "Econophysics" groups. Almost all empirical results describe returns even when they have finite variance, as power-law distributed (Sornette, Gabaix et al.,2003, Stanley et al.,2000). Further, Weron (2001) showed that infinite variance processes can masquerade as having an $\alpha>2$, again, from small sample effects.

In part because of the influence of research on non-Gaussian stable distributions, it seemed crucial that fat tails were paired with the focus on values of $\alpha<2$, where the variance is "infinite". But Taleb (2009b) showed that the requirement for power-law effects such as fat tails is not the same requirement of finite or infinite variance: power-laws are power-laws, and their mere acceptance invalidates the results of Markowitz (1959) portfolio theory. Simply, power laws have infinite moments higher than α, meaning that even if variance is the highest moment needed

for the Markowitz derivations, this is merely an asymptotic property that breaks down anywhere before infinity. The results of Markowitz cannot accommodate power laws, finite or infinite variance (though derivatives pricing is not affected at all by such argument).

The empirical tests in the literature point to a "cubic" power-law, with an α around 2.7.

Let us look at the properties of aggregated (log) returns and determine whether the return properties in Treynor (2011) may be consistent with fat tails at the weekly level or the annual level.

Let x be daily log returns, and take x randomly generated following the distribution with tail exponent α= 2.7

$$\frac{\left(\frac{\alpha}{x^2+\alpha}\right)^{\frac{1+\alpha}{2}}}{\sqrt{\alpha}\ \text{Beta}\left[\frac{\alpha}{2},\frac{1}{2}\right]}$$

we take the sum to get y, the annual return,

$$y_{j,z} = \sum_{i=1}^{252} x_{i,z}$$

and

$$Y_z = \left\{y_{j,z}\right\}_{j=1}^{195}$$

We calculate the kurtosis of the distribution of annual returns for a given run *z*

$$K(z) = \frac{\frac{1}{195}\sum_{j=1}^{195}\left(y_{j,z}-\bar{y}_z\right)^4}{\left(\frac{1}{195}\sum_{j=1}^{195}\left(y_{j,z}-\bar{y}_z\right)^2\right)}$$

Finally, we end with the vector of length M

$$\left\{K(z)\right\}_{z=1}^{M}$$

Now by standard result, the expectation of K(z) is infinite. But small (by which we mean finite) samples may yield a different result, which may obscure the unboundness of the kurtosis for this distribution. We therefore conduct a sampling exercise.

We generated close to 4 billion, 3.82 10^7 daily runs of *x*, for a total M of 78 10^3. The median kurtosis observed is 3.36. For a purely Gaussian process the median kurtosis observed for a sample of 200 is 2.95; the probability of observing a kurtosis of 3.36 for a sample size of 200 is approximately .88. Thus, even with infinite higher moments, we may find that observed moments may be consistent with the moments from "thin" tailed distributions.

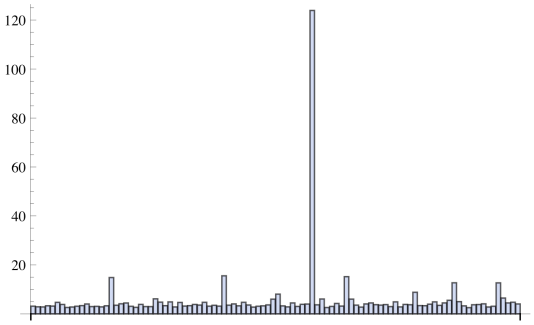


Figure 1— The "small" Monte Carlo run illustrating the distribution of Kurtosis, here M=100.

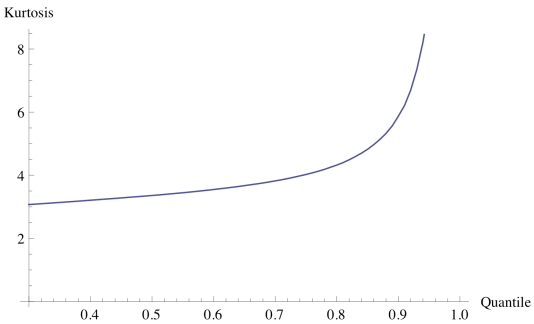


Figure 2, Distribution M=78,000 from 3.8 billion simulations of daily returns x

Data: 3.8 10^7
Median: 3.35
Mean:4.8231
Quantile (.98): 19.94
Max: 189.84

What is one of many the lesson from this exercise? Moment-based empirical reasoning is an unreliable basis for determining the distributional properties of the underlying return generating process. If our concern is risk, then we should focused on tail-regarding measures of risk, rather than take misplaced comfort in moment-based measures.

We note further, that for an N of 52, we have a width that is 1.3 times the aggregate standard deviation, thus indicating even for finite variance distributions, under power-laws the rate at which tails become Gaussian is very slow relative to sample sizes. Even taking 195x52 weeks of data, or more than 10,000 observations, the Gaussian component of the distribution is only 2.0 times the aggregate standard deviation. Assuming aggregation of daily data, we have the Gaussian

component of the distribution through 2.16 times the aggregate standard deviation.

Conclusion

Treynor (2011) argues from historical data and an imprecise interpretation of the central limit theorem that equity market return data is, when aggregated, sufficiently well behaved to employ volatility-based risk measures derived from price returns. However, we argue that the line of reasoning he pursues – namely misplaced reliance on the central limit theorem – and the evidence he presents (195 annual returns for equity markets), do not in any way contradict the presence of fat-tailedness in financial market returns, even when aggregated to be considered annually. Moreover, we argue that moment-based measures of risk are insufficient for, particularly for passive/static exposures to, financial market assets.

The harder conclusion is that portfolio theory (and other results flowing from it) are alas unusable in the real world, whether for daily, monthly, annual, or centennial returns .

***

Boucheaud, J.-P. and M. Potters, 2003, *Theory of Financial Risks and Derivatives Pricing, From Statistical Physics to Risk Management,* 2 nd Ed., Cambridge University Press.

Gabaix,X., P. Gopikrishnan, V.Plerou & H.E. Stanley, 2003, A theory of power-law distributions in financial market fluctuations, *Nature,* 423, 267-270.

Markowitz, Harry, *Portfolio Selection: Efficient Diversi*fication of Investments, Wiley (1959).

Mandelbrot, B. 1963, The variation of cer- tain speculative prices. *The Journal of Business,* 36(4):394–419.

Mandelbrot, B. and Taleb, N. N. (2010) "Random Jump, not Random Walk", in The Known, the Unknown, and the Unknowable, Frank Diebold, Neil Doherty, and Richard Herring,eds., Princeton University Press

Officer, R. R., 1972, "The Distribution of Stock Returns." *Journal of the American Statistical Association* 340(67): 807–812.

Sornette, D., 2004, 2 nd Ed., *Critical Phenomena in Natural Sciences, Chaos, Fractals, Self-organization and Disorder: Concepts and Tools,* (Heidelberg: Springer)

Stanley, H.E. L.A.N. Amaral, P. Gopikrishnan, and V. Plerou, 2000, Scale invariance and universality of economic fluc- tuations, *Physica A,* 283,31-41

Taleb, N. N. (2009a). "Errors, robustness, and the fourth quadrant." *International Journal of Forecasting* 25(4): 744-759.

Taleb, N. N. (2009b), Finiteness of variance is irrelevant in the practice of quantitative finance. *Complexity*, 14: 66–76

Treynor, Jack L. (2011), What Can Taleb Learn From Markowitz, *Journal Of Investment Management,* Vol. 9, No. 4, (2011), pp. 5–9

Weron, R., 2001,Levy-stable distributions revisited: tail index > 2 does not exclude the Levy-stable regime. *International Journal of Modern Physics C* (2001) 12(2), 209-223.

Notes: Platonic Convergence and the Central Limit Theorem

1) An erroneous notion of limit:

Take the standard formulation of the Central Limit Theorem (Feller 1971, Vol. II;Grimmet & Stirzaker, 1982):

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with mean m & variance σ^2 satisfying $m < \infty$ and $0 < \sigma^2 < \infty$, then

$$\frac{\sum_{i=1}^{N} X_i - N m}{\sigma \sqrt{N}} \xrightarrow{D} \text{Gaussian}(0, 1) \text{ as } N \to \infty$$

Where $\xrightarrow{D}$ is converges "in distribution".

Taking convergence for granted provides a plain illustration of the severe disease of Platonicity --or working backwards from theory to practice. Effectively we are dealing with a double problem.

1) The first, as uncovered by Jaynes, comes from to the abuses of formalism & measure theory:

- Jaynes 2003 (p.44):"The danger is that the present measure theory notation presupposes the infinite limit already accomplished, but contains no symbol indicating which limiting process was used (...) Any attempt to go directly to the limit can result in nonsense".

  Granted Jaynes is still too Platonic in general and idealizes his convergence process (he also falls headlong for the Gaussian by mixing thermodynamics and information). But we accord with him on this point --along with the definition of probability as information incompleteness, about which in later sessions.

2) The second problem is that we do not have a "clean" limiting process --the process cannot be idealized. It is very rare to find permanent idealized conditions that allow for temporal aggregation.

Now how should we look at the Central Limit Theorem? Let us see how we arrive to it assuming "independence".

2) The Problem of Convergence

The CLT works in a specific way:It does not fill-in uniformily,but in a near-Gaussian way--indeed,disturbingly so. Simply,whatever your distribution (assuming one mode),your sample is going to be skewed to deliver more central observations,and fewer tail events.The consequence is that,under aggregation,the sum of these variables will converge "much" faster in the body of the distribution than in the tails. As *N*,the number of observations increases,the Gaussian zone should cover more grounds... but not in the "tails".

You can see it very easily with two very broad uniform distributions,say with a lower bound a and an upper bound *b*, *b-a* very large. As you convolute, you will see the peakedness in the center, which means that more observations will fall there (see Appendix).

This quick note shows the intuition of the convergence and presents the difference between distributions. (See Appendix)

Take the sum of of random independent variables X_i with **finite variance** under distribution $\varphi(X)$. Assume 0 mean for simplicity (and symmetry, absence of skewness to simplify).

A better formulation of the Central Limit Theorem (Kolmogorov et al,x)

$$P\left[-u \le Z = \frac{\sum_{i=0}^{n} X_i}{\sqrt{n}\ \sigma} \le u\right] = \frac{1}{\sqrt{2\,\pi}} \int_{-u}^{u} e^{-\frac{z^2}{2}}\, dZ$$

So the distribution is going to be:

$$\left(1 - \int_{-u}^{u} e^{-\frac{z^2}{2}}\, dZ\right) \text{for} -u \le z \le u$$

inside the "tunnel" [-u,u] --the odds of falling inside the tunnel itself

and

$$\int_{-\infty}^{u} \varphi'[n]\,(Z)\,dz + \int_{u}^{\infty} \varphi'[n]\,(Z)\,dz$$

outside the tunnel [-u,u]

Where φ'[n] is the n-summed distribution of φ.

How φ'[n] behaves is a bit interesting here --it is distribution dependent. And it depends on the initial distribution!

Bouchaud-Potters Treatment of Width of the Tunnel [-u,u]

(in class derivation)

■ 3) Using Log Cumulants & Observing Gaussian Convergence

The normalized cumulant of order *n*, C(n) is the derivative of the log of the characteristic function ϕ which we convolute N times divided by the second cumulant (i,e., second moment).

$$C(n, N) = \frac{(-i)^n\, \partial^n \log(\phi^N)}{\left(-\partial^2 \log(\phi)^N\right)^{n-1}} /. z \to 0$$

Since C(N+M)=C(N)+C(M), the additivity of the Log Characteristic function under convolution makes it easy to see the speed of the convergence to the Gaussian.

Fat tails implies that higher moments implode --not just the 4th .

Table of Normalized Cumulants -Speed of Convergence (Dividing by σ^n where n is the order of the cumulant).

Distribution	Normal[μ, σ]	Poisson(λ)	Exponential(λ)	$\Gamma(a, b)$
PDF	$\frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\,\sigma}$	$\frac{e^{-\lambda}\,\lambda^x}{x!}$	$e^{-x\lambda}\,\lambda$	$\frac{b^{-a}\,e^{-\frac{x}{b}}\,x^{a-1}}{\Gamma(a)}$
N−convoluted Log Characteristic	$N\log\left(e^{i\,z\,\mu-\frac{z^2\,\sigma^2}{2}}\right)$	$N\log(e^{(-1+e^{i\,z})\,\lambda})$	$N\log(\frac{\lambda}{\lambda-i\,z})$	$N\log((1-i\,b\,z)^{-a})$
2nd Cum	1	1	1	1
3 rd	0	$\frac{1}{N\,\lambda}$	$\frac{2\,\lambda}{N}$	$\frac{2}{a\,b\,N}$
4 th	0	$\frac{1}{N^2\,\lambda^2}$	$\frac{3!\,\lambda^2}{N^2}$	$\frac{3!}{a^2\,b^2\,N^2}$
5 th	0	$\frac{1}{N^3\,\lambda^3}$	$\frac{4!\,\lambda^3}{N^3}$	$\frac{4!}{a^3\,b^3\,N^3}$
6 th	0	$\frac{1}{N^4\,\lambda^4}$	$\frac{5!\,\lambda^4}{N^4}$	$\frac{5!}{a^4\,b^4\,N^4}$
7 th	0	$\frac{1}{N^5\,\lambda^5}$	$\frac{6!\,\lambda^5}{N^5}$	$\frac{6!}{a^5\,b^5\,N^5}$
8 th	0	$\frac{1}{N^6\,\lambda^6}$	$\frac{7!\,\lambda^6}{N^6}$	$\frac{7!}{a^6\,b^6\,N^6}$
9 th	0	$\frac{1}{N^7\,\lambda^7}$	$\frac{8!\,\lambda^7}{N^7}$	$\frac{8!}{a^7\,b^7\,N^7}$
10 th	0	$\frac{1}{N^8\,\lambda^8}$	$\frac{9!\,\lambda^8}{N^8}$	$\frac{9!}{a^8\,b^8\,N^8}$

Distribution	Mixed Gaussians (Stoch Vol)	StudentT(3)	StudentT(4)	□
PDF	$p\,\frac{e^{-\frac{x^2}{2\,{\sigma_1}^2}}}{\sqrt{2\pi}\,\sigma_1}+(1-p)\,\frac{e^{-\frac{x^2}{2\,{\sigma_2}^2}}}{\sqrt{2\pi}\,\sigma_2}$	$\frac{6\sqrt{3}}{\pi\,(x^2+3)^2}$	$12\,(\frac{1}{x^2+4})^{5/2}$	□
$N-$convoluted Log Characteristic	$N\log\left(p\,e^{-\frac{z^2\,{\sigma_1}^2}{2}}+(1-p)\,e^{-\frac{z^2\,{\sigma_2}^2}{2}}\right)$	$N\left(\log\left(\sqrt{3}\,\lvert z\rvert+1\right)-\sqrt{3}\,\lvert z\rvert\right)$	$N\log(2\,\lvert z\rvert^2\,K_2(2\,\lvert z\rvert))$	□
2nd Cum	1	1	1	□
3 rd	0	Ind	□	□
4 th	$-\left(3\,(-1+p)\,p\,(\sigma_1^2-\sigma_2^2)^2\right)/\left(N^2\,(p\,\sigma_1^2-(-1+p)\,\sigma_2^2)^3\right)$	Ind	Ind	□
5 th	0	Ind	Ind	□
6 th	$\left(15\,(-1+p)\,p\,(-1+2\,p)\,(\sigma_1^2-\sigma_2^2)^3\right)/\left(N^4\,(p\,\sigma_1^2-(-1+p)\,\sigma_2^2)^5\right)$	Ind	Ind	□

Note: On "Infinite Kurtosis"- Discussion

Note on Chebyshev's Inequality and upper bound on deviations under finite variance

A lot of idiots talk about finite variance not considering that it still does not mean much. Consider Chebyshev's inequality:

$$P[X > \alpha] \leq \frac{\sigma^2}{\alpha^2}$$

$$P[X > n\,\sigma\,] \leq \frac{1}{n^2}$$

Which effectively accommodate power laws but puts a bound on the probability distribution of large deviations --but still significant.

The Effect of Finiteness of Variance

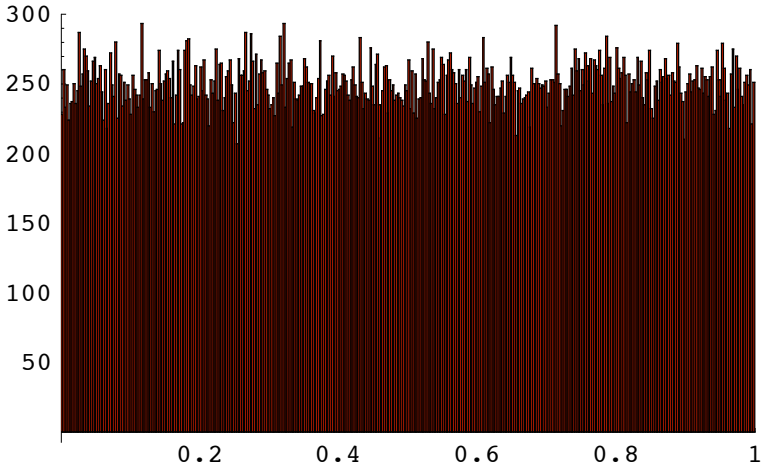
This table shows the probability of exceeding a certain σ for the Gaussian and the lower on probability limit for any distribution with finite variance.

Deviation	Gaussian	Chebyshev Upper Bound
3	$7.\times 10^{2}$	9
4	$3.\times 10^{4}$	16
5	$3.\times 10^{6}$	25
6	$1.\times 10^{9}$	36
7	$8.\times 10^{11}$	49
8	$2.\times 10^{15}$	64
9	$9.\times 10^{18}$	81
10	$1.\times 10^{23}$	100

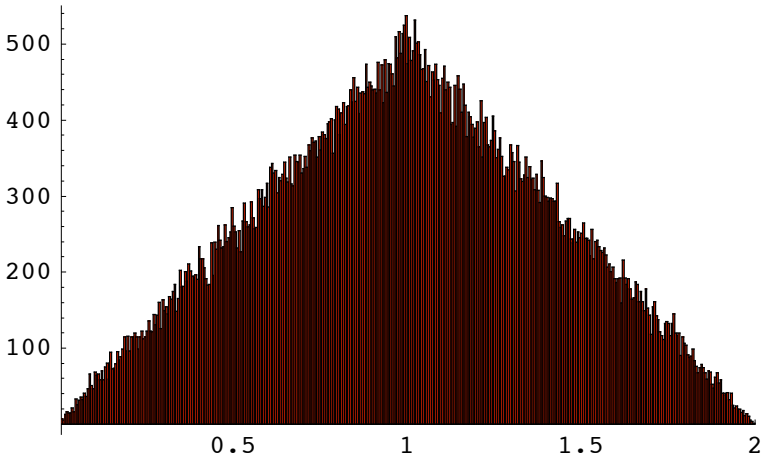
■ Calculations

■ Calculations

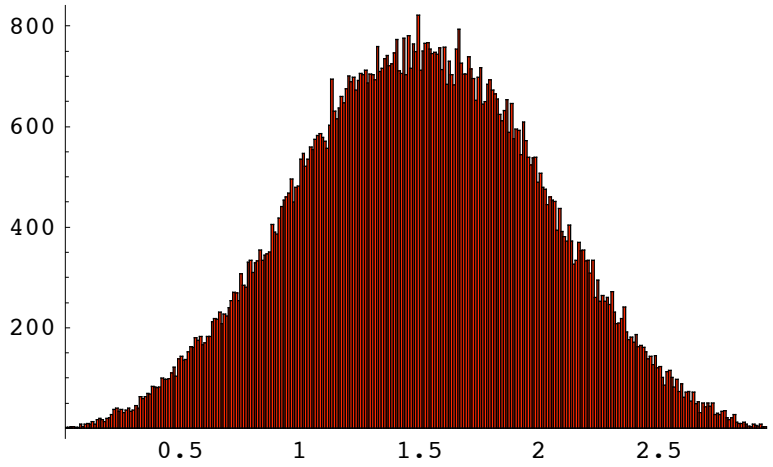
■ APPENDIX to 2- How We Converge Mostly in the Center - A Tutorial



N=2



N=3



Out[417]= - Graphics -

Example: Uniform Distribution

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \end{cases}$$

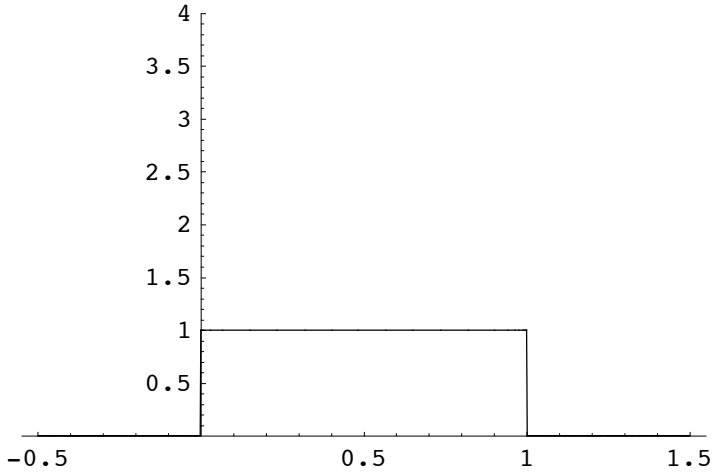
By Convoluting 2, 3, 4 times iteratively:

$$f2(z2) = \int_{-\infty}^{\infty} (f(z-x))\,(f\,x)\,dx = \begin{cases} 2 - z2 & 1 < z2 < 2 \\ z2 & 0 < z2 \le 1 \end{cases}$$

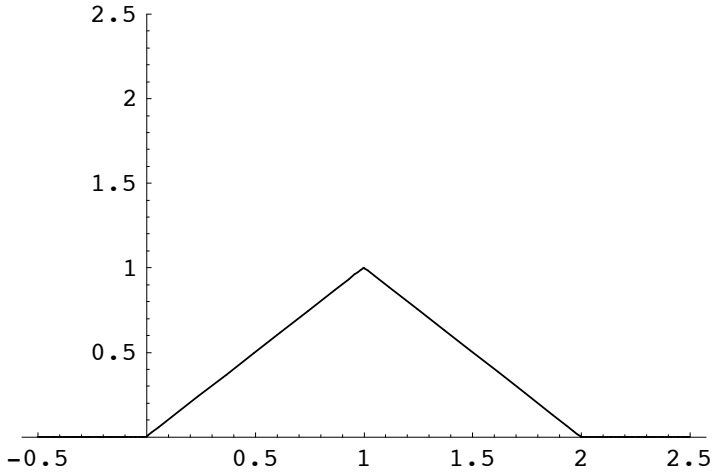
$$f3\,z3 = \int_{0}^{3} (f2(z3-x2))\,(f\,x2)\,dx2 = \begin{cases} \frac{z3^2}{2} & 0 < z3 \le 1 \\ -(z3-3)\,z3 - \frac{3}{2} & 1 < z3 < 2 \\ -\frac{1}{2}(z3-3)(z3-1) & z3 = 2 \\ \frac{1}{2}(z3-3)^2 & 2 < z3 < 3 \end{cases}$$

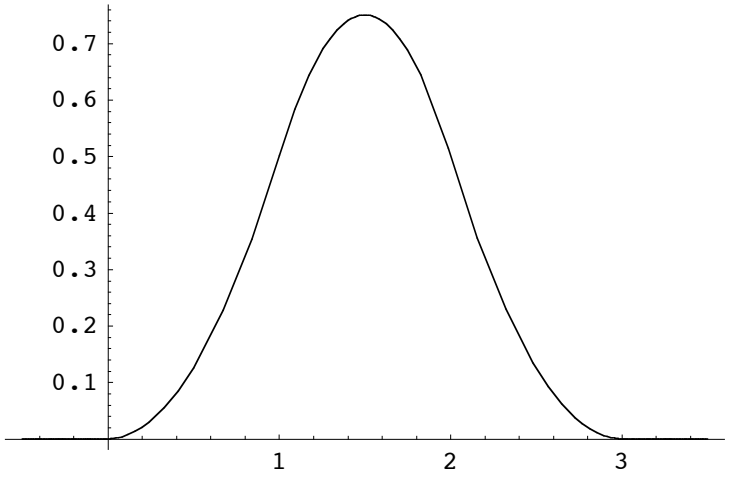
$$f4\,x = \int_{0}^{4} (f3(z4-x))\,(f\,x3)\,dx3 = \begin{cases} \frac{1}{4} & z4 = 3 \\ \frac{1}{2} & z4 = 2 \\ \frac{z4^2}{4} & 0 < z4 \le 1 \\ \frac{1}{4}(-z4^2 + 4\,z4 - 2) & 1 < z4 < 2 \vee 2 < z4 < 3 \\ \frac{1}{4}(z4-4)^2 & 3 < z4 < 4 \end{cases}$$

A simple Uniform Distribution



We can see how quickly, after one single addition, the net probabilistic "weight" is going to be skewed to the center of the distribution, and the vector will weight future densities..





Benoit Mandelbrot and Nassim Nicholas Taleb

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: $P_{>x} = K\, x^{-a}$ where $P_{>x}$ is the probability of exceeding a variable x and a 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 a. This maximum is well-known to behave like $n^{1/a}$. 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 a–2.

If a > 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 infinite 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 finds that the relative contribution of the largest addend behaves like $n^{1/a-½}$. In the example of a=3, this becomes $n^{-1/6}$. Again asymptotically for n®¥, this ratio tends to 0 — as expected — but the convergence is exquisitely slow. For comparison, examine for EX≠0 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≠0 — 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 a<2 for which the variance is infinite. The maximum's behavior is still $n^{1/a}$, but the — subtracting nEX —sum's behavior changes from $n^{1/2}$ to the "anomalous"$n^{1/a}$. Therefore, the relative contribution of the largest addend is of the order $n^{1/a-1/a}=n^0$. 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 a=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 a<2 but preasymptotically it continues to hold if a is not far above 2.

Lecture:The fundamental problem of the 0^{th} moment and the irrelevance of "naked probability"

The Nonbinary problem: Decisions (by humans) are rarely made based on probability except in the case of strictly binary bets: those on win/lose, in which the agent is focused on the outcome of naked probability $\int_D p(x)\,dx$ rather than $\int_D f(x)\,p(x)\,dx$ where f(x) is a function of the random variable and *D* the area of integration. So for the expectation, i.e. "impact", most common criterion of concern, f(x)= x.

The agent might discuss the "probable" and "improbable" but not know that he does not really mean it. It is just a proxy for something else -"consequential" or "inconsequential".

Note:We will see (section x) that we do not care about some part of the 0^{th} moment *(probability),* burt some part of the first moment, and **just that** --or some complicated function of it, causing the dependence on the $\mathcal{L}^1$ norm. We have no reasons (except computational) to worry about $\mathcal{L}^2$ or higher ones (higher moments). Anyway, we will be using f for the expectation of scaling of the outcomes. $f(x) = |x|$ (mean deviation) or $f(x)=x^2$ for the variance, etc., for higher moments, One can even include some "utility" function as part of *f*, whatever that means --it is not necessary as it can be embedded in p(x).

One aspect of this irrelevance of probability is that "fatter tails" does not necessarily mean a higher **incidence** *(i.e. frequency) of rare events; it means a higher* **contribution** *of these events which generally corresponds to a lower incidence of* **some** *tail events (and a rise of others further out in the distribution). Given the same scaling, a higher fourth moment "fatter tails" decreases the probability of exceeding K, i.e.,* $\int_K^\infty f(x)\,dx$ *while increasing the contribution* $\int_K^\infty x\,f(x)\,dx$

Example: Naive Fattening of the Gaussian

Create a naive fat-tailed Gaussian. We pick a dual Gaussian mixture, both mixes equiprobable ($\frac{1}{2}$) with a "low" variance $(\sigma(1-v))^2$ and a "high" one $\left(\sigma\sqrt{-v^2+2v+1}\right)^2$ selecting a single v so that the total variance remains the same. With 1 >v≥ 0, the total standard deviation

$$\sigma = \sqrt{\frac{1}{2}\left((\sigma(1-v))^2 + \left(\sigma\sqrt{-v^2+2v+1}\right)^2\right)}$$

v=0, v= 1/2 fattens the tails up to 1 standard deviation

Illustration: Time spent in the "tunnel" between -1 and 1 "sigmas" for the deterministic and mild Gaussian mixture

We can see that as v increases (therefore volatility is more stochastic), the time spend betwen +1 and -1 standard deviations increases. So events , like P>1σ , with 16% probability have actually 12% of occuring.

	v	Time ± 1 std
1	0	0.682689
2	0.1	0.687089
3	0.2	0.698764
4	0.3	0.715553
5	0.4	0.73477
6	0.5	0.752404
7	0.6	0.763293

Beyond some "sigma" the effect reverses -- here rather quickly: 3 standard deviations. So fatter tails imply fewer 1 sigma events, and more 3 sigma ones. Simply, we are not dealing with very fat tails as these do not fill out too far outside the central region.

	v	Time ± 3 MAD
1	0	0.983319
2	0.1	0.98198
3	0.2	0.978556
4	0.3	0.973975
5	0.4	0.969164
6	0.5	0.964807
7	0.6	0.961187

Stopping Time & Fattening of the tails of a Brownian Motion

Consider the distribution of the time it takes for a continuously monitored Brownian motion S to exit from a "tunnel" with a lower bound L and an upper bound H. Counterintuitively, fatter tails makes an exit (at some sigma) take longer. You are likely to spend more time inside the tunnel --since exits are far more dramatic.

ψ is the distribution of exit time *t,* where $t \equiv inf\{t: S \notin [L,H]\}$

From Taleb (1997) we have the following approximation

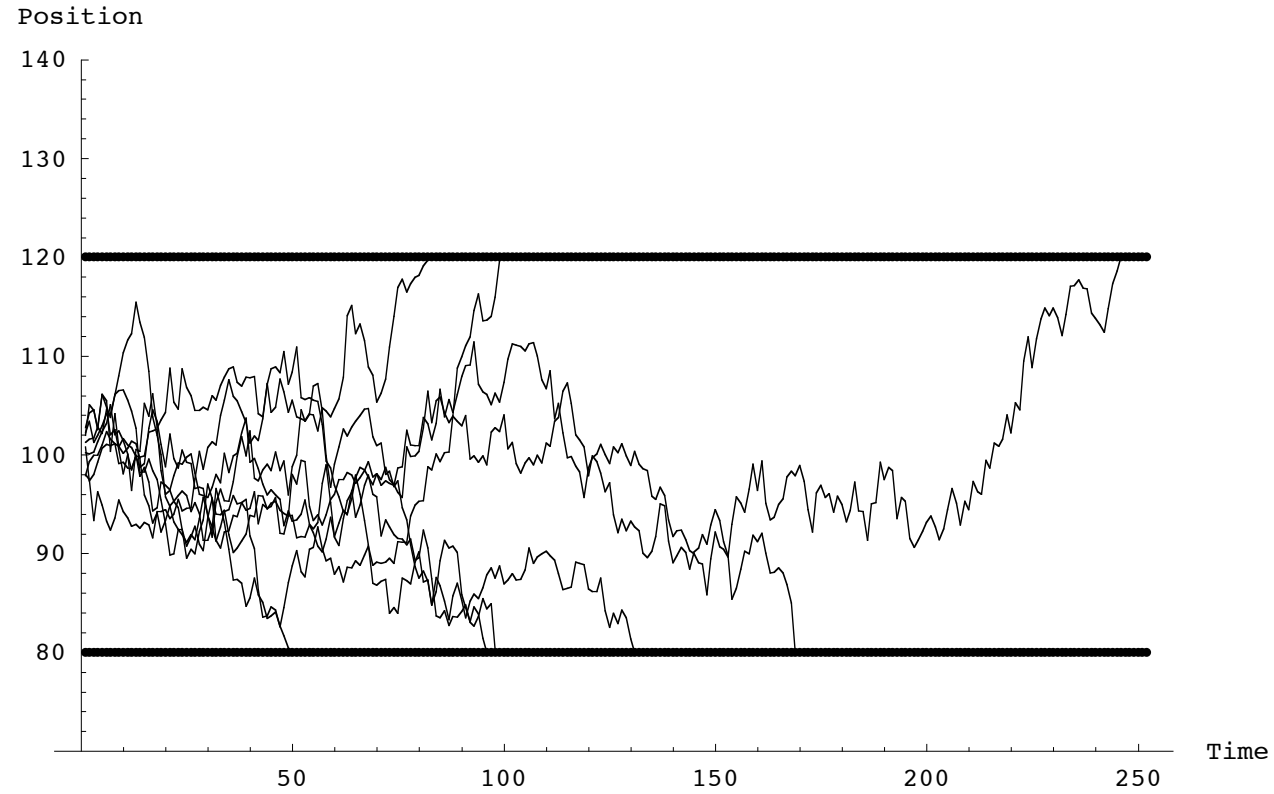
$$\psi(t \mid \sigma) = \frac{1}{(\log(H) - \log(L))^2}$$

$$\left(e^{-\frac{1}{8}(t\sigma^2)}\, \pi \sigma^2 \sum_{n=1}^{m} \frac{(-1)^n\, e^{-\frac{n^2 \pi^2 t \sigma^2}{2(\log(H)-\log(L))^2}}\, n \sqrt{S} \left(\sqrt{L} \sin\left(\frac{n\pi(\log(L)-\log(S))}{\log(H)-\log(L)} \right) - \sqrt{H} \sin\left(\frac{n\pi(\log(H)-\log(S))}{\log(H)-\log(L)} \right) \right)}{\sqrt{H}\sqrt{L}} \right)$$

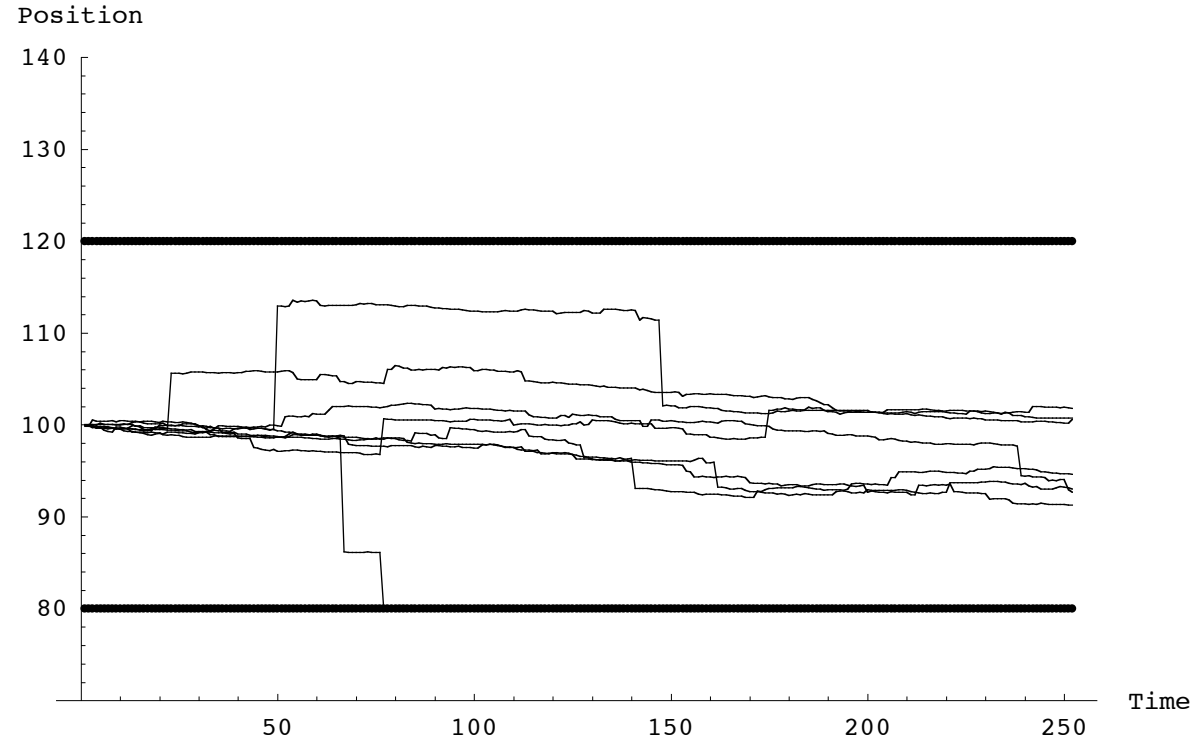
and the fatter-tailed distribution from mixing Brownians with σ^2 separared by a coefficient v:

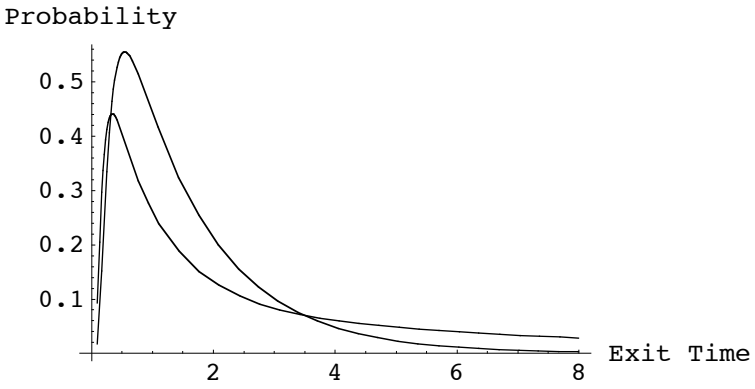
$$\psi(t \mid \sigma, v) = \frac{1}{2} p(t \mid \sigma(1-v)) + \frac{1}{2} p\left(t \,\middle|\, \sigma\sqrt{-v^2+2v+1}\right)$$

Stochastic paths terminating upon hitting barriers H (high) H=120 and L (low) L=80. Time to exit is extended by the fattening of the tails.

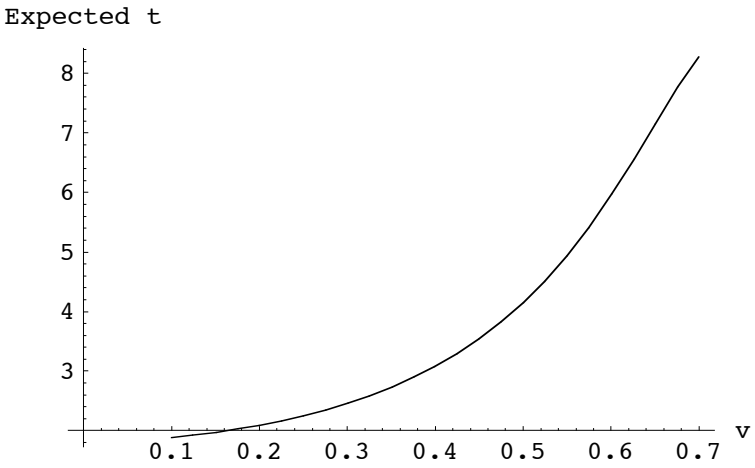


With very fat tails (almost-Cauchy)





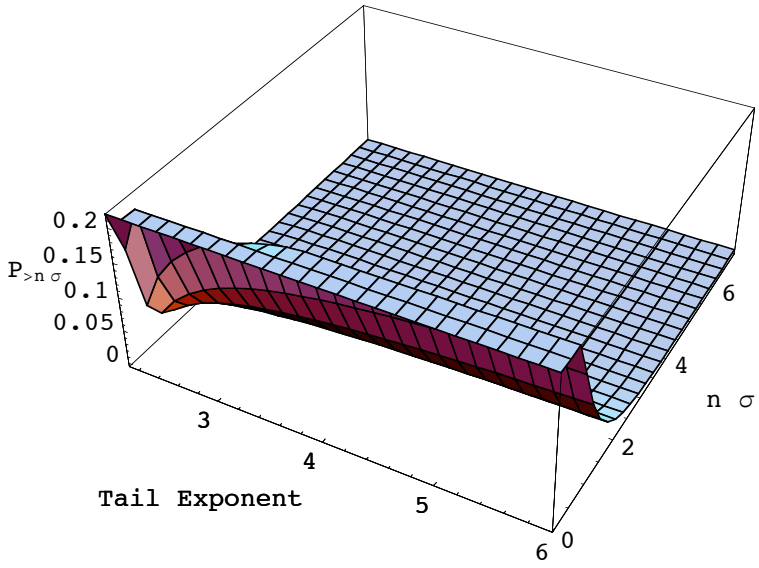
Expected stopping time explosion with v



The Implication: Visibly it "takes longer" to capture the statistical properties. How much "longer"? I don't know --it is, as we saw earlier, an inverse problem.

Fatter tails, or "truer" fat tails

Here I use a finite variance scalable distribution. Clearly at some exponent, the probability of exceeding drops for for n σ low and rises for higher n σ



Discrete Calc

Derivatives, Prediction and *True* Fat Tails (i.e. Fractal), Part 1: The Fragility of Option Pricing

Nassim Nicholas Taleb

This is a working notebook --it cannot be quoted in this present version.

Derivatives that depends on the high consequential large deviation are marred with huge sampling error. I examine the sensitivity of the derivatives to the parameters, the sampling error of the estimations or "predictions", then look at the empirical stability of these parameters.
Organization: First 1) I do the math of distribution & derivatives, as there is no intelligent literature on the subject outside of the inapplicable Levy-Stable, 2) I show the magnitudes errors w.r. to some parameters (mainly the tail exponent α) 3) I discuss the error in the estimation of these parameters.
Main point: For options on remote events, a small change in the tail exponent say α betwen 1.5 and 2, well within the estimation errors, make the option change in value: a .5 change in exponent makes the error on the event vary by a factor >10, often >100. Moral: don't play with tail estimations, and don't believe that options can estimate anything.

.

1- True Fat Tails and Derivatives Pricing

Definition: true fat tails (see lecture x) are as follows $P_{>nX}/P_{>X}$ depends on n, not X for X large enough.

First, we select a distribution without a tail-characteristic scale for x on the real line -∞ and ∞, which consists in a fractal tails with exponent α and a multiplying scale. Typical Student T

$$\phi(u) = \frac{1}{\sqrt{\alpha}\,\beta\left(\frac{\alpha}{2},\frac{1}{2}\right)}\left(\frac{\alpha}{\alpha+u^2}\right)^{\frac{1+\alpha}{2}},\quad u\in[-\infty,\infty],\ \alpha\geq 1$$

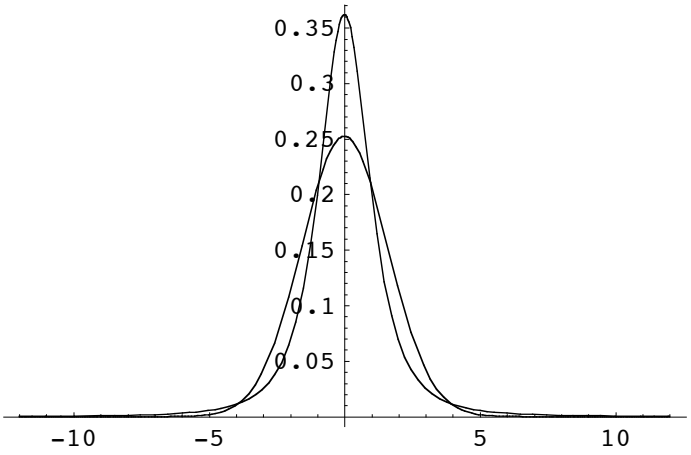
So for large u "in the tails", we can see that it behaves K $u^{-\alpha-1}$, where K is a constant.

Where $\beta(.)$ is **Euler beta function** $\beta(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}\,(1-t)^{b-1}\,dt$.

ϕ(.) has fractal tails with exponent α on both sides.
Note: I ignore the designation "Levy-stable"

α= 5/2; Comparison with a Gaussian N(0,$\sqrt{5/2}$)

Comparison with an equivalent Gaussian



Now consider the multiplicative (monoperiodic) process X= X_0 (1+s u -c), where u is ϕ distributed. $\frac{X-X0-cX0}{X0}\frac{1}{s}$ is a straight relative price change with a "drift" term c and a "dispersion" constant s to scale by the "volatility", simplified as a multiple of mean deviation (for a given period between an initial 0 and T.) The problem is that we cannot take a fractal tailed distribution for Log[$\frac{X}{X_0}$] for obvious reasons (too unwieldy; I tried), so we have to be content with relative price changes.

By change of stochastic variables, I am able to get the distribution of X, conditional on X_0.

(If x has distribution f then y=z(x) has density $\frac{f(g(x))}{f'(g(x))}$ where g is the inverse function of z).

$$f(X) = \frac{1}{\sqrt{\alpha}\, s\, X_0\, \beta(\frac{\alpha}{2}, \frac{1}{2})} \left(\frac{\alpha}{\frac{(X - C\, X_0 - X_0)^2}{s^2\, v^2} + \alpha} \right)^{\frac{\alpha+1}{2}}, \; X > 0$$

Caveat 1 and Renormalization: The distribution *f(X)* may have minutely small mass for *X*<0, when *(1+s u)* turns negative, *s u< 1*. This requires an atrocially huge volatility and can be compensated by a truncating effect and renormalization of the mass with f'[x]= f(x) $\frac{1}{1-\int_{-\infty}^{0} f(X)\, dX}$. I left it out as it does not affect the exercise.

Indeed

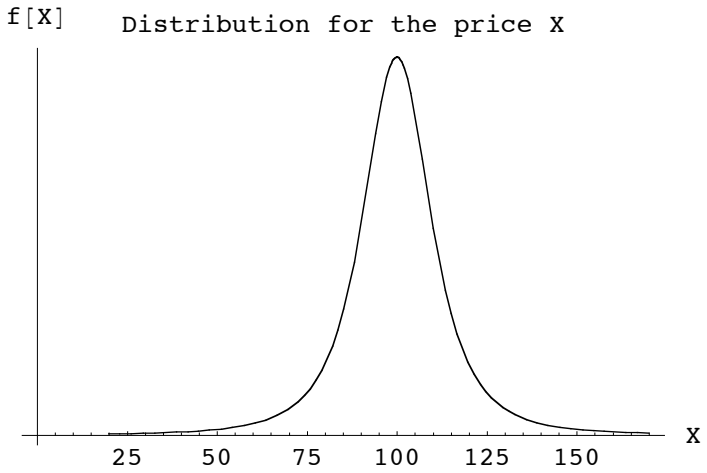
$$\int_{-\infty}^{0} f(X) \Big|_{\alpha=2} dX = \frac{1}{2} - \frac{1}{2\sqrt{1+2s^2}}$$

and

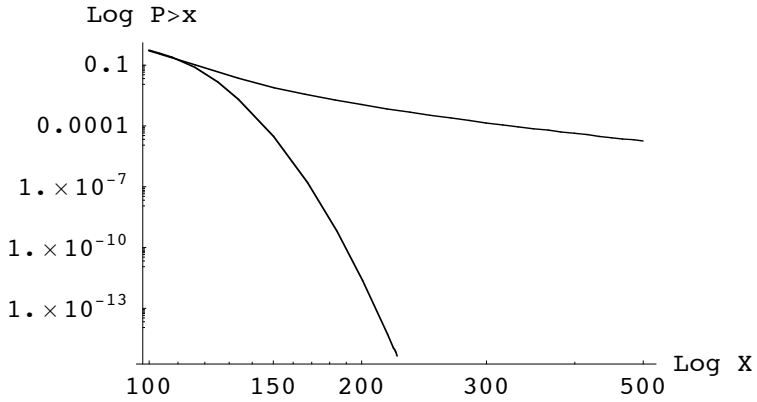
$$\int_{-\infty}^{0} f(X) \mid_{\alpha=3} dX\, (6s^2+2) \sin^{-1}\left(\frac{\sqrt{\sqrt{3s^2+1}-1}}{\sqrt{2}\, \sqrt[4]{3s^2+1}} \right) - \frac{\sqrt{3}\, s}{3\pi s^2 + \pi}$$

Which is very small: of the order of < .01% for high volatility environments when α=3, and .05% when α=2 --thus justifying ignoring the renormalization.

Caveat 2 and Explosive Mean: Likewise the mean may become explosive upwards, in which case the compensation can be part of the drift c just like the lognormal is compensated by a negative $\frac{-1}{2}\ \sigma^2$ (where σ is the Gaussian standard deviation. But,for the purposes of the exercise f(.) works well in addressing option errors.



Scalable as far as meets the eye (unlike the Lognormal)



I was not able to find a close solution except for α=2,3. At α=∞ we get the standard Bachelier-Thorp (a.k.a. Black-Scholes) equation.

Note on moments: With no drift *c*,

Whith finite variance α=3

$$\alpha = 3,$$

$$E[X] = X_0 \left(\frac{2\sqrt{3}\ s + 2\cot^{-1}\left(\sqrt{3}\ s\right) + \pi}{2\pi} \right) \simeq X_0$$

where $\cot^{-1}(z)$ is the Arc Cotangent of z

$$E\left[\left(\frac{X - X0}{X0}\right)^2\right] = \frac{3\,s^2}{2}$$

$$E\left(\left|\frac{X - X0}{X0}\right|\right) = \frac{\sqrt{3}\;s\,(2 + 3\,s^2)}{\pi + 3\,\pi\,s^2}$$

With infinite variance (borderline) α=2

$$\alpha = 2,\ E[X] = \frac{X_0\left(s + \sqrt{2\,s^4 + s^2}\right)}{2\,s} \simeq X_0$$

$$E\left(\left|\frac{X - X0}{X0}\right|\right) = s\left(\sqrt{2} - \frac{s}{\sqrt{1 + 2\,s^2}}\right)$$

■ Call Options Under Different Parametrizations

a- Call Option Price C with a Cubic α

$$\text{Call Price } C = \int_K^\infty (X - K)\,f(X)\,dX \;\Big|\; \alpha = 3$$

$$C_3 = \frac{s\,X0\left(\pi\sqrt{\frac{1}{s^2\,X0^2}}\,(-K + C\,X0 + X0) + 2\sqrt{3}\right) + 2\,(-K + C\,X0 + X0)\cot^{-1}\left(\frac{\sqrt{3}\;s\,X0}{-K + C\,X0 + X0}\right)}{2\,\pi}$$

I apologize for the inelegance but I can't do better

b- Call Option Price with α= 5/2

$$C_{5/2} = \frac{1}{6\,5^{3/4}\,\sqrt{\pi}\,\Gamma(\frac{5}{4})}$$
$$\left(\left(2\sqrt{2}\left(5\sqrt{s\,X0}\,\sqrt[4]{5\,s^2\,X0^2 + 2\,(-K + C\,X0 + X0)^2} + \frac{\sqrt[4]{5}\;s\,X0\,(-K + C\,X0 + X0)^2\,(\zeta 1 + 5)\,\zeta 2}{5\,s^2\,X0^2 + 2\,(-K + C\,X0 + X0)^2}\right) + \frac{5^{3/4}\,\sqrt{\pi}\,(-K + C\,X0 + X0)\,\Gamma(\frac{1}{4})}{\Gamma(\frac{3}{4})}\right)\Gamma\left(\frac{7}{4}\right)\right)$$

$$\zeta 1 = \frac{2\,(-K + C\,X0 + X0)^2}{s^2\,X0^2}$$

$$\zeta 2 = {}_2F_1\left(\frac{1}{2}, \frac{3}{4}; \frac{3}{2}; -\frac{2\,(-K + C\,X0 + X0)^2}{5\,s^2\,X0^2}\right)$$

where ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} (a)_k\,(b)_k / (c)_k\,z^k / k!$.

c- Call Option Price with square α

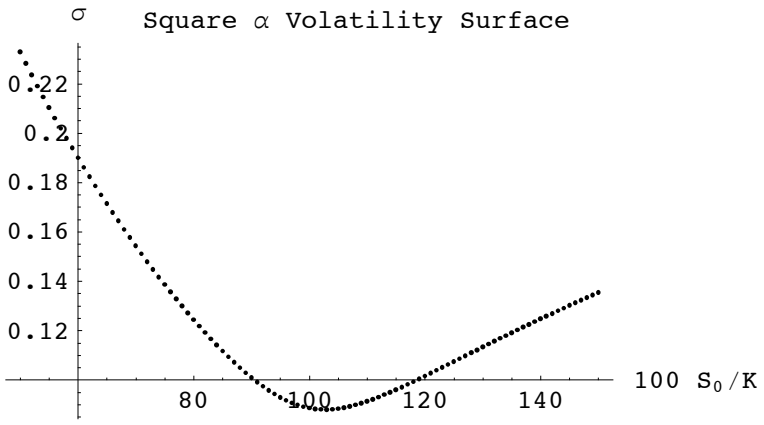
$$C_2 = \frac{1}{2}\left(\zeta 3\, K^2 + \left(-2\,(C+1)\,\mathrm{X0}\,\zeta 3 - 1\right) K + \mathrm{X0}\left(C + \left((C+1)^2 + 2\,s^2\right)\mathrm{X0}\,\zeta 3 + 1\right)\right)$$

where

$$\zeta 3 = \sqrt{\frac{1}{K^2 - 2\,(1+C)\,K\,\mathrm{X0} + \left((1+C)^2 + 2\,s^2\right)\mathrm{X0}^2}}$$

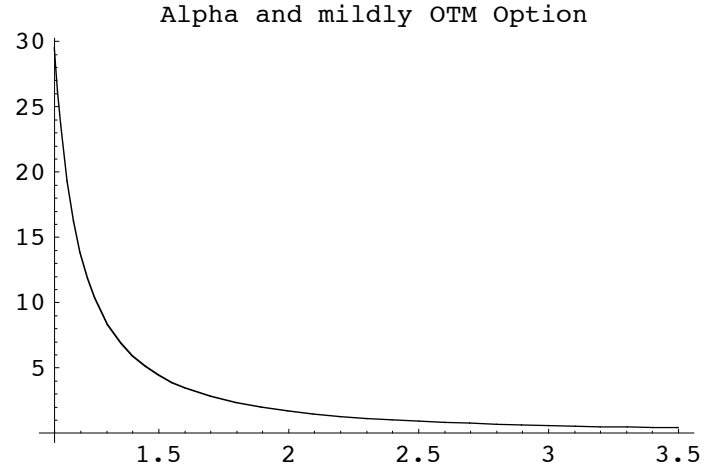
■ Comparison to the Volatility Smile (Bachelier-Thorp, a.k.a. Black Scholes)

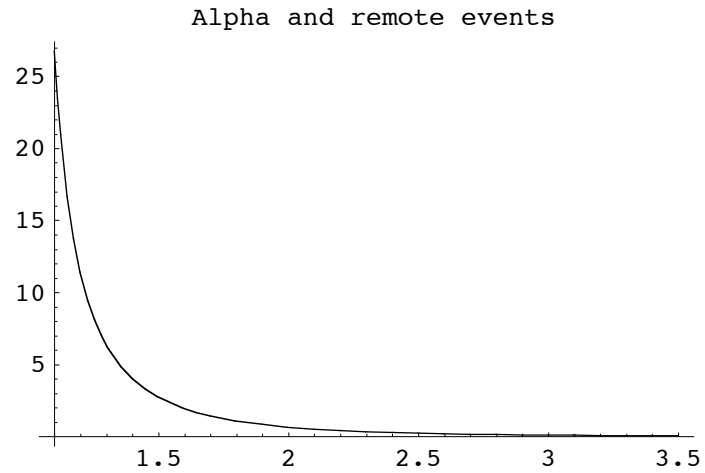
The Infinite Variance Case: $\alpha \leqslant 2$ does not mean anything for option pricing, it generates a volatility surface --so long as the scaling s is calibrated on the absolute first moment.



2- *True* Fat Tails and Derivatives Errors

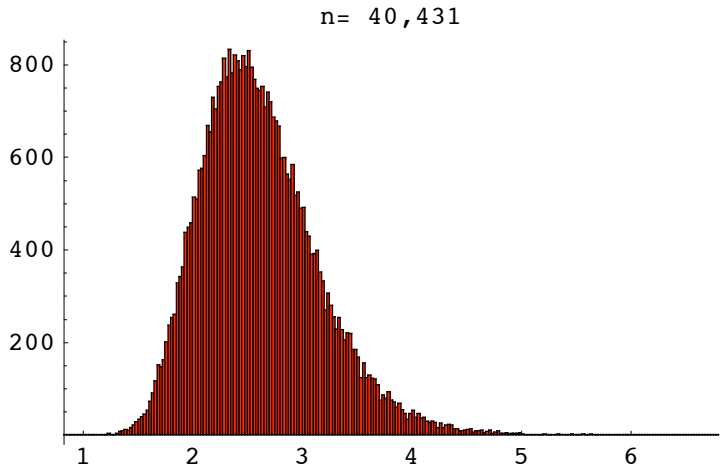
By changing α and maintaining the rest constant, we can do guess the consequence of a small error in the tail exponent on the option value.





Note that almost all options converge to the same price (minus moniness) when alpha drops to close to 1.

Errors in Alpha Estimation



The Mean Deviation= .42 for an estimated alpha of 2.62 (using the Hill Estimator).

ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

international journal of forecasting

www.elsevier.com/locate/ijforecast

Introduction

Decision making and planning under low levels of predictability

Spyros Makridakis[a,*], Nassim Taleb[b,1]

[a] *INSEAD, Boulevard de Constance, 77305 Fontainebleau, France*
[b] *Polytechnic Institute of NYU, Department of Finance and Risk Engineering, Six MetroTech Center, Rogers Hall 517, Brooklyn, NY 11201, USA*

Abstract

This special issue aims to demonstrate the limited predictability and high level of uncertainty in practically all important areas of our lives, and the implications of this. It summarizes the huge body of solid empirical evidence accumulated over the past several decades that proves the disastrous consequences of inaccurate forecasts in areas ranging from the economy and business to floods and medicine. The big problem is, however, that the great majority of people, decision and policy makers alike, still believe not only that accurate forecasting is possible, but also that uncertainty can be reliably assessed. Reality, however, shows otherwise, as this special issue proves. This paper discusses forecasting accuracy and uncertainty, and distinguishes three distinct types of predictions: those relying on patterns for forecasting, those utilizing relationships as their basis, and those for which human judgment is the major determinant of the forecast. In addition, the major problems and challenges facing forecasters and the reasons why uncertainty cannot be assessed reliably are discussed using four large data sets. There is also a summary of the eleven papers included in this special issue, as well as some concluding remarks emphasizing the need to be rational and realistic about our expectations and avoid the common delusions related to forecasting.

Keywords: Forecasting; Accuracy; Uncertainty; Low level predictability; Non-normal forecasting errors; Judgmental predictions

1. Introduction

The unknown future is a source of anxiety, giving rise to a strong human need to predict it in order to reduce, or ideally eliminate, its inherent uncertainty. The demand for forecasts has created an ample supply of "experts" to fulfill it, from augurs and astrologists to economists and business gurus. Yet the track record of almost all forecasters is dismal. Worse, the accuracy of "scientific" forecasters is often no better than that of simple benchmarks (e.g. today's value, or some average). In addition, the basis of their predictions is often as doubtful as those of augurs and astrologists. In the area of economics, who predicted the subprime and credit crunch crises, the Internet bubble, the Asian contagion, the real estate and savings and loans crises, the Latin American lending calamity, and the other major disasters? In business, who "predicted" the collapse of Lehman Brothers, Bear Stearns, AIG, Enron or WorldCom (in the USA), and Northern Rock,

* Corresponding editor. Tel.: +30 6977661144.
E-mail addresses: smakrid@otenet.gr (S. Makridakis), nnt@fooledbyrandomness.com (N. Taleb).
[1] Tel.: +1 718 260 3599; fax: +1 718 260 3355.

0169-2070/$ - see front matter © 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2009.05.013

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

Royal Bank of Scotland, Parmalat or Royal Ahold (in Europe); or the practical collapse of the entire Iceland economy? In finance, who predicted the demise of LTCM and Amaranth, or the hundreds of mutual and hedge funds that close down every year after incurring huge losses? And these are just the tip of the iceberg.

In the great majority of situations, predictions are never accurate. As is mentioned by Orrell and McSharry (this issue), the exception is with mechanical systems in physics and engineering. The predictability of practically all complex systems affecting our lives is low, while the uncertainty surrounding our predictions cannot be reliably assessed. Perpetual calendars in handheld devices, including watches, can show the exact rise and set of the sun and the moon, as well as the phases of the moon, up to the year 2099 and beyond. It is impressive that such small devices can provide highly accurate forecasts. For instance, they predict that on April 23, 2013, in Greece:

> The sun will rise at 5:41 and set at 7:07
> The moon will rise at 4:44 and set at 3:55
> The phase of the moon will be more than 3/4 full, or 3 days from full moon.

These forecasts are remarkable, as they concern so many years into the future, and it is practically certain that they will be perfectly accurate so many years from now. The same feeling of awe is felt when a spaceship arrives at its exact destination after many years of traveling through space, when a missile hits its precise target thousands of kilometers away, or when a suspension bridge spanning 2000 m can withstand a strong earthquake, as predicted in its specifications.

Physics and engineering have achieved amazing successes in predicting future outcomes. By identifying exact patterns and precise relationships, they can extrapolate or interpolate them, to achieve perfect, error free forecasts. These patterns, like the orbits of celestial objects, or relationships like those involving gravity, can be expressed with exact mathematical models that can then be used for forecasting the positions of the sun and the moon on April 23, 2013, or firing a missile to hit a desired target thousands of kilometers away. The models used make no significant errors, even though they are simple and can often be programmed into hand-held devices.

Predictions involving celestial bodies and physical law type relationships that result in near-perfect, error free forecasts are the exception rather than the rule—and forecasting errors are of no serious consequence, thanks to the "thin-tailedness" of the deviations. Consider flipping a coin 10 times; how many heads will appear? In this game there is no certainty about the outcome, which can vary anywhere from 0 to 10. However, even with the most elementary knowledge of probability, the best forecast for the number of heads is 5, the most likely outcome, which is also the average of all possible ones. It is possible to work out that the chance of getting exactly five heads is 0.246, or to compute the corresponding probability for any other number.

The distribution of errors, when a coin is flipped 10 times and the forecast is 5 heads, is shown in Fig. 1, together with the actual results of 10,000 simulations. The fit between the theoretical and actual results is remarkable, signifying that uncertainty can be assessed correctly when flipping a coin 10 times.

Games of chance like flipping coins, tossing dice, or spinning roulette wheels have an extremely nice property: the events are independent, while the probability of success or failure is constant over all trials. These two conditions allow us to calculate both the best forecast and the uncertainty associated with various occurrences. Moreover, when n, the number of trials, is large, the central limit theorem applies, guaranteeing that the distribution around the mean, the most likely forecast, can be approximated by a normal curve, knowing that the larger the value of n the better the approximation. Even when a coin is tossed 10 times ($n = 10$), the distribution of errors, with a forecast of 5, can be approximated pretty well with a normal distribution, as can be seen in Fig. 1.

With celestial bodies and physical law relationships, we can achieve near-perfect predictions. With games of chance, we know that there is no certainty, but we can figure out the most appropriate forecasts and estimate precisely the uncertainty involved. In the great majority of real life situations, however, there is always doubt as to which is the "best" forecast, and, even worse, the uncertainty surrounding a forecast cannot be assessed, for three reasons. First, in most cases, errors are not independent of one another; their variance is not constant, while their distribution cannot be assured to follow a normal curve—which means that the variance itself will be either intractable or a poor indicator of potential errors, what has been

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

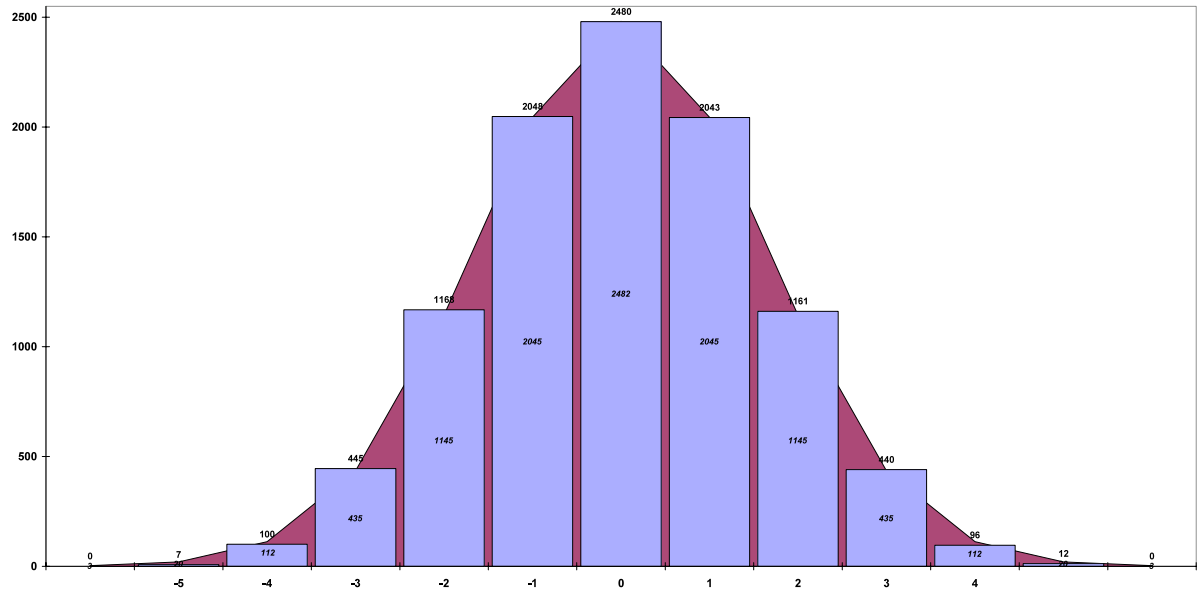


Fig. 1. The erros assuming 5 heads when a coin is flipped 10 times (10,000 replications).

called "wild randomness" by Mandelbrot (1963). Second, there is always the chance of highly unlikely or totally unexpected occurrences materializing — and these can play a large role (Taleb, 2007). Third, there is a severe problem outside of artificial setups, such as games: probability is not observable, and it is quite uncertain which probabilistic model to use.

In addition, we must remember that we do not forecast for the sake of forecasting, but for some specific purpose, so we must realize that some forecast errors can cause harm or missed opportunities, while others can be benign. So to us, any analysis of forecasting needs to take the practical dimension into account: both the consequences of forecast errors and the fragility and reliability of predictions. In the case of low reliability, we need to know what to do, depending on the potential losses and opportunities involved.

2. The accuracy and uncertainty in forecasting

This section examines each of two distinct issues associated with forecasting: the accuracy of predictions and the uncertainty surrounding them. In doing so, it distinguishes three types of predictions: (a) those involving patterns, (b) those utilizing relationships, and (c) those based primarily on human judgment. Each of these three will be covered using information from empirical studies and three concrete examples, where ample data are available.

2.1. The accuracy when forecasting patterns

The M-Competitions have provided abundant information about the accuracy of all major time series forecasting methods aimed at predicting patterns. Table 1 lists the overall average accuracies for all forecasting horizons for the 4004 series used in the M-Competition (Makridakis et al., 1982) and the M3-Competition (Makridakis & Hibon, 2000). The table includes five methods. Naïve 1 is a simple, readily available benchmark. Its forecasts for all horizons up to 18 are the latest available value. Naïve 2 is the same as Naïve 1 except that the forecasts are appropriately seasonalized for each forecasting horizon. Single exponential smoothing is a simple method that averages the most recent values, giving more weight to the latest ones, in order to eliminate randomness. Dampen exponential smoothing is similar to single, except that it first smooths the most recent trend in the data to remove randomness and then extrapolates and dampens, as its names implies, such a smoothed trend. Single smoothing was found to be highly accurate in the M- and M3-Competitions, while dampen was one

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

Table 1
MAPE[a] (average absolute percentage error) of various methods and percentage improvements.

	MAPEs: Forecasting horizons				Improvement (in Avg. MAPE)	% Improvement in Avg. MAPE:			
	1st	6th	18th	Avg. MAPE (1–18 horizons)	over Naïve1	Naïve2 over Naïve1	Single over Naïve2	Dampen over Single	Box–Jenkins over Dampen
Naïve1	11.7%	18.9%	24.6%	17.9%					
Naïve2	10.2%	16.9%	22.1%	16.0%	1.9%	11.6%			
Single exponential smoothing	9.3%	16.1%	21.1%	15.0%	2.9%		6.4%		
Dampen exponential smoothing	8.7%	15.0%	19.2%	13.6%	4.3%			8.1%	
The Box–Jenkins methodology to ARIMA models	9.2%	14.9%	19.8%	14.2%	3.7%				**−2.5%**

[a] All MAPEs and % improvements are symmetric; that is, the divisor is: (Method1 − Method2)/(0.5∗Method1 + 0.5∗Method2).

of the best methods in each of these competitions. Finally, the Box–Jenkins methodology with ARIMA models, a statistically sophisticated method that identifies and fits the most appropriate autoregressive and/or moving average model to the data, was less accurate overall than dampen smoothing.

Table 1 shows the MAPEs of these five methods for forecasting horizons 1, 6 and 18, as well as the overall average of all 18 forecasting horizons. The forecasting errors start at around 10% for one period ahead forecasts, and almost double for 18 periods ahead. These huge errors are typical of what can be expected when predicting series similar to those of the M- and M3-Competitions (the majority consisting of economic, financial and business series). Table 1 also shows the improvements in MAPE of the four methods over Naïve 1, which was used as a benchmark. For instance, Naïve 2 is 1.9% more accurate than Naïve 1, a relative improvement of 11.6%, while dampen smoothing is 4.3% more accurate than Naïve 1, a relative improvement of 27.2%.

The right part of Table 1 provides information about the source of the improvements in MAPE. As the only difference between Naïve 1 and Naïve 2 is that the latter captures the seasonality in the data, this means that the 11.6% improvement (the biggest of all) brought by Naïve 2 is due to predicting the seasonality in the 4004 series. An additional improvement of 6.4% comes from single exponential smoothing, which averages the most recent values in order to eliminate random noise. The final improvement of 8.1%, on top of seasonality and randomness, is due to dampen smoothing, which eliminates the randomness in the most recent trend (we can call this trend the momentum of the series). Finally, the Box–Jenkins method is less accurate than dampen smoothing by 0.6%, or, in relative terms, has a decrease of 2.5% in overall forecasting accuracy.

As dampen smoothing cannot predict turning points, we can assume that the Box–Jenkins does not either, as it is less accurate than dampen. In addition, dampen smoothing is considerably more accurate than Holt's exponential smoothing (not shown in Table 1), which extrapolates the most recent smoothed trend, without dampening. This finding indicates that, on average, trends do not continue uninterrupted, and should not, therefore, be extrapolated. Cyclical turns, for instance, reverse established trends, with the consequence of huge errors if such trends are extrapolated assuming that they will continue uninterrupted.

2.2. *The uncertainty when forecasting patterns*

What is the uncertainty in the MAPEs shown in Table 1? Firstly, uncertainty increases together with the forecasting horizon. Secondly, such an increase is bigger than that postulated theoretically. However, it has been impossible to establish the distribution of forecasting errors in a fashion similar to that shown in Fig. 1 or Table 1, as the number of observations in the series in the M-Competitions is not large enough. For this reason, we will demonstrate the uncertainty in

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

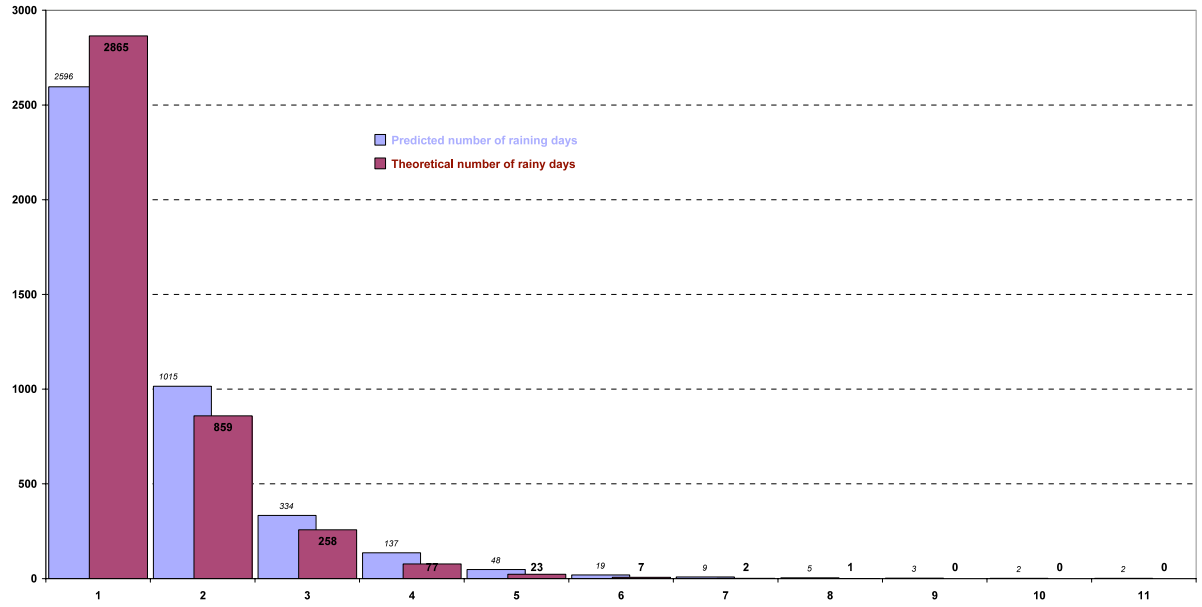


Fig. 2. Predicted and theoretical number of rainy days.

forecasting by using four long series, allowing us to look at the distributions of forecasting errors.

Rainfall data from January 1, 1971 to May 6, 2008 ($n = 13{,}648$) in Amsterdam show that the chance of rain on any given day is very close to that of flipping a coin (0.506, to be precise). Since it rains more during some periods than during others (i.e. events are not independent), we can use Naïve 1 to improve our ability to forecast. By doing so, we increase the probability of correctly predicting rain from 0.506, assuming that rainy days are independent of each other, to 0.694. Fig. 2 shows the theoretical and actual forecasting errors using Naïve 1. The fit between the theoretical and actual errors is remarkable, indicating that we can estimate the uncertainty of the Naïve 1 model with a high degree of reliability when using the theoretical estimates. It seems that in binary forecasting situations, such as rain or no rain, uncertainty can be estimated reliably.

Fig. 3 shows the average daily temperatures in Paris for each day of the year, using data from January 1, 1900 to December 31, 2007. Fig. 3 shows a smooth pattern, with winter days having the lowest temperatures and summer days the highest ones, as expected. Having identified and estimated this seasonal pattern, the best forecast suggested by meteorologists for, say, January 1, 2013, is the average of the temperatures for all 108 years of data, or 3.945 °C.

However, it is clear that the actual temperature on 1/1/2013 will, in all likelihood, be different from this average. An idea of the possible errors or uncertainty around this average prediction can be inferred from Fig. 4, which shows the 108 errors if we use 3.945, the average for January 1, as the forecast. These errors vary from −13 to 8 degrees, with most of them being between 7 and 11 °C. The problem with Fig. 4, however, is that the distribution of errors does not seem to be well behaved. This may be because we do not have enough data (a problem with most real life series) or because the actual distribution of errors is not normal or even symmetric. Thus, we can say that our most likely prediction is 3.945 degrees, but it is difficult to specify the range of uncertainty in this example with any degree of confidence.

The number of forecasting errors increases significantly when we make short term predictions, like the temperature tomorrow, and use Naïve 1 as the forecast (meteorologists can improve the accuracy of predicting the weather over that of Naïve 1 for up to three days ahead). If we use Naïve 1, the average error is zero, meaning that Naïve 1 is an unbiased forecasting model, with a standard deviation of 2.71 degrees and a range of errors from −11.2 to 11 degrees. The

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

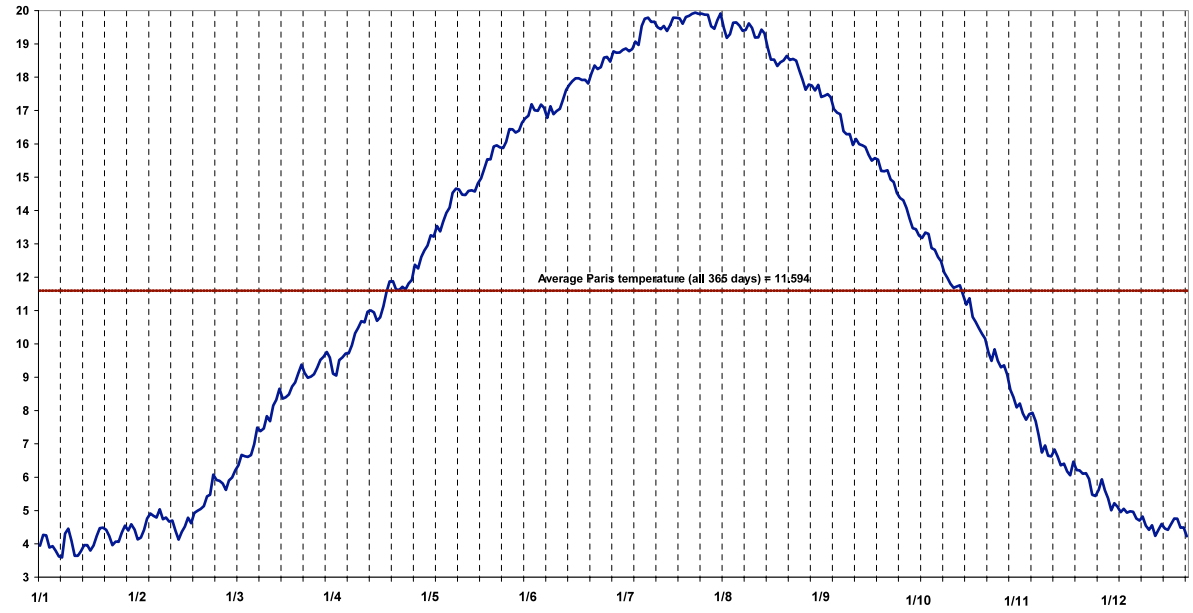


Fig. 3. Average daily temperatures in Paris: 1900 to 2007.

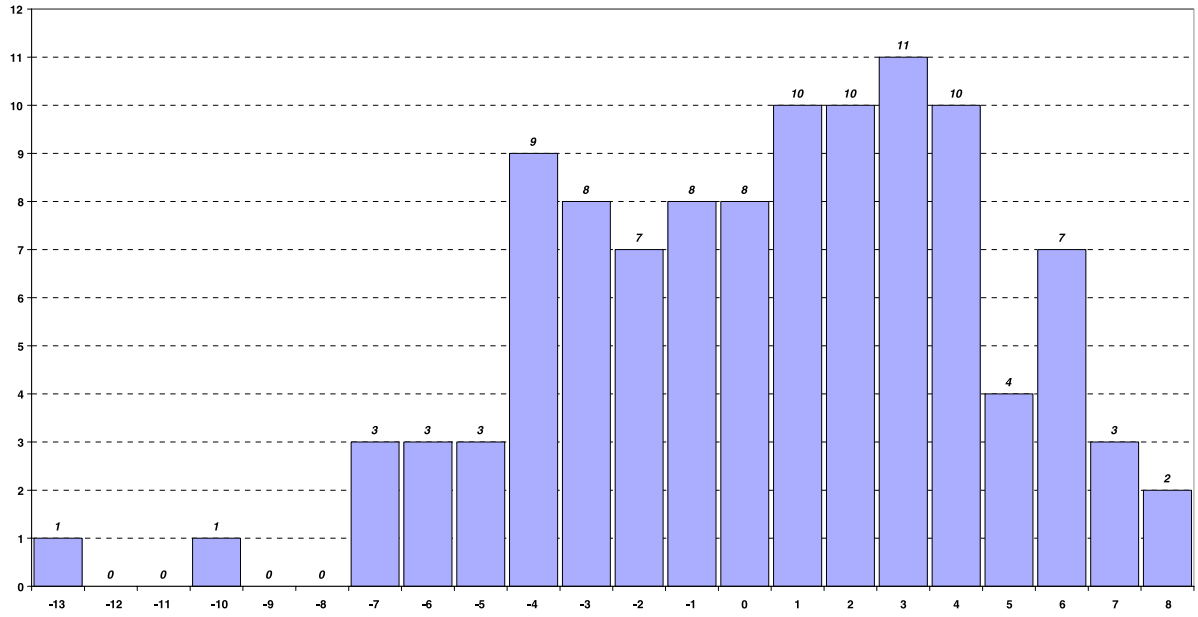


Fig. 4. Errors from the mean in daily temperatures (in Celsius) on January 1st: 1900–2007.

distribution of these errors is shown in Fig. 5, superimposed on a normal curve.

Two observations come from Fig. 5. First, there are more errors in the middle of the distribution than postulated by the normal curve. Second, the tails of the error distribution are much fatter than if they were following a normal curve. For example, there are 14 errors of temperature less than −8.67 degrees, corresponding to more than 4 standard deviations from the mean. This is a practical impossibility if the actual

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

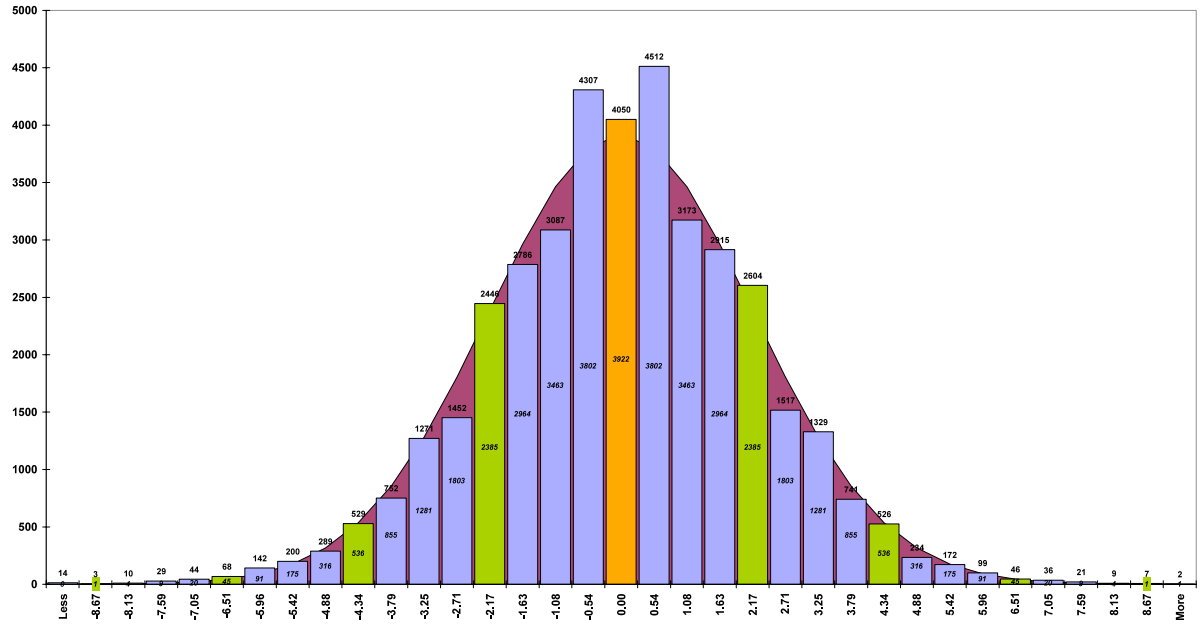


Fig. 5. Paris temperatures 1900–2007: Daily changes.

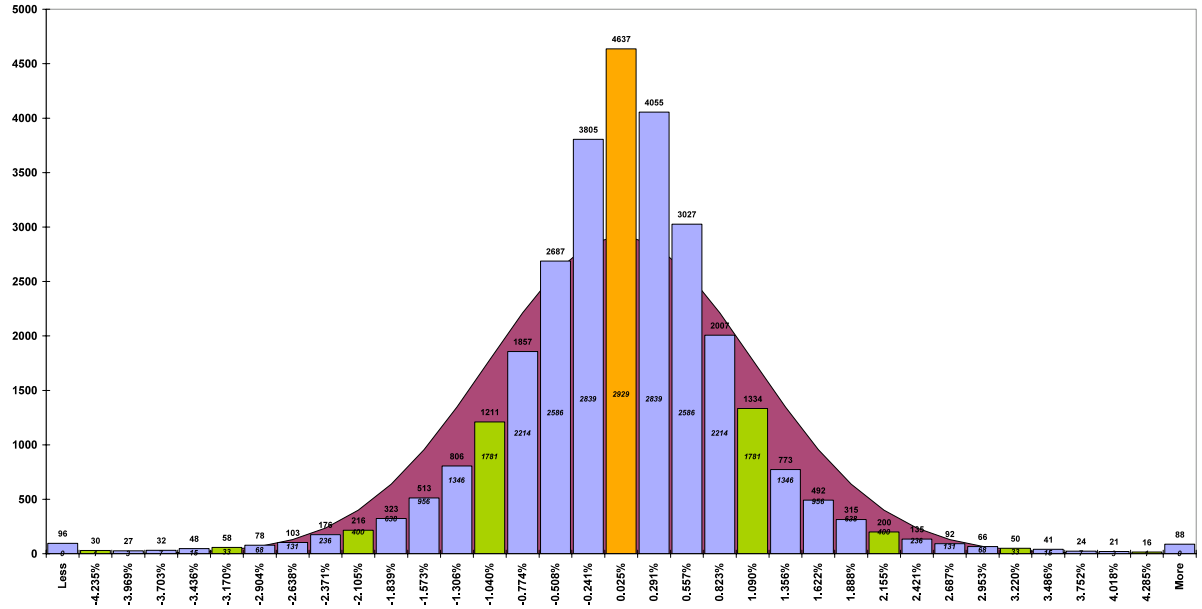


Fig. 6. The daily forecasting errors for the DJIA, 1900–2007.

distribution was a normal one. Similarly, there are 175 errors outside the limits of the mean ±3 standard deviations, versus 69 if the distribution was normal. Thus, can we say that the distribution of errors can be approximated by a normal curve? The answer is complicated, even though the differences are not as large as those of Fig. 6, describing the errors of the next example: the DJIA.

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

Table 2
DJIA 1900–2000: Worst-best daily returns.

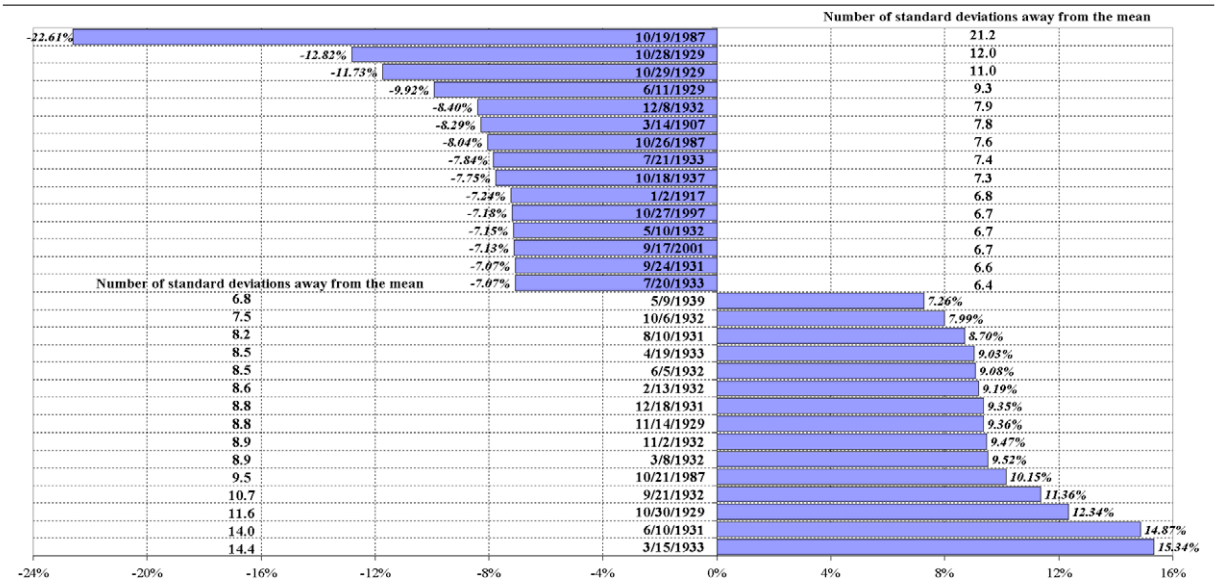


Fig. 6 shows the same information as Fig. 5, except that it refers to the values of the DJIA when Naïve 1 is used as the forecasting model. The data ($n = 29{,}339$) cover the same period as the Paris temperatures, January 1, 1900 to December 31, 2007 (there are fewer observations because the stock market is not open during weekends and holidays). The actual distribution of Fig. 6 also does not follow a normal curve. The middle values are much higher than those of Fig. 5, while there are many more values outside the limits of ±4 standard deviations from the mean. For instance, there are 184 values below and above 4 standard deviations, while there should not be any such values if the distribution was indeed normal.[2]

Table 2 further illustrates the long, fat tails of the errors of Fig. 6 by showing the 15 smallest and largest errors and the number of standard deviations away from the mean such errors correspond to (they range from 6.4 to 21.2 standard deviations). Such large errors could not have occurred in many billions of years if they were part of a normal distribution.

The fact that the distribution of errors in Fig. 6 is much more exaggerated than that of Fig. 5 is due to the human ability to influence the DJIA, which is not the case with temperatures. Such an ability, together with the fact that humans overreact to both good and bad news, increases the likelihood of large movements in the DJIA. There is no other way to explain the huge increases/decreases shown in Table 2, as it is not possible for the capitalization of all companies in the DJIA to lose or gain such huge amounts in a single day by real factors.

Another way to explain the differences between the two figures is that temperature is a physical random variable, subject to physical laws, while financial markets are informational random variables that can take any value without restriction—there are no physical impediments to the doubling of a price. Although physical random variables can be non-normal owing to nonlinearities and cascades, they still need to obey some structure, while informational random variables do not have any tangible constraint.

[2] Departure from normality is not accurately measured by counting the number of observations in excess of 4, 5, or 6 standard deviations (sigmas), but in looking at the contribution of large deviations to the total properties. For instance, the Argentine currency froze for a decade in the 1990s, then had a large jump. Its kurtosis was far more significant than the Paris weather, although we only had one single deviation in excess of 4 sigmas. This is the problem with financial measurements that discard the effect of a single jump.

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

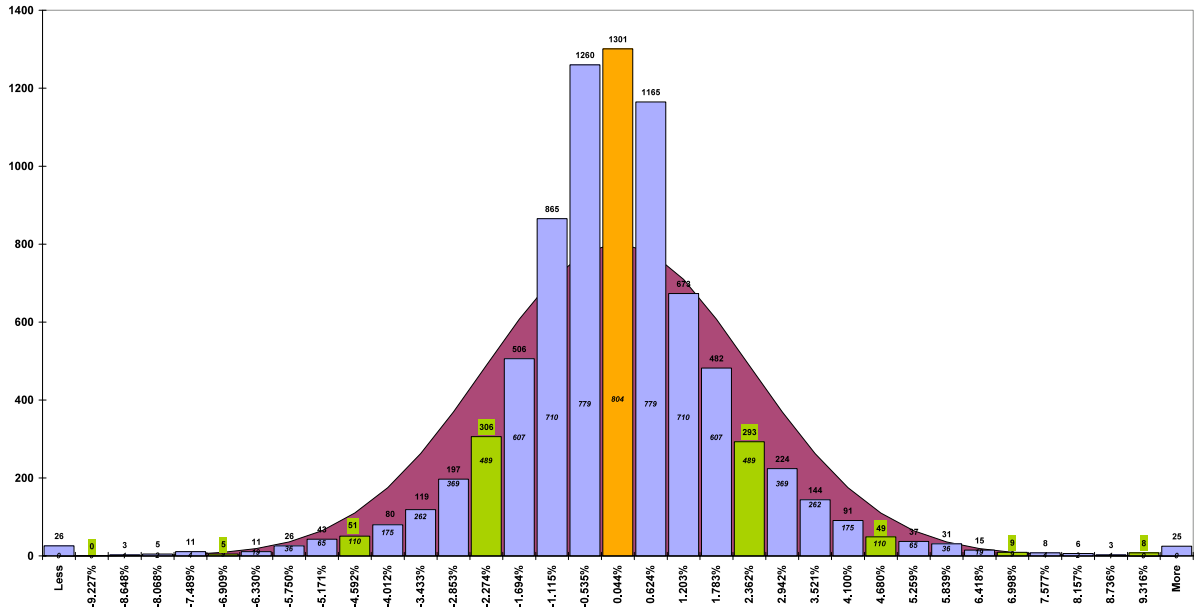


Fig. 7. The daily forecasting errors for Citigroup, 1977–2008.

Non-normality gets worse where individual stocks are concerned, as the recent experience with bank stocks has shown. For instance, the price of Citigroup dropped 34.7% between September 9 and 17, 2008, and then increased by 42.7% on the two days of September 18 and 19. These are huge fluctuations that are impossible to explain assuming independence and well behaved errors (the mean daily return of Citigroup is 0.044% and the standard deviation is 2.318%). Therefore, the uncertainty surrounding future returns of Citigroup cannot be also assessed either, as the distribution has long, fat tails (see Fig. 7), and its errors are both proportionally more concentrated in the middle, and have proportionally more extreme values in comparison to those of the DJIA shown in Fig. 6.

2.3. The accuracy and uncertainty when forecasting relationships

There is no equivalent of the M-Competitions to provide us with information about the post-sample forecasting accuracy of relationships. Instead, econometricians use the R^2 value to determine the goodness of fit of how much better the average relationship is in comparison to the mean (used as a benchmark).

Estimating relationships, like patterns, requires "averaging" of the data to eliminate randomness. Fig. 8 shows the heights of 1078 fathers and sons,[3] as well as the average of such a relationship passing through the middle of the data.

The most likely prediction for the height of a son whose father's height is 180 cm is 178.59 cm, given that the average relationship is:

$$\begin{aligned}\text{Height Son} &= 86.07 + 0.514(\text{Height Father})\\ &= 178.59. \end{aligned} \quad (1)$$

Clearly, it is highly unlikely that the son's height will be exactly 178.59, the average postulated by the relationship, as the pairs of heights of fathers and sons fluctuate a great deal around the average shown in Fig. 8. The errors, or uncertainty, in the predictions depend upon the sizes of the errors and their distribution. These errors, shown in Fig. 9, fluctuate from about −22.5 to +22.8 cm, with the big majority being between −12.4 and +12.4. In addition, the distribution of forecast errors seems more like a normal curve, although there are more negative

[3] These are data introduced by Karl Pearson, a disciple of Sir Francis Galton.

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

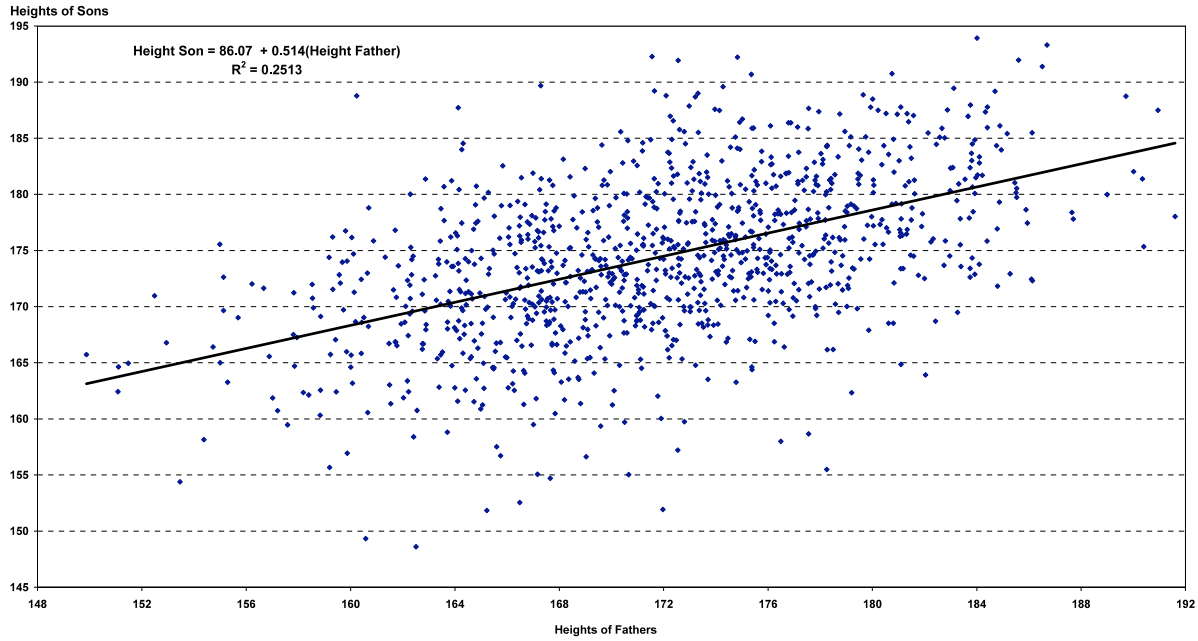


Fig. 8. Heights: Fathers and sons.

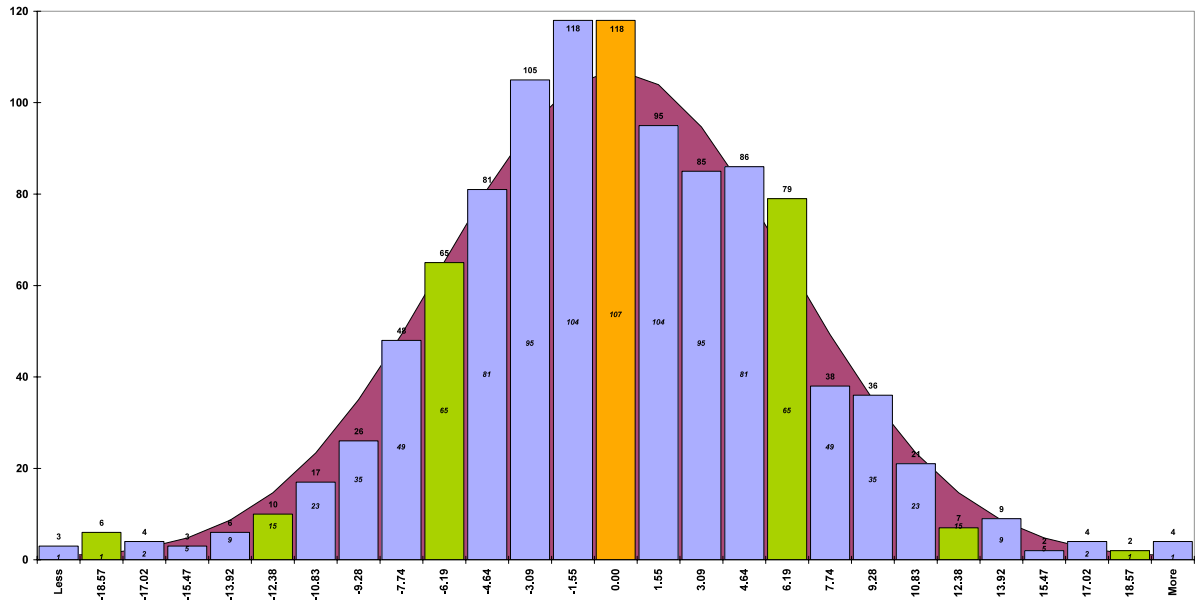


Fig. 9. The residual errors of the relationship height of fathers/sons.

errors close to the mean than postulated by the normal distribution, and more very small and very large ones. Given such differences, if we can assume that the distribution of errors is normal, we can then specify a 95% level of uncertainty as being:

$$\text{Height Son} = 86.07 + 0.514(\text{Height Father}) \pm 1.96(6.19) \quad (2)$$

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

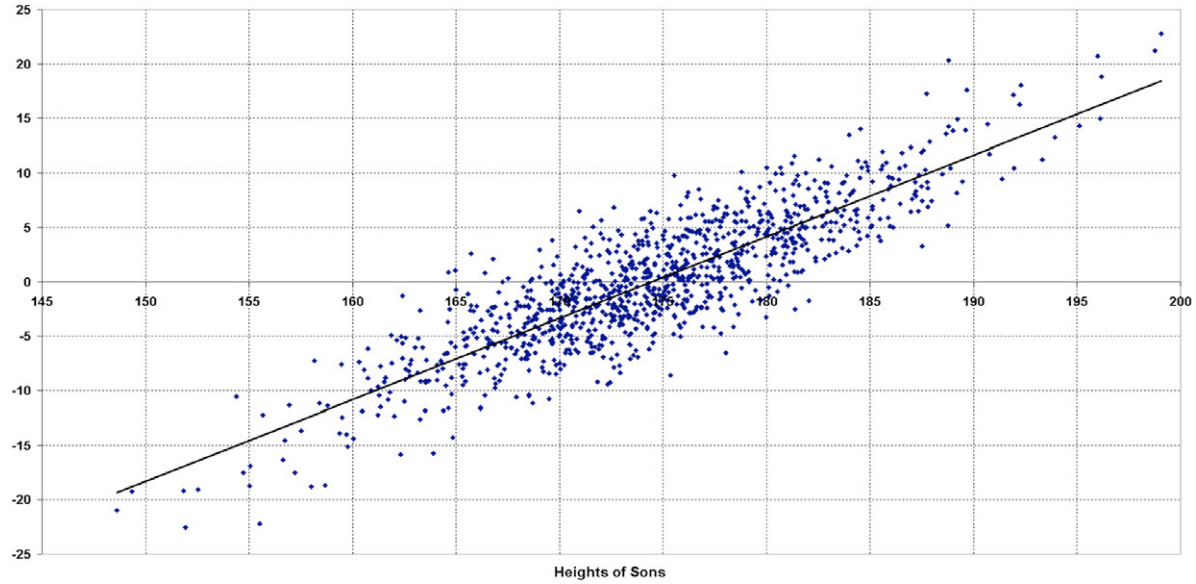


Fig. 10. Residual errors vs heights of sons.

(6.19 is the standard deviation of residuals).

Thus,

Height of Son $= 178.59 \pm 12.3$.

Even in this simple example, ± 12.3 cm indicates a lot of uncertainty in the prediction, which also suffers from the fact that the distribution of errors is not entirely normal. In addition, there is another problem that seriously affects uncertainty. If the errors are plotted against the heights of the sons (Fig. 10), they show a strong correlation, implying that expression (1) underestimates short heights and overestimates tall ones. It is doubtful, therefore, that the forecast specified by expression (1) is the best available for the heights of sons, while the uncertainty shown in expression (2) cannot be estimated correctly, as the errors are highly correlated. Finally, there is an extra problem when forecasting using relationships: the values of the independent variables must, in the great majority of cases, be predicted (this is not the case with (1) as the height of the father is known), adding an extra level of uncertainty to the desired prediction.

Forecasts from econometric models used to be popular, giving rise to an industry with revenues in the hundreds of millions of dollars. Today, econometric models have somewhat fallen out of fashion, as empirical studies have showed that their predictions were less accurate than those of time series methods like Box–Jenkins. Today, they are only used by governmental agencies and international organizations for simulating policy issues and better understanding the consequences of these issues. Their predictive ability is not considered of value (see Orrell & McSharry, this issue), as their limitations have been accepted by even the econometricians themselves, who have concentrated their attention on developing more sophisticated models that can better fit the available data.

Taleb (2007) revisits the idea that such consequences need to be taken into account in decision making. He shows that forecasting has a purpose, and it is the purpose that may need to be modified when we are faced with large forecasting errors and huge levels of uncertainty that cannot be assessed reliably.

2.4. *Judgmental forecasting and uncertainty*

Empirical findings in the field of judgmental psychology have shown that human judgment is even less accurate at making predictions than simple statistical models. These findings go back to the fifties with the work of psychologist Meehl (1954), who reviewed some 20 studies in psychology and discovered that the "statistical" method of diagnosis was superior to the traditional "clinical" approach.

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

When Meehl published a small book about his research findings in 1954, it was greeted with outrage by clinical psychologists all over the world, who felt professionally diminished and dismissed his findings. Many subsequent studies, however, have confirmed Meehl's original findings. A meta-analysis by Grove, Zald, Lebow, Snitz, and Nelson (2000) summarized the results of 136 studies comparing clinical and statistical predictions across a wide range of environments. They concluded by stating:

"*We identified no systematic exceptions to the general superiority (or at least material equivalence) of mechanical prediction. It holds in general medicine, in mental health, in personality, and in education and training settings. It holds for medically trained judges and for psychologists. It holds for inexperienced and seasoned judges*".

A large number of people can be wrong, and know that they can be wrong, brought about by the comfort of a system. They continue their activities "because other people do it". There have been no studies examining the notion of the diffusion of responsibility in such problems of group error.

As Goldstein and Gigerenzer (this issue) and Wright and Goodwin (this issue) point out, the biases and limitations of human judgment affect its ability to make sound decisions when optimism influences its forecasts. In addition, it seems that the forecasts of experts (Tetlock, 2005) are not more accurate than those of other knowledgeable people. Worse, Tetlock found out that experts are less likely to change their minds than non-experts, when new evidence appears disproving their beliefs.

The strongest evidence against the predictive value of human judgment comes from the field of investment, where a large number of empirical comparisons have proven, beyond the slightest doubt, that the returns of professional managers are not better than a random selection of stocks or bonds. As there are around 8500 investment funds in the USA, it is possible that a fund can beat, say, the S&P500, for 13 years in a row. Is this due to the ability of its managers or to chance? If we assume that the probability of beating the S&P 500 each year is 50%, then if there were 8192 funds, it would be possible for one of them to beat the S&P500 for 13 years in a row by pure chance. Thus, it is not obvious that the funds that outperform the market for many years in a row do so by the ability of their managers and rather than because they happen to be lucky. So far there is no empirical evidence that has conclusively proven that professional managers have consistently outperformed the broad market averages due to their own skills (and compensation). In addition to the field of investments, Makridakis, Hogarth, and Gaba (2009) have concluded that in the areas of medicine, as well as business, the predictive ability of doctors and business gurus is not better than simple benchmarks. These findings raise the question of the value of experts: why pay them to provide forecasts that are not better than chance, or than simple benchmarks like the average or the latest available value?

Another question is, how well can human judgment assess future uncertainty? Empirical evidence has shown that the ability of people to correctly assess uncertainty is even worse than that of accurately predicting future outcomes. Such evidence has shown that humans are overconfident of positive expectations, while ignoring or downgrading negative information. This means that when they are asked to specify confidence intervals, they make them too tight, while not considering threatening possibilities like the consequences of recessions, or those of the current subprime and credit crisis. This is a serious problem, as statistical methods also cannot predict recessions and major financial crises, creating a vacuum resulting in surprises and financial hardships for large numbers of people, as nobody has provided them with information to enable them to consider the full range of uncertainty associated with their investments or other decisions and actions.

3. A summary of the eight papers of this issue

This introductory paper by Makridakis and Taleb demonstrates the limited predictability and high level of uncertainty in practically all important areas of our lives, and the implications of this. It presents empirical evidence proving this limited predictability, as well as examples illustrating the major errors involved and the high levels of uncertainty that cannot be adequately assessed because the forecasting errors are not independent, normally distributed and constant. Finally, the paper emphasizes the need to be rational and realistic about our expectations from forecasting,

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

and avoid the common illusion that predictions can be accurate and that uncertainty can be assessed correctly.

The second paper, by Orrell and McSharry, states that complex systems cannot be reduced to simple mathematical laws and be modeled appropriately. The equations that attempt to represent them are only approximations to reality, and are often highly sensitive to external influences and small changes in parameterization. Most of the time they fit past data well, but are not good for predictions. Consequently, the paper offers suggestions for improving forecasting models by following what is done in systems biology, integrating information from disparate sources in order to achieve such improvements.

The third paper, by Taleb, provides evidence of the problems associated with econometric models, and proposes a methodology to deal with such problems by calibrating decisions, based on the nature of the forecast errors. Such a methodology classifies decision payoffs as simple or complex, and randomness as thin or fat tailed. Consequently, he concentrates on what he calls the fourth quadrant (complex payoffs and fat tail randomness), and proposes solutions to mitigate the effects of possibly inaccurate forecasts based on the nature of complex systems.

The fourth paper, by Goldstein and Gigerenzer, provides evidence that some of the fast and frugal heuristics that people use intuitively are able to make forecasts that are as good as or better than those of knowledge-intensive procedures. By using research on the adaptive toolbox and ecological rationality, they demonstrate the power of using intuitive heuristics for forecasting in various domains, including sports, business, and crime.

The fifth paper, by Ioannidis, provides a wealth of empirical evidence that while biomedical research is generating massive amounts of information about potential prognostic factors for health and disease, few prognostic factors have been robustly validated, and fewer still have made a convincing difference in health outcomes or in prolonging life expectancy. For most diseases and outcomes, a considerable component of the prognostic variance remains unknown, and may remain so in the foreseeable future. Ioannidis suggests that in order to improve medical predictions, a systematic approach to the design, conduct, reporting, replication, and clinical translation of prognostic research is needed. Finally, he suggests that we need to recognize that perfect individualized health forecasting is not a realistic target in the foreseeable future, and we have to live with a considerable degree of residual uncertainty.

The sixth paper, by Fink, Lipatov and Konitzer, examines the accuracy and reliability of the diagnoses made by general practitioners. They note that only 10% of the results of consultations in primary care can be assigned to a confirmed diagnosis, while 50% remain "symptoms", and 40% are classified as "named syndromes" ("picture of a disease"). In addition, they provide empirical evidence collected over the last fifty years showing that less than 20% of the most frequent diagnoses account for more than 80% of the results of consultations. Their results prove that primary care has a severe "black swan" element in the vast majority of consultations. Some critical cases involving "avoidable life-threatening dangerous developments" such as myocardial disturbance, brain bleeding and appendicitis may be masked by those often vague symptoms of health disorders ranked in the 20% of most frequent diagnoses. They conclude by proposing that (1) primary care should no longer be defined only by "low prevalence" properties, but also by its black-swan-incidence-problem; (2) at the level of everyday practice, diagnostic protocols are necessary to make diagnoses more reliable; and (3) at the level of epidemiology, a system of classifications is crucial for generating valid information by which predictions of risks can be improved.

The seventh paper, by Makridakis, Hogarth and Gaba, provides further empirical evidence that accurate forecasting in the economic and business world is usually not possible, due to the huge uncertainty, as practically all economic and business activities are subject to events which we are unable to predict. The fact that forecasts can be inaccurate creates a serious dilemma for decision and policy makers. On the one hand, accepting the limits of forecasting accuracy implies being unable to assess the correctness of decisions and the surrounding uncertainty. On the other hand, believing that accurate forecasts are possible means succumbing to the illusion of control and experiencing surprises, often with negative consequences. They suggest that the time has come for a new attitude towards dealing with the future that accepts our limited ability to make predictions in the economic and business environment, while also providing a framework

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

that allows decision and policy makers to face the future — despite the inherent limitations of forecasting and the huge uncertainty surrounding most future-oriented decisions.

The eighth paper, by Wright and Goodwin, looks at scenario planning as an aid to anticipation of the future under conditions of low predictability, and examines its success in mitigating issues to do with inappropriate framing, cognitive and motivational bias, and inappropriate attributions of causality. They consider the advantages and limitations of such planning and identify four potential principles for improvement: (1) challenging mental frames, (2) understanding human motivations, (3) augmenting scenario planning through adopting the approach of crisis management, and (4) assessing the flexibility, diversity, and insurability of strategic options in a structured option-against-scenario evaluation.

The ninth paper, by Green, Armstrong and Soon, proposes a no change, benchmark model for forecasting temperatures which they argue is the most appropriate one, as temperatures exhibit strong (cyclical) fluctuations and there is no obvious trend over the past 800,000 years that Antarctic temperature data from the ice-core record is available. These data also show that the temperature variations during the late 1900s were not unusual. Moreover, a comparison between the *ex ante* projections of the benchmark model and those made by the Intergovernmental Panel on Climate Change at 0.03 °C-per-year were practically indistinguishable from one another in the small sample of errors between 1992 through 2008. The authors argue that the accuracy of forecasts from the benchmark is such that even perfect prediction would be unlikely to help policymakers in getting forecasts that are substantively more accurate than those from a no change, benchmark model.

Because global warming is an emotional issue, the editors believe that whatever actions are taken to reverse environmental degradation cannot be justified on the accuracy of predictions of mathematical or statistical models. Instead, it must be accepted that accurate predictions are not possible and uncertainty cannot be reduced (a fact made obvious by the many and contradictory predictions concerning global warming), and whatever actions are taken to protect the environment must be justified based on other reasons than the accurate forecasting of future temperatures.

The tenth paper, by the late David Freedman, shows that model diagnostics have little power unless alternative hypotheses can be narrowly defined. For instance, independence of observations cannot be tested against general forms of dependence. This means that the basic assumptions in regression models cannot be inferred from the data. The same is true with the proportionality assumption, in proportional-hazards models, which is not testable. Specification error is a primary source of uncertainty in forecasting, and such uncertainty is difficult to resolve without external calibration, while model-based causal inference is even more problematic to test. These problems decrease the value of our models and increase the uncertainty of their predictions.

The final paper of this issue, written by the editors, is a summary of the major issues surrounding forecasting, and also puts forward a number of ideas aimed at a complex world where accurate predictions are not possible and where uncertainty reigns. However, once we accept the inaccuracy of forecasting, the critical question is, how can we plan, formulate strategies, invest our savings, manage our health, and in general make future-oriented decisions, accepting that there are no crystal balls? This is where the editors believe that much more effort and thinking is needed, and where they are advancing a number of proposals to avoid the negative consequences involved while also profiting from the low levels of predictability.

4. The problems facing forecasters

The forecasts of statistical models are "mechanical", unable to predict changes and turning points, and unable to make predictions for brand new situations, or when there are limited amounts of data. These tasks require intelligence, knowledge and an ability to learn which are possessed only by humans. Yet, as we saw, judgmental forecasts are less accurate than the brainless, mechanistic ones provided by statistical models. Forecasters find themselves between Carybdis and Scylla. On the one hand, they understand the limitations of the statistical models. On the other hand, their own judgment cannot be trusted. The biggest advantage of statistical predictions is their objectivity,

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

Table 3
Values of daily statistics for DJIA and Paris temperatures for each decade from 1900 to 2008.

	Daily DJIA values							Daily Paris temperatures						
Decade	Mean	St. Dev	Min	Max	Skewness	Kurtosis	n	Mean	St. Dev	Min	Max	Skewness	Kurtosis	n
1900 - 1910	0.019%	1.03%	-8.29%	6.69%	-0.36	4.75	2992	-0.001	2.238	-10.4	10.6	0.04	0.81	3650
1910 - 1920	0.018%	0.98%	-7.24%	5.47%	-0.50	4.74	2876	0.001	2.174	-9.8	8.2	-0.05	0.58	3650
1920 - 1930	0.034%	1.11%	-12.82%	12.34%	-0.94	21.41	2986	0.000	2.158	-10.1	8.6	0.04	0.62	3650
1930 - 1940	0.000%	1.85%	-8.40%	15.34%	0.63	6.35	2988	-0.002	2.148	-7.8	8.8	0.04	0.24	3650
1940 - 1950	0.013%	0.74%	-6.80%	4.73%	-1.16	10.84	2918	0.001	2.265	-10.1	11.0	0.00	0.64	3650
1950 - 1960	0.049%	0.66%	-6.54%	4.13%	-0.84	6.76	2598	0.002	2.204	-9.3	8.0	-0.02	0.59	3650
1960 - 1970	0.009%	0.65%	-5.71%	4.69%	0.02	5.45	2489	-0.003	2.162	-9.0	7.4	-0.16	0.40	3650
1970 - 1980	0.006%	0.93%	-3.50%	5.08%	0.33	1.89	2526	0.002	2.090	-9.4	7.7	-0.26	0.49	3650
1980 - 1990	0.054%	1.13%	-22.61%	10.15%	-3.08	68.81	2528	-0.001	2.158	-11.2	6.5	-0.27	0.46	3650
1990 - 2000	0.061%	0.89%	-7.18%	4.98%	-0.31	4.88	2528	0.002	2.140	-10.3	6.8	-0.30	0.20	3650
2000 - 2008	-0.004%	1.30%	-7.87%	11.08%	0.26	9.12	2264	0.005	2.088	-7.7	7.4	-0.18	0.32	3163
1900 - 2008	0.023%	1.08%	-22.61%	15.34%	0.18	18.89	29693	0.000	2.167	-11.2	11.0	-0.09	0.51	39663

which seems to be more important than the intelligence, knowledge and ability of humans to learn. The problem with humans is that they suffer from inconsistency, wishful thinking and all sorts of biases that diminish the accuracy of their predictions. The biggest challenge and only solution to the problem is for humans to find ways to exploit their intelligence, knowledge and ability to learn while avoiding their inconsistencies, wishful thinking and biases. We believe that much work can be done in this direction.

Below, we summarize the problem of limited predictability and high levels of uncertainty using the daily values of the DJIA and the Paris temperatures. The availability of fast computers and practically unlimited memory has allowed us to work with long series and study how well they can forecast and identify uncertainty. Table 3 shows various statistics for the daily % changes in the DJIA and the daily changes in Paris temperatures, for each decade from 1900 to 2008 (the 2000 to 2008 period does not cover the whole decade). Table 3 allows us to determine how well we can forecast and assess uncertainty for the decade 1910–1920, given the information for the decade 1900–1910, for the decade 1920–1930 given the information for 1910–1920, and so on.

4.1. The mean percentage change of the DJIA and the average change in Paris temperature

The mean percentage change in the DJIA for the decade 1900–1910 is 0.019%. If such a change had been used as the forecast for the decade 1910–1920, the results would have been highly accurate. In addition, the volatility in the daily percentage changes from 1900–1910 would have been an excellent predictor for 1910–1920. The same is true with both the means and the standard deviations of the changes in daily temperatures, as they are very similar in the decades 1900–1910 and 1910–1920. Starting from the decade 1920–1930 onwards, however, both the means and the standard deviations of the percentage daily changes in the DJIA vary a great deal, from 0.001% in the 1930s to 0.059% in the 1990s (this means that $10,000 invested at the beginning of 1930 would have become $10,334 by the end of 1939, while the same amount invested at the beginning of 1990 would have grown to $44,307 by the end of 1999). The differences are equally large for the standard deviations, which range from 0.65% in the 1960s to 1.85% in the 1930s. On the other hand, the mean daily changes in temperatures are small, except possibly for the 2000–2008 period, when they increased to 0.005 of a degree. In addition, the standard deviations have remained pretty much constant throughout all eleven decades.

Table 3 conveys a clear message. Forecasting for some series, like the DJIA, cannot be accurate, as the assumption of constancy of their patterns, and possibly relationships, is violated. This means that predicting for the next decade, or any other forecasting horizon, cannot be based on historical information, as both the mean and the fluctuations around the mean vary too much from one decade to another. Does the increase to 0.005 in the changes in daily Paris temperature for the period of 2000–2008 indicate global warming? This is a question we will not attempt to answer, as it has been dealt with in the paper by Green et al. in this issue. However, the potential exists that even in series like temperature we have to worry about a possible change in the long term trend.

Another technique for looking at differences is departures from normality. Consider the kurtosis of the two variables. The 5 largest observations in the temperature represent 3.6% of the total kurtosis. For

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

the Dow Jones, the 5 largest observations represent 38% of the kurtosis (e.g., the kurtosis in the decade 1970–1980 is 1.89, while that of the following decade is an incredible 68.84—see Table 3). Furthermore, under aggregation (i.e., by taking longer observation intervals of 1 week, 1 fortnight, or 1 month), the kurtosis of the temperature drops, while that of the stock market does not change.

In real life, most series behave like the DJIA; in other words, humans can influence their patterns and affect the relationships involved by their actions and reactions. In such cases, forecasting is extremely difficult or even impossible, as it involves predicting human behavior, something which is practically impossible. However, even with series like the temperature human intervention is also possible, although there is no consensus in predicting its consequences.

4.2. The uncertainty in predicting changes in DJIA and Paris temperatures

Having data since 1900 provides us with a unique opportunity to break it into sub-periods and obtain useful insights by examining their consistency (see Table 3), as we have already done for the mean, and we can now assess the uncertainty in these two series. The traditional approach to assessing uncertainty assumes normality and then constructs confidence intervals around the mean. Such an approach cannot work for the percentage changes in the DJIA for three reasons. First, the standard deviations are not constant; second, the means also change substantially from one decade to another (see Table 3); and finally, the distribution is not normal (see Fig. 6). Assessing the uncertainty in the changes in Paris temperatures does not suffer from the first or second problem, as the means and standard deviations are fairly constant. However, the distribution of changes is not quite normal (see Fig. 5), as there are a considerable number of extremely large and small changes, while there are more values around the mean than in a normal curve.

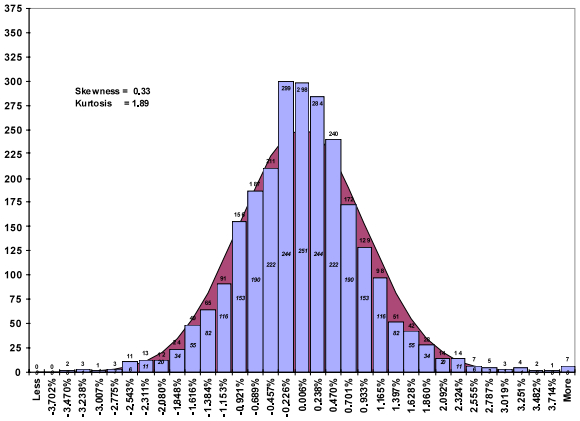
There is an additional problem when attempting to assess uncertainty. The distribution of changes also varies a great deal, as can be seen in Fig. 11. Worse, this is true not only in the DJIA data, but also in the temperature data. In the 1970s, for instance, the distribution of the DJIA percentage changes was close to normal with not too fat tails (the skewness and kurtosis of the distribution were 0.33 and 1.89 respectively), while that of the 1980s was too tall in the middle (the kurtosis was 68.84, versus 1.89 in the 1970s) with considerable fat tails on both ends. Given the substantial differences in the distributions of changes, or errors, is it possible to talk about assessing uncertainty in statistical models when (a) the distributions are not normal, even with series like temperatures; (b) the means and standard deviations change substantially; and (c) the distributions or errors are not constant? We believe that the answer is a strong no, which raises serious concerns about the realism of financial models that assume that uncertainty can be assessed assuming that errors are well behaved, with a zero mean, a constant variance, a stable distribution and independent errors.

The big advantage of series like the DJIA and the Paris temperatures is the extremely large number of available data points that allows us to extract different types of information, such as that shown in Table 3, which is based on more than 2500 observations in the case of the DJIA, and 3650 for the temperatures. Real life series, however, seldom exceed a few hundred observations at most, making it impossible to construct distributions similar to those of Table 3. In such a case we are completely unable to verify the assumptions required to assure ourselves that there are not problems with the assessment of uncertainty. Finally, there is another even more important assumption, that of independence, that also fails to hold true, and negatively affects both the task of forecasting and that of assessing uncertainty. For instance, it is interesting to note that between September 15 and December 1, 2008, 52.7% of the daily fluctuations in the DJIA were greater than the mean ± 3 (standard deviations). In the temperature changes there are fewer big concentrations of extreme values, but since 1977 we can observe that the great majority of such values are negative, again obliging us to question the independence of series like temperatures, which seem to be also influenced by non-random runs of higher and lower temperatures.

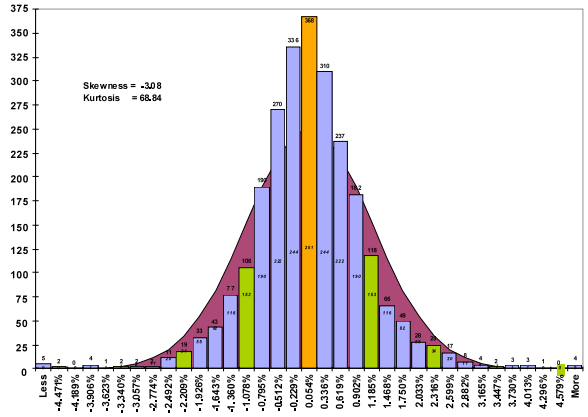
5. Conclusions

Forecasting the future is neither easy nor certain. At the same time, it may seem that we have no choice. But

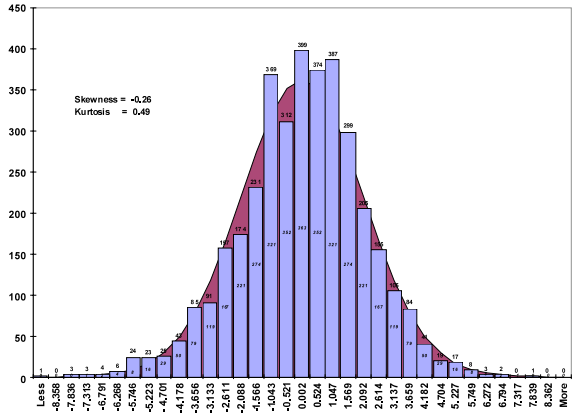
Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013



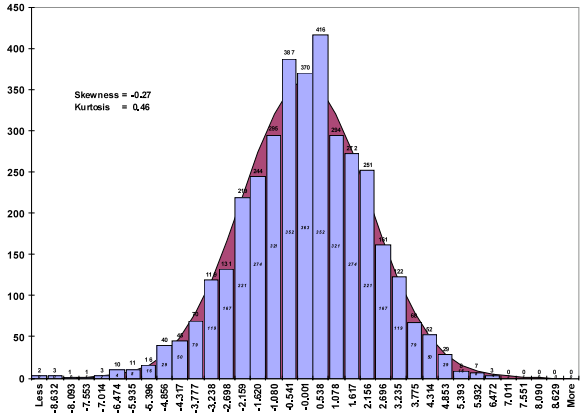
(a) The distribution of daily percentage changes in the DJIA in the 1970s.



(b) The distribution of daily percentage changes in the DJIA in the 1980s.



(c) The distribution of daily changes in the Paris temperatures in the 1970s.



(d) The distribution of daily changes in the Paris temperatures in the 1980s.

Fig. 11. The distribution of daily changes in the DJIA and Paris temperatures.

in reality we do have a choice: we can make decisions based on the potential sizes and consequences of forecasting errors, and we can also structure our lives to be robust to such errors. In a way, which is the motivation of this issue, we can make deep changes in the decision process affected by future predictions.

This paper has outlined the major theme of this special issue of the *IJF*. Our ability to predict the future is limited, with the obvious consequence of high levels of uncertainty. It has proved such limited predictability using empirical evidence and four concrete data sets. Moreover, it has documented our inability to assess uncertainty correctly and reliably in real-life situations, and has discussed the major problems involved. Unfortunately, patterns and relationships are not constant, while in the great majority of cases: (a) errors are not well behaved, (b) their valiance is not constant, (c) the distribution of errors are not stable, and, worst of all, (d) the errors are not independent of each other.

References

Goldstein, D., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, this issue (doi:10.1016/j.ijforecast.2009.05.010).

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30.

Makridakis, S., Hogarth, R., & Gaba, A. (2009). *Dance with chance: Making luck work for you*. Oxford: Oneworld.

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolative (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: The University of Minnesota Press.

Orrell, D., & McSharry, P. (2009). System economics: Overcoming the pitfalls of forecasting models via a multidisciplinary approach. *International Journal of Forecasting*, this issue (doi:10.1016/j.ijforecast.2009.05.002).

Taleb, N. (2007). *The black swan: The impact of the highly improbable*. Random House (US) and Penguin (UK).

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.

Wright, G., & Goodwin, P. (2009). Decision making and planning under low levels of predictability: Enhancing the scenario method. *International Journal of Forecasting*, this issue (doi:10.1016/j.ijforecast.2009.05.019).

Please cite this article in press as: Makridakis, S., & Taleb, N. Decision making and planning under low levels of predictability. International Journal of Forecasting (2009), doi:10.1016/j.ijforecast.2009.05.013

We Don't Quite Know What We Are Talking About

When we talk about volatility.

Daniel G. Goldstein and Nassim Nicholas Taleb

There is no particular normative reason to express or measure volatility in one of several possible ways, provided we remain consistent. Once we express it in a particular way, however, substituting one measure for another will lead to a consequential mistake.

Suppose we measure volatility in root mean square deviations from the mean, as in conventional statistics. It would then be an error to substitute a definition and consider it mean deviation in the course of decision-making, opinion formation, or descriptions of the property of the process. Yet people do make this mistake.

This brief note provides experimental evidence that participants with varied backgrounds in financial markets err in interpreting a physical linear description (in mean absolute returns per day) as a calculated non-linear measure (standard deviation). We illustrate the confusion, and discuss its implications for financial decision-making and portfolio risk management.

EXPERIMENTS

To investigate the common understanding of mean absolute deviation, we asked professionals and students of finance the question:

> A stock (or a fund) has an average return of 0%. It moves on average 1% a day in absolute value; the average up move is 1%, and the average down move is 1%. This does not mean that all up moves are 1%—some are 0.6%, others 1.45%, and so on. Assume that we live in the Gaussian world where the returns (or daily percentage moves) can be safely

DANIEL G. GOLDSTEIN
is an assistant professor of Marketing at London Business School, London, UK.
dgoldstein@london.edu

NASSIM NICHOLAS TALEB
is an visiting professor at London Business School, London, UK.

modeled using a normal distribution. Assume that a year has 256 business days. Our questions concern the standard deviation of returns (i.e., of the percentage moves), the sigma that is used for volatility in financial applications. What is the daily sigma? What is the yearly sigma?

Our suspicion that there would be considerable confusion is fed by years of hearing options traders make statements like "an instrument that has a daily standard deviation of 1% should move 1% a day on average." Not so. In the Gaussian world, where x is a random variable, assuming a mean of 0, in expectation, the ratio of standard deviation to mean deviation should satisfy the equality:

$$\frac{\sum |x|}{\sqrt{\sum x^2}} = \sqrt{\frac{2}{\pi}}$$

Since mean absolute deviation is about 0.8 times the standard deviation, in our problem the daily sigma should be 1.25%, and the yearly sigma should be 20.00% (which is the daily sigma annualized by multiplying by 16, the square root of the number of business days).

To test the hypothesis that mean absolute deviation is confused with standard deviation, we ran studies with three groups: 1) 97 portfolio managers, assistant portfolio managers, and analysts at investment management companies who were taking part in a professional seminar; 2) 13 Ivy League graduate students preparing for a career in financial engineering; and 3) 16 investment professionals working for a major bank. The question was presented in writing and explained orally to make sure definitions were clear.

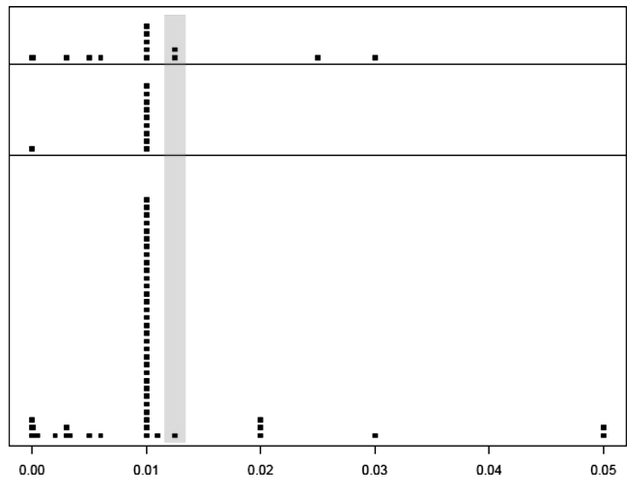
All respondents in the latter two groups turned in responses, compared to 58 in the first group. One might expect this sort of self-selection to improve accuracy.

RESULTS

Exhibit 1 shows frequency histograms of responses to the daily sigma question, converted to decimal notation. Only 3 of the 87 respondents arrived at the correct answer of 0.0125. The modal answer of 0.01 made up more than half of the responses received. Nine people gave responses of blanks or question marks. Far from symmetrical, the ratio of underestimations of volatility to overestimations was 65 to 10.

Performance for yearly sigmas was even worse. No one submitted a correct response. Here, the modal answer was 0.16, which appears to be the correct annualization of the incorrect daily volatility. Eleven people responded with blanks or question marks. The ratio of underestimations to overestimations was 76 to 9.

EXHIBIT 1
Estimates of Daily Standard Deviation



The top plot represents the responses of investment professionals at a major bank. The middle plot represents responses of graduate students in financial engineering (excluding one outlier at 0.1). The bottom plot represents responses of professional portfolio managers and analysts. The correct answer under the stated Gaussian assumptions (0.0125) is shaded in gray.

CONCLUSION

The error we point out is more consequential than it seems. The dominant response of 1% shown in the Exhibit suggests that even financially and mathematically savvy decision-makers treat mean absolute deviation and standard deviation as the same thing. Although a Gaussian random variable that has a daily percentage move in absolute terms of 1.00% has a standard deviation of about 1.25%, it can reach up to 1.90% in empirical distributions (emerging market currencies and bonds). Mean absolute deviation is by Jensen's inequality lower than (or equal to) standard deviation.*

In a world of fat tails, the bias increases dramatically. Consider the vector of dimension 10^6, composed of 999,999 elements of 0 and a single one of 10^6: $V = \{0,0,0,0,\ldots,10^6\}$. Here the standard deviation would be 1,000 times the average move.

For a Student-t with 3 degrees of freedom (often used to model returns in financial markets in value at

risk simulations), standard deviation is 1.57 times mean deviation:

$$\frac{\sum |x|}{\sqrt{\sum x^2}} = \frac{2}{\pi}$$

See Bouchaud and Potters [2003] and Glasserman, Heidelberger, and Shahabuddin [2002].

Conversations with our respondents revealed that they rarely had an immediate understanding of the error when we pointed it out. Yet when we asked them to present the equation for "standard deviation," they expressed it flawlessly as the root mean square of deviations from the mean. Whatever the respondents' reason for the error, it did not result from ignorance of the concept. Indeed, most participants would have failed a basic statistics course had they not been aware of the mathematical definition, but when given data that are clearly not a standard deviation, they treat it as one.

Kahneman and Frederick [2002] discuss a similar problem: that statisticians make basic statistical mistakes outside the classroom:

> The mathematical psychologists who participated in the survey not only should have known better—they did know better . . . . Most of them would have computed the correct answers on the back of an envelope.

Why is this problem relevant? This sloppiness in translation between mathematics and applications can have severe effects, considering that practitioners speak with co-workers, customers, and the media about volatility on a regular basis. We know of instances in the financial media where journalists make the same mistake in explaining the volatility index VIX to the general public.

Either we have the wrong intuition about the right volatility, or the right intuition but the measure of volatility is the wrong one. Two roads lead out of this unfortunate situation. Either we can continue defining volatility as we do, and conduct further research to see if the error can be made to disappear with training. Or we can do as the probabilists of the Enlightenment age did. When the intuitions of *hommes éclairés* did not align with the valuations of the expected value formula, it was the mathematicians who dreamed up something more intuitive, namely, expected utility (see Daston [1988]).

Perhaps some day finance will adopt a more natural metric than standard deviation. Until then, users should rely on definitions, not intuitions, where volatility is concerned.

ENDNOTE

*More technically, the norm $L^p = \left(\frac{1}{n}\Sigma |x|^p\right)^{1/p}$ increases with p.

REFERENCES

Bouchaud, J.-P., and M. Potters. *Theory of Financial Risks and Derivatives Pricing: From Statistical Physics to Risk Management*, 2nd ed. Cambridge: Cambridge University Press, 2003.

Daston, L. J. *Classical Probability in the Enlightenment*. Princeton: Princeton University Press, 1988.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. "Portfolio Value at Risk with Heavy Tailed Risk Factors." *Mathematical Finance*, 12 (2002), pp. 232-269.

Kahneman, D., and S. Frederick. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In T. Gilovich, D. Griffin, and D. Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*. New York: Cambridge University Press, 2002.

Statistical Intuitions and Domains:
The Telescope Test

Daniel G. Goldstein & Nassim N. Taleb

DRAFT; 2 Sept 2010

ABSTRACT

There are two different probabilistic structures: some with “typical” large deviations, others without a “typicality” of these deviations. Many human judgment errors can come from the application of one intuition for one domain to take decisions in the second domain, leading to, among others, an increase in risks. We test whether humans have a natural intuition for the difference.

Statistical Intuitions and Domains:
The Telescope Test

Introduction

There are two types of randomness, "thin tailed and "fat tailed", the first being nonscalable, the second scalable". Missing on the difference (by mistaking the fat tailed domain into a thin-tailed one) can result in a severe flaw in decision-making as the agent might discount the role of the large deviation in determining the statistical properties.

One environment (thin-tailed) allows for intrapolation; the other (fat-tailed) requires extrapolation. Mistaking one domain for another is best illustrated by a passage from the Latin philosopher and poet Lucretius to the effect that people project the largest possible mountain as equal to the largest mountain they've seen in their past, clearly the point there is that they had to have known that previous to encountering such large mountain, the largest they had seen was considerably smaller, yet they remained oblivious to second order thinking. Indeed it is not uncommon for professionals (not just the general public) to miss on these second order effect —the crash of 1987, in which the market went down close to 23 percent in a single day, could not have been guessed in an interpolative way or methods matching "similarities" from its worst predecessor, a one-day loss of around 10 percent; yet operators today work on protecting themselves, thanks to "stress tests" for moves around such level, not extrapolating into larger losses. Similarly, one sees books written by economists on the structure of past debt crises (such as Reinhard and Rogoff, 2009) that can only be informative if one does not consider the Lucretius effect, that the largest mountain to be seen will be equal, or similar, to the next one. Worse of all much of economic risk management by governments worldwide at the time of writing is based on "stress testing" equally plagued with such lack of rigor —the crisis today having not been detected by previous stress testing, should not lead to interpolative methods.

Distinction Between Domains: More formally, our distinction between two domains, nonscalable and scalable along the criterion of a class of distribution that exhibits convergence to the Central Limit Theorem in applicable time (i.e., with acceptable preasymptotics).

For nonscalable domains, the conditional expectation of a random variable X, conditional on its exceeding a number K (henceforth “the boundary”), converge to K for larger values of K.

$$\lim_{K \to \infty} E\left[X \mid_{X>K}\right] = K$$

For instance, the expectation for a Gaussian variable of mean 0, conditional that it exceeds 0, is approximately .8 standard deviations. However, when the boundary K equals 6 standard deviations, the conditional expectation converges to 6 standard deviations: the difference between the conditional expectation and the boundary goes to zero. The same applies to all the random variables that do not have power-law tails. This induces some “typicality” of large moves.

For scalable random variables, such limit does not seem to hold:

$$\lim_{K\to\infty} E\left[X\,|_{X>K}\right] = K\,c$$

where c >1, not necessarily of known parametrization.

Method

Each participant completed one scalable domain item and one non-scalable domain item. Scalable domain items A and C were completed on paper by 68 members of the London Business School participant pool. Non-scalable domain items B and D were completed online by 157 members of the School's online research panel. Item text is as follows:

Scalable domain items

Item A (height): Participants were first asked whether they preferred thinking about height in feet and inches or meters and centimeters, and given an appropriate questionnaire for this preference. The item for the metric formulation is as follows.

- *We computed the average height of men in the USA who are taller than 1.95 meters. Estimate the average height of men in the USA who are taller than 1.95 meters: _____meters*
- *We computed the average height of men in the USA who are taller than 2.05 meters. Estimate the average height of men in the USA who are taller than 2.05 meters: _____meters*
- *We computed the average height of men in the USA who are taller than 2.15 meters. Estimate the average height of men in the USA who are taller than 2.15 meters: _____meters*
- *We computed the average height of men in the USA who are taller than 2.25 meters. Estimate the average height of men in the USA who are taller than 2.25 meters: _____meters*
- *We computed the average height of men in the USA who are taller than 2.35 meters. Estimate the average height of men in the USA who are taller than 2.35 meters: _____meters*

The corresponding values in feet and inches were: 6'4", 6'8", 7'0", 7'4" and 7'8".

Item B (life expectancy): The basic item took the form:

- *We computed the average age at death of women in the USA who lived more than 85 years. Estimate the average age at death of women in the USA who lived more than 85 years: _____ years.*

As with Item A, the basic question was asked four more times by substituting in the following ages: 90, 95, 100, and 105 years.

Non-scalable domain items

Item C (market capitalization): Participants were read "*Market capitalization is defined as "an estimation of the value of a business that is obtained by multiplying the number of shares of stock outstanding by the current price of a share."* The basic item took the form:

- *We computed the average market capitalization of companies in the USA with a market capitalization of greater than 5 billion dollars. Estimate the average market capitalization of companies in the USA with a market capitalization of greater than 5 billion dollars: _____ billion dollars*

The second and third sentences were then repeated, substituting in the following values: 10, 15, 20, and 25 billion dollars.

Item D (stock s): The basic item took the form:

- *We computed the average percentage increase in the price of a typical individual U.S. stock on days when its price increased MORE THAN 10%. Estimate the average percentage increase of a typical individual U.S. stock on days when it increased MORE THAN 10%: _____ %*

The same question was then repeated, substituting in the following percentages: 20%, 30%, 40%, and 50%.

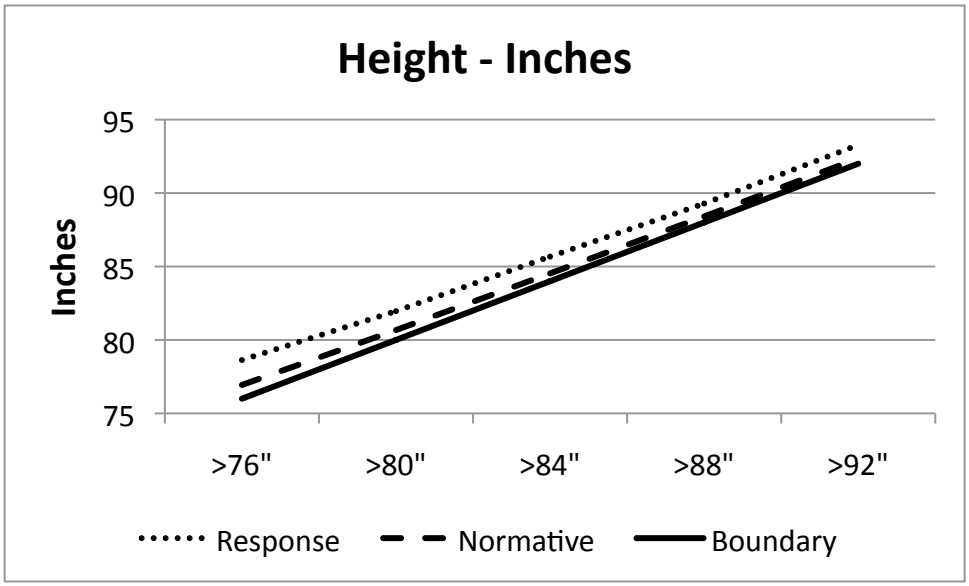
Before data analysis took place, 14 of the 450 responses (3.1%) were excluded for being incomplete, ambiguous or less than the condition given.

Results

Scalable domain items

Item A (height)

Figure AAA plots in the top panel data of the 49 participants who preferred to respond in feet and inches, and those of the 13 participants who preferred to respond in meters in the bottom panel. The normative lines were generated from a normal distribution of mean 69.2 inches (175.77 cm) and standard deviation 2.85 inches (7.24 cm), which has an excellent fit to the height of American male adults (Brainard & Burmaster, 1992). In both cases, the data show that average responses get closer to the boundary in this scalable domain. In the inch data, overestimation of the normative response ranges from 1.70 down to .92 inches, and in the metric data from 5.5 down to 1.8 cm.



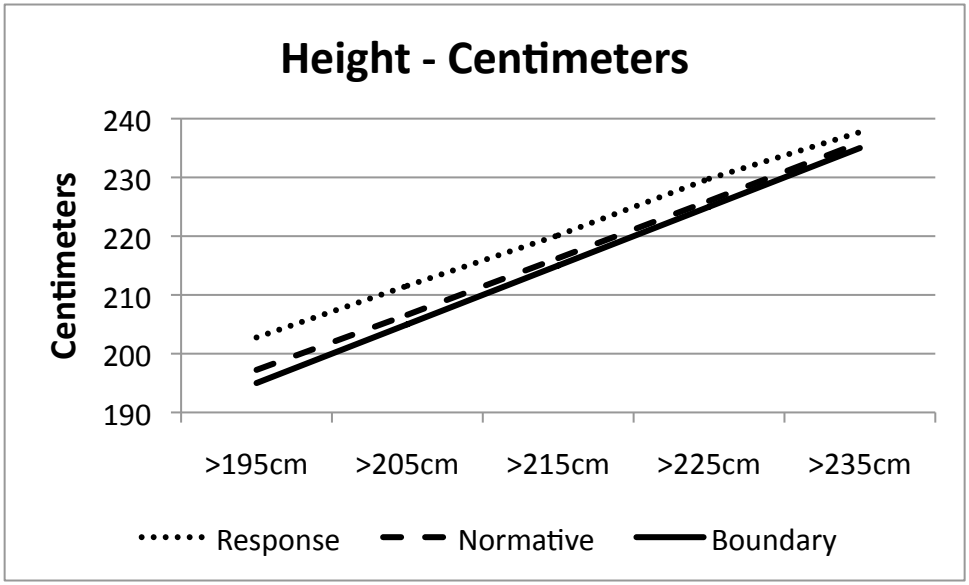


Figure AAA: Estimates of conditional expectation of height of US men in inches (top panel) and centimeters (bottom panel).

Item B (life expectancy)

Figure BBB shows average responses for the life expectancy item from 155 participants. Normative data were taken from governmental actuarial tables (Bell & Miller, 2005, Table 6). These data show slight underestimation ranging from 2.29 down to .03 years. The sample standard error of the mean for the responses is .19, .18. .20, .13, and .12 years for the five categories as they are plotted from left to right.

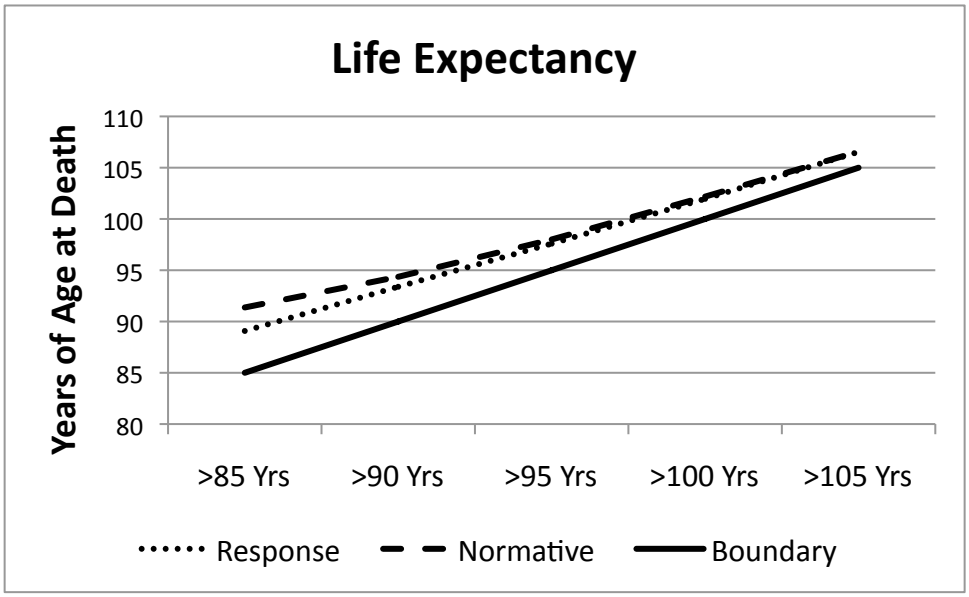


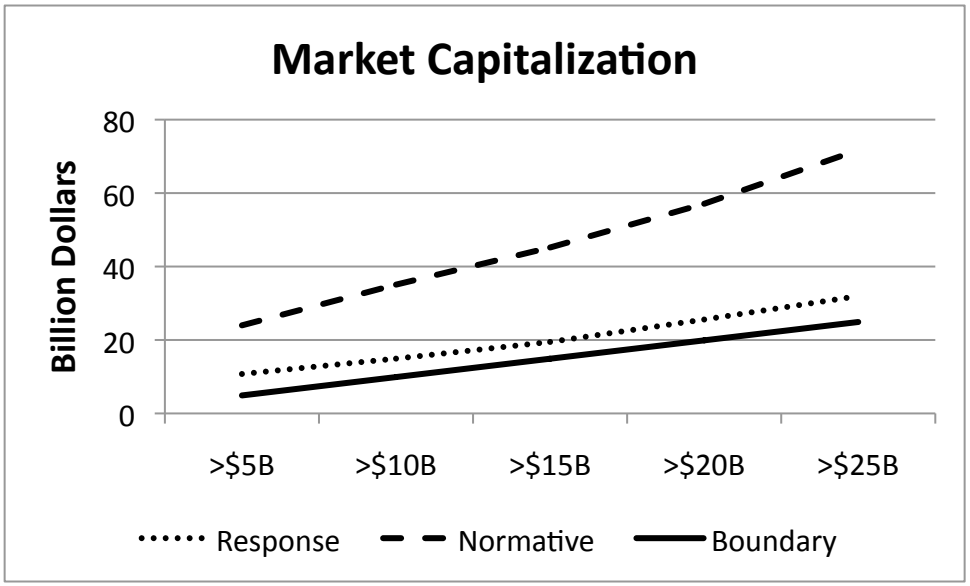
Figure BBB: Estimates of conditional expectation of age at death of US women.

Non-Scalable domain items

Where in the scalable domain items, the normative answers tend towards the boundary, in the non scalable items below, they tend away from it. Do people pick up this difference?

Item C (market capitalization):

The average responses of the 66 participants are shown in Figure CCC. Normative data were computed from US stocks with a market capitalization of greater than 5 billion dollars on April 2, 2009. On the average, there is no appreciation that the conditional expectation departs from the boundary as the boundary increases. The average responses underestimate the normative estimate from 13.2 to 39.8 billion dollars. Moving from left to right in the chart, the standard error of the responses is 1.12, .83, .64, .96, and 1.32 billion dollars.



Item D (stock s):

Figure DDD depicts average responses for 153 participants. Normative data are computed from using real data and extrapolating using the power-law tail exponent obtained from the data. As with the previous items, the average response underestimates the actual situation and fails to move increasingly far from the boundary. Underestimation ranges from 1.52 to 29.15 percentage points. The standard error of the 5 responses, as plotted from left to right, are .53, .43, .46, .52, and .68 percentage points.

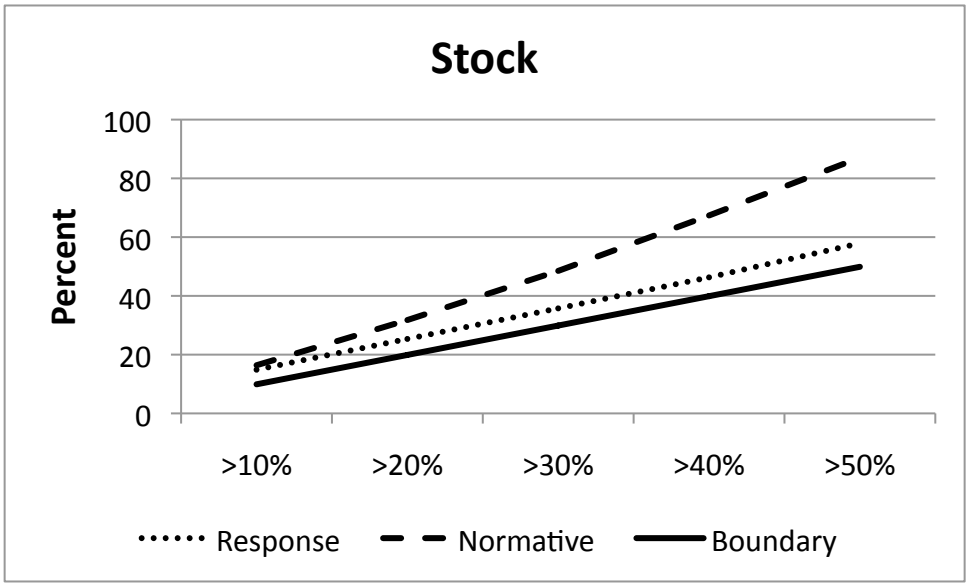


Figure DDD: Estimates of conditional expectation of stock price increases.

Averages mask individual differences . We noticed three general trends among a larger variety of response strategies undertaken by individual participants. In general, responses tend to approach, depart from, or maintain a constant distance from the boundary as the boundary increases. To classify individuals objectively, the differences between each participant's responses and the boundary were fit by a regression line. The slopes of individual fitting lines are categorized as positive, negative or zero, for each item in Figure EEE. About 80% of participants in the scalable domain items give responses that tend toward the boundary, as heights and life expectancies actually do. In the non-scalable domain items, there is a greater variance in strategy adoption, with only 52% and 36% of participants giving responses that fall into the normatively correct category.

Finer categories of response categories can be identified. We classified each participants responses to each item as: increasing, flat, decreasing, u-shaped, inverse-u-shaped, or other. To be classified as increasing, for instance, the series must increase at least once and never decrease. Using this stricter criterion, in the scalable domain 40% (height) and 46% (life expectancy) of participants adopted the normatively correct (decreasing slope) strategy. However, in the non-scalable domain, only 17% (market capitalization) and 20% (stock) of participants gave responses that could be categorized into the correct (increasing slope) strategy.

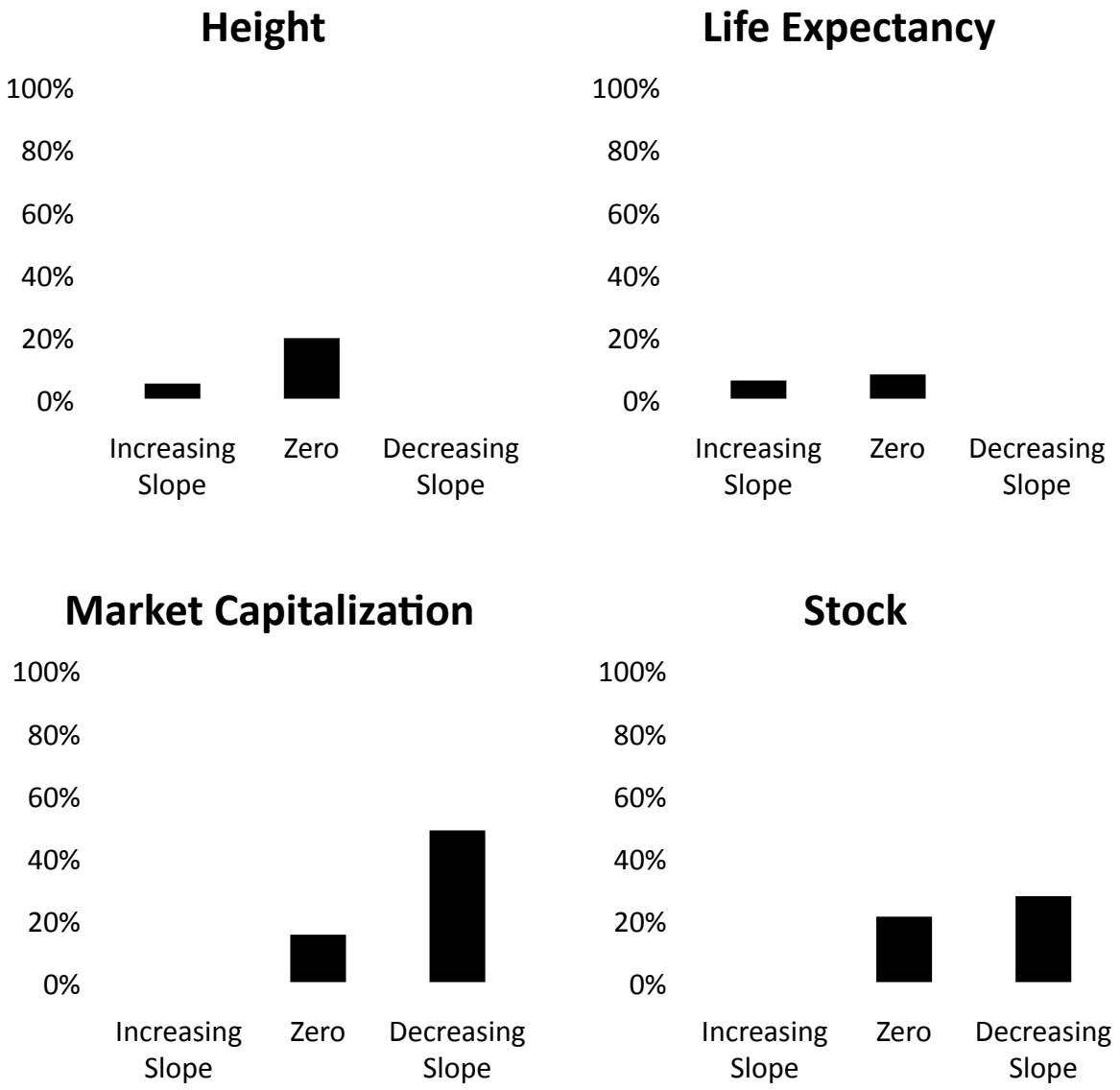


Figure EEE: Each subplot shows the percentage of participants classified into each strategy. Category labels refer to the slope of the regression line fitting the differences between each participant's responses and the boundary. The normatively correct response categories are plotted in white bars.

Discussion

References

Bell, Felicitie C. & Michael L. Miller (2005). Life Tables for the United States Social Security Area 1900-2100. Actuarial Study No. 120. Social Security Administration, Office of the Chief Actuary. SSA Publication Number 11-11536

Brainard, Jennifer & David E. Burmaster (1992). Bivariate Distributions for Height and Weight of Men and Women in the United States. Risk Analysis 12(2), 267-275.

Reinhart, C. M., and K. Rogoff. 2009. "This Time It's Different: Eight Hundred Years of Financial Folly." Princeton, NJ: Princeton University Press.

Black Swans and the Domains of Statistics

Nassim Nicholas TALEB

1. INTRODUCTION

The Black Swan: The Impact of the Highly Improbable (hence *TBS*) is only critical of statistics, statisticians, or users of statistics in a very narrow (but consequential) set of circumstances. It was written by a veteran practitioner of uncertainty whose profession (a mixture of quantitative research, derivatives pricing, and risk management) estimates and deals with exposures to higher order statistical properties. Derivatives depend on some nonlinear function of random variables (often square or cubes) and are therefore extremely sensitive to estimation errors of the higher moments of probability distributions. This is the closest to *applied statistician* one can possibly get. Furthermore, *TBS* notes the astonishing success of statistics as an engine of scientific knowledge in (1) some well-charted domains such as measurement errors, gambling theory, thermodynamics, and quantum mechanics (these fall under the designation of "mild randomness"), or (2) some applications in which our vulnerability to errors is small. Indeed, statistics has been very successful in "low moment" applications such as "significance testing" for problems based on probability, not expectation or higher moments. In psychological experiments, for instance, the outlier counts as a single observation, and does not cause a high impact beyond its frequency.

TBS is critical of some statistics in the following areas:

1. The unrigorous use of statistics, and reliance on probability in *domains* where the current methods can lead us to make consequential mistakes (the "high impact") where, on logical grounds, we need to force ourselves to be suspicious of inference about low probabilities.

2. The psychological effects of statistical numbers in lowering risk consciousness and the suspension of healthy skepticism—in spite of the unreliability of the numbers produced about low-probability events.

3. Finally *TBS* is critical of the use of commoditized metrics such as "standard deviation," "Sharpe ratio," "mean-variance," and so on in fat-tailed domains where these terms have little practical meaning, and where reliance by the untrained has been significant, unchecked and, alas, consequential.

Let me summarize the aims of *TBS*. What one of the reviewers calls "philosophy" (a term that generally alludes to the sterile character of some of the pursuits in philosophy departments), owing perhaps to the lack of quantitative measures in *TBS*, I tend to call "risk management." That is, practical wisdom and translation of knowledge into responsible decision making. Again, for a practitioner "philosophy" is, literally, "wisdom," not empty talk. As put directly in *TBS*, it is about how "not to be a sucker." My aim of the book is "how to avoid being the turkey." It cannot get more practical (and less "philosophical" in the academic sense) than that.

Accordingly, *TBS* is meant to provide a roadmap for *dealing* with tail events by exposing areas where our knowledge can be deemed fragile, and where tail events can have extreme impacts. It presents methods to avoid such events by not venturing into areas where our knowledge is not rigorous. In other words, it offers a way to live safely in a world we do not quite understand. It does not get into the trap of offering another precise model to replace another precise model; rather it tells you where we should have the courage to say "I don't know," or "I know less."

2. CONFIDENCE ABOUT SMALL PROBABILITIES

I will next outline the "inverse problem" of the real world. Life is not an artificial laboratory in which we are supplied with probabilities. Nor is it an urn (alas) as in elementary statistics textbooks. Nor is it a casino where the state authorities monitor and enforce some probabilistic transparency (i.e., try to eliminate the uncertainty about the probabilities). Empirical estimation of probabilities poses a problem in domains with unbounded or near-unbounded payoffs. (*I am not assuming, which is key, that an upper or lower bound does not exist, only that we do not know where it is*.)

Suppose that you are deriving probabilities of future occurrences from the data, assuming (in the "rosy" case) that the past is representative of the future. An event can be a market crash, a banking crisis, a loss for an insurance company, a riot, people affected in an epidemic, an act of terrorism, and so on. The severity of the event here will be inversely proportional to its expected frequency: the so-called 10-year flood will be more frequent than the 100-year flood, and the 100-year flood will be more devastating. In these events, we are not sampling from a problem-style closed urn of known composition and impacts. We don't even know if there is a 200-year flood, and what impact it may have. We are now subjected to the classical problem of induction: making bold claims about the unknown based on assumed properties of the known. So (1) the smaller the probability, the larger we need the sample size to be in order to make inferences, and the smaller the probability, the higher the relative error in estimating this probability. (2) Yet in these domains, *the smaller the probability, the more consequential* the impact of the absolute probability error on the moments of the distribution.

Estimation errors for tail probabilities are very important when their large impact is considered. The pair probability *times* impact is a rectangle that gets thinner as probabilities becomes smaller, but its area can become more stochastic if the probabilities do not drop too quickly as the impact becomes larger. This is clearly intractable. It can be solved on paper, of course, by assuming a priori a certain class of distributions. Indeed the choice

Nassim Nicholas Taleb is a veteran derivatives trader and researcher, London Business School and Empirica Laboratory Limited (E-mail: *gamma@fooledbyrandomness.com*).

 DOI: 10.1198/000313007X219996

of distributions with characteristic scale—that is, what Mandelbrot defined as "mild randomness," more on that later—appears to conveniently push such problems under the rug.

3. SELF-REFERENCE

This problem has been seemingly dealt away with the use of "off-the-shelf" probability distributions. But distributions are self-referential. Do we have enough data? If the distribution is, say, the traditional Gaussian, then yes, we may be able to say that we have sufficient data—for instance, the Gaussian *itself* tells us how much data we need. But if the distribution is not from such a well-bred family, then we may not have enough data. But how do we know which distribution we have on our hands? Well, *from the data itself*.

So we can state the problem of self-reference of statistical distributions in the following way. If (1) one needs data to obtain a probability distribution to gauge knowledge about the future behavior of the distribution from its past results, and if, at the same time, (2) one needs a probability distribution to gauge data sufficiency and whether or not it is predictive outside its sample, then we are facing a severe regress loop. We do not know what weight to put on additional data. And unlike many problems of regress, this one can have severe consequences when we talk about risk management.

4. NOT ANY FAT TAILS WOULD DO

Although they do not share some aspects of the style of the message, the four discussants appear to agree with *TBS* about the role of outliers and their primacy over the ordinary in determining the statistical properties. The discussants advocate the following: robust statistics, stochastic volatility or GARCH, or Extreme Value Theory. These approaches either do not solve the problem of confidence about small probability, or they create new ones: many of these are tools, not solutions. Robust statistics are certainly more natural tools (Goldstein and Taleb 2007), but I fail to see how robust statistics will produce more information about the probability of events that are not in the sample of the past realizations (see Freedman and Stark 2003). Moreover, there is a major methodological difference between our standpoints: I do not believe in using *any* distribution that naively produces *some* extreme event (or calibrates one from past data). From an operational (and risk management) standpoint, not any fat tails would do.

The central idea of *TBS* concerns the all-too-common logical confusion of *absence of evidence* with *evidence of absence*, associated with the error of confirmation. It tries to avert this logical error in the interpretation of statistical information. As it is impossible to make precise statements about unseen events, those that lie outside the sample set, we need to make the richest possible scenarios about them. For this *TBS* uses, on both logical and empirical grounds, the classification made by Mandelbrot (1963) between two classes of probability distributions: those that have "true fat tails" and others that do not. I had difficulty understanding why the statistical literature has neglected for so long the Mandelbrotian classification.

True fat-tailed distributions have a scale-free or fractal property that I can simplify as follows: for X large enough, (i.e., "in the tails"), $P[X > nx]/P[X > x]$ depends on n, not on x. In financial securities, say, where X is a monthly return, there is no reason for $P[X > 20\%]/P[X > 10\%]$ to be different from $P[X > 15\%]/P[X > 7.5\%]$. This self-similarity at all scales generates power-law, or Paretian, tails; that is, above a crossover point, $P[X > x] = Kx^{-\alpha}$. (Note that the same properties hold for $P[X < x]$ in the negative domain.)

The standard Poisson and stochastic volatility models are not scale-invariant. There is a known value of x beyond which these distributions become thin-tailed—when in reality we do not know what the upper bound is. Further, the Poisson lends itself to in-sample overfitting: you can always use a Poisson jump to fit, in past samples, the largest realization of a fractal fat-tailed process. But it would fail out of sample. For instance, before the 23% drop in the stock market crash of 1987, the worst previous in-sample move was close to 10%. Calibrating a Poisson jump of 10% would not have prepared the risk manager for the ensuing large drop. On the other hand, for someone using the framework of Mandelbrot (1963), the crash of 1987 would not have been surprising—nor would hundreds of large moves we've had in currencies and stocks (*TBS* presents an overview of the literature on dozens of empirical tests across socioeconomic random variables).

Unless there are logical reasons to assume "Mediocristan," or mild randomness, *TBS* advocates using a fractal distribution for the tails *as a default*, which is the opposite of what I've seen practiced. Why? There is a logical asymmetry: a true fat-tailed distribution can camouflage as thin-tailed in small samples; the opposite is not true. If I see a "20-sigma" event, I can be convinced that the data are not Gaussian. If I see no such deviation I cannot make statements that the tails are necessarily thin—in fat-tailed distributions, nothing eventful takes place most of the time. The burden of proof is not on a fat-tailed distribution.

Decision makers are mostly concerned about the cost of mistakes, rather than exact knowledge about the statistical properties. We are dealing with plenty of invisibles, so I do not use power-law tails as a way to estimate precise probabilities—since the parameter α is not easily computed—rather as an aid to make decisions. How?

First, we use power laws as risk-management tools; they allow us to quantify sensitivity to left- and right-tail measurement errors and rank situations based on the *full* effect of the unseen. We can effectively get information about our vulnerability to the tails by varying the power-law exponent α and looking at the effect on the moments or the shortfall (expected losses in excess of some threshold). This is a fully structured stress testing, as the tail exponent α decreases, all possible states of the world are encompassed. And skepticism about the tails can lead to action and allow ranking situations based on the fragility of knowledge; as these errors are less consequential in some areas than others. I explain as follows. If your left tail is "organically" truncated (i.e., the state of the world is not possible or cannot affect you), then you may not worry about negative low-probability events and look forward to positive ones. In a business that benefits from the rare event (bounded left-tail exposure, unbounded right one), rare events that the past did not reveal are almost certainly going

to be good for you. When you look at past biotech revenues, for example, you do not see the superblockbuster in them, and owing to the potential for a cure for a disease, there is a small probability that the sales in that industry may turn out to be far larger than what was revealed from past data. This is illuminated by thickening the right tail: varying the α to gauge the effect of the unseen.

On the other hand, consider businesses negatively exposed to rare events (bounded right tails). The track record you see is likely to overestimate the properties—and any thickening of the left tail lowers your expectation. *TBS* discusses the 1982 blowup of banks that lost a century of profits in a single episode: on the eve of the episode, they appeared to the naïve observer to be more profitable than they seemed.

The second reason I advocate the "true fat tails" method of Mandelbrot (1963) in finance and economics is, as I said, empirical. As we saw with the crash of 1987, events have remained consistent with statistics since then—unlike other methods (Poisson or stochastic volatility) that failed us out of sample. But methods allowing for "wild randomness" are not popular in economics and the disciplines that rely on times series analyses because they do away with the measure called "variance," embedded in the consciousness, and so necessary for many applications.

5. CONCLUSION

To conclude, I am exposing the fragility of knowledge about the tails of the distributions in domains where errors can be consequential. I discuss my operational reasons to select scalable laws, that is, "true fat tails" as default distributions and as tools to minimize exposure to such consequential errors. It is only in these cases of lessened tail dependence that statistics are safe—and that is where its strength lies.

Finally I would like to thank the discussants and *The American Statistician* for their open-mindedness and for giving me the opportunity to explain myself. This makes me extremely proud to be an applied statistician.

REFERENCES

Freedman, D. A., and Stark, P. B. (2003), "What is the Probability of an Earthquake?" in *Earthquake Science and Seismic Risk Reduction*, NATO Science Series IV: Earth and Environmental Sciences, vol. 32, eds. F. Mulargia and R. J. Geller, Dordrecht, The Netherlands: Kluwer.

Goldstein, D. G., and Taleb, N. N. (in press), "We Don't Quite Know What We are Talking About When We Talk About Volatility," *Journal of Portfolio Management*.

Mandelbrot, B., (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

——— (1997), *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, New York: Springer-Verlag.

——— (2001), "Scaling in Financial Prices," *Quantitative Finance*, 1, 113–123, 124–130, 427–440, and 641–649.

Taleb, N. N. (1997), *Dynamic Hedging: Managing Vanilla and Exotic Options*, New York: Wiley.

——— (2007), *The Black Swan: The Impact of the Highly Improbable*, New York: Random House and London: Penguin.

Quantitative Finance, Vol. 5, No. 4, August 2005, 323–326

Routledge
Taylor & Francis Group

The illusions of dynamic replication

EMANUEL DERMAN*† and NASSIM NICHOLAS TALEB‡

†Columbia University and Prisma Capital Partners LP
‡University of Massachusetts, Amherst and Empirica LLC

(*Received 24 May 2005; in final form 24 June 2005*)

1. Introduction

How well does options pricing theory really work, and how dependent is it on the notion of dynamic replication? In this note we describe what many practitioners know from long and practical experience: (i) dynamic replication doesn't work as well as students are taught to believe; (ii) most derivatives traders rely on it as little as possible; and (iii) there is a much simpler way to derive many option pricing formulas: many of the results of dynamic option replication can be obtained more simply, by regarding (as many practitioners do) an options valuation model as an interpolating formula for a hybrid security that correctly matches the boundary values of the ingredient securities that constitute the hybrid.

2. Replication

The logic of replication is that a security whose payoff can be replicated purely by the continuous trading of a portfolio of underlying securities is *redundant*; its value can be derived from the value of the underlying replicating portfolio, requiring no utility function or risk premium applied to expected values. The fair value of the replicated security follows purely from riskless arbitrage arguments.

The method of static replication for valuing securities was well known, but prior to Black and Scholes (1973) the possibility of dynamic replication was unexplored, although there had been hints of the approach, as in Arrow (1953). What distinguishes the Black–Scholes–Merton model is the dynamic replication of the portfolio and the economic consequences of this argument, rather than, as is frequently asserted in the literature, the option pricing equation *per se*.

We shall show that the Black–Scholes option pricing formula could have been derived much earlier by requiring that a portfolio consisting of a long position in a call and a short position in a put, valued by the traditional discounted expected value of their payoffs, must statically replicate a forward contract.

3. Arguments for skepticism

There are a variety of empirical arguments that justify some skepticism about the efficacy of dynamic hedging as a framework for options valuation.

- Options are currently priced and traded on myriads of instruments—live commodities, agricultural products, perishable goods, and extremely illiquid equity securities—where dynamic replication cannot possibly be achieved. Yet these options are priced with the *same* models and software packages as are options on those rare securities where dynamic replication is feasible.
- Even where dynamic replication is feasible, the theory requires continuous trading, a constraint that is unachievable in practice. The errors resulting from discrete hedging, as well as the transaction costs involved, are prohibitive, a point that has been investigated extensively in the literature (see, for example, Taleb (1997, 1998)).
- In addition, market-makers, who are in the business of manufacturing long and short option positions for their clients, do not hedge every option dynamically; instead they hedge only their extremely small *net* position. Thus, the effect of the difference between dynamic and static hedging on their portfolio is extremely small.
- Dynamic replication assumes continuous asset price movements, but real asset prices can move discontinuously, destroying the possibility of accurate replication and providing a meaningful likelihood of bankruptcy for any uncovered option seller who does not have *unlimited* capital.

*Corresponding author. Email: emanuel.derman@mac.com

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online
http://www.tandf.co.uk/journals
DOI: 10.1080/14697680500305105

- All manner of exotic and even hybrid multidimensional derivative structures have proliferated in the past decade, instruments of such complexity that dynamic replication is clearly practically impossible. Yet they are priced using extensions of standard options models.

Hakansson's so-called paradox (Hakansson 1979, Merton 1992) encapsulates the skepticism about dynamic replication: if options can only be priced because they can be replicated, then, since they can be replicated, why are they needed at all?

4. The logic of dynamic hedging

Let us review the assumptions about dynamic replication that lead to the Black–Scholes equation for European options on a single stock.

In the Black–Scholes picture a stock S is a primitive security, primitive in the sense that its payoff cannot be replicated by means of some other security. An option C whose payoff depends through a specified payoff function of S at some expiration time T is a derivative security.

Assume that the underlying stock price S undergoes geometric Brownian motion with expected return μ and return volatility σ. A short position in the option C with price $C(S, t)$ at time t can be hedged by purchasing $\partial C/\partial S$ shares of stock against it.

The hedged portfolio $\Pi = -C + \partial C/\partial S$ S consisting of a short position in the option and a long position in Δ shares of the underlying stock will have no instantaneous linear exposure to the stock price S.

Note that the immediate effect of this hedge is to remove all immediate dependence of the value of portfolio Π on the expected return μ of the stock.

$$E[\Delta\Pi] = -\partial C/\partial S E[\Delta S] - \frac{1}{2}\partial^2 C/\partial S^2 E[\Delta S^2]$$

$$- \partial C/\partial t \Delta t + \partial C/\partial S E[\Delta S].$$

We† can see how the first and last terms cancel each other, eliminating $E[\Delta S]$ from the expectation of the variations in the hedged portfolio.

The portfolio of option and stock has not yet becomes a *riskless* instrument whose return is determined. We need another element, the stream of subsequent dynamic hedges.

With continuous rehedging, the instantaneous profit on the portfolio per unit time is given by

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t},$$

assuming for simplicity that the riskless interest rate is zero.

If the future return volatility σ of the stock is known, this profit is deterministic and riskless. If there is to be no arbitrage on any riskless position, then the instantaneous profit must be zero, leading to the canonical Black–Scholes equation

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + \frac{\partial C}{\partial t} = 0, \quad (1)$$

which can be solved for boundary conditions corresponding to a simple European call to yield the Black–Scholes formula.

Note that the Nobel committee upon granting the Bank of Sweden Prize in honour of Alfred Nobel, provided the following citation: 'Black, Merton and Scholes made a vital contribution by showing that it is in fact not necessary to use any risk premium when valuing an option. This does not mean that the risk premium disappears; instead it is already included in the stock price.'‡ It is for having removed the effect of μ on the value of the option, *and not for rendering the option a deterministic and riskless security*, that their work is cited.

The effect of the subsequent stream of secondary dynamic hedges is to render the option riskless, not, as it is often assumed, to remove the risk of the exposure to the underlying security. The more we hedge, the more the option becomes (under the Black–Scholes assumptions) a deterministic payoff—but, again, under a set of very precise and idealized assumptions, as we will see next.

5. Dynamic hedging and its discontents

The Black–Scholes–Merton formalism relies upon the following central assumptions:

(1) constant (and known) σ;
(2) constant and known carry rates;
(3) no transaction costs;
(4) frictionless (and continuous) markets.

Actual markets violate all of these assumptions.

- Most strikingly, the implied volatility smile is incompatible with the Black–Scholes–Merton model, which leads to a flat implied volatility surface. Since the option price is incompatible with the Black–Scholes formula, the correct hedge ratio is unknown.
- One cannot hedge continuously. Discrete hedging causes the portfolio Π to become risky before the next rebalancing. One can think of this as a sampling error of order $1/(\sqrt{2N})$ in the stock's volatility, where N is the number of rebalancings. Hedge 50 times on a three-month option rather than continuously, and the standard deviation of the error in the replicated option price is about 10%, a significant mismatch.

†We are taking the equality in expectation because we are operating in discrete time not at the limit of Δt going to 0.
‡See www.Nobel.se

- In addition to the impossibility of continuous hedging, transaction costs at each discrete rehedging impose a cost that make an options position worth less than the Black–Scholes value.
- Future carry rates are neither constant nor known.
- Furthermore, future volatility is neither constant nor known.
- More radically, asset price distributions have fat tails and are inadequately described by the geometric Brownian motion assumed by Markowitz's mean-variance theory, the Capital Asset Pricing Model and options theory itself.

Furthermore, practitioners know from bitter experience that dynamic replication is a much more fragile procedure than static replication: a trading desk must deal with transactions costs, liquidity constraints, the need for choosing price evolution models and the uncertainties that ensue, the confounding effect of discontinuous asset price moves, and, last but by no means least, the necessity for position and risk management software.

6. Options valuation by expectations and static replication

Practitioners in derivatives markets tend to regard options models as interpolating formulas for hybrid securities. A convertible bond, for example, is part stock, part bond: it becomes indistinguishable from the underlying stock when the stock price is sufficiently high, and equivalent to a corporate bond when the stock price is sufficiently low. A convertible bond valuation model provides a formula for smoothly interpolating between these two extremes. In order to provide the correct limits at the extremes, the model must be calibrated by static replication. A convertible model that doesn't replicate a simple corporate bond at asymptotically low stock prices is fatally suspect.

One can view the Black–Scholes formula in a similar light. Assume that a stock S that pays no dividends has future returns that are lognormal with volatility σ. A plausible and time-honoured *actuarial* way to estimate the value at time t of a European call C with strike K expiring at time T is to calculate its expected discounted value, which is given by

$$C(S,t) = e^{-r(T-t)}(E[S-K]_+)$$
$$= e^{-r(T-t)}\left\{S e^{\mu(T-t)}N(d_1) - KN(d_2)\right\}, \quad (2)$$

where r is the appropriate but unknown discount rate, still unspecified and μ is the unknown expected growth rate for the stock.

The analogous actuarial formula for a put P is given by

$$P(S,t) = e^{-r(T-t)}(E[K-S]_+)$$
$$= e^{-r(T-t)}\left\{KN(-d_2) - Se^{\mu(T-t)}N(-d_1)\right\}, \quad (3)$$

where

$$d_{1,2} = \frac{\ln[S e^{\mu(T-t)}/K] \pm [\sigma^2(T-t)/2]}{\sigma\sqrt{T-t}}. \quad (4)$$

A dealer or market-maker in options, however, has additional consistency constraints. As a manufacturer rather than a consumer of options, the market-maker must stay consistent with the value of his raw supplies. He must notice that a portfolio $F = C - P$ consisting of a long position in a call and a short position in a put with the same strike K has exactly the same payoff as a forward contract with expiration time T and delivery price K whose fair current value is

$$F = S - K e^{-R(T-t)}, \quad (5)$$

where R is the zero-coupon riskless discount rate for the time to expiration.

The individual formulas of equations (2) and (3) must be calibrated to be consistent with equation (5). If they are not, the market-maker will be valuing his options, stock and forward contracts inconsistently, despite their underlying similarity. What conditions are necessary to satisfy this?

Combining equations (2) and (3) we obtain

$$F = C - P = e^{-r(T-t)}\left\{S e^{\mu(T-t)} - K\right\}. \quad (6)$$

The requirement that equations (5) and (6) be consistent dictates that both the appropriate discount rate r and the expected growth rate μ for the stock in the options formula be the zero-coupon discount rate R. These choices make equation (2) equivalent to the Black–Scholes formula.

A similar consistency argument can be used to derive the values of more complex derivatives, dependent on a larger number of underlyers, by requiring consistency with the values of all tradable forwards contracts on those underlyers. For an application of this method to valuing quanto options, see Derman *et al.* (1998).

7. From Bachelier to Keynes

Let us zoom back into the past. Assume that in 1973, there were puts and calls trading in the market-place. The simple put–call parity argument would have revealed that these can be combined to create a forward contract.

John Maynard Keynes was the first to show that the forward need not be priced by the expected return on the stock, the equivalent of the μ we discussed earlier, but by the arbitrage differential, namely, the equivalent of $r - d$. This follows the exposition of the formula that was familiar to every institutional foreign exchange trader.

> If by lending dollars in New York for one month the lender could earn interest at the rate of $5\frac{1}{2}\%$ per annum, whereas by lending sterling in London for

> one month he could only earn interest at the rate of 4%, then the preference observed above for holding funds in New York rather than in London is wholly explained. That is to say, forward quotations for the purchase of the currency of the dearer money market tend to be cheaper than spot quotations by a percentage per month equal to the excess of the interest which can be earned in a month in the dearer market over what can be earned in the cheaper. Keynes (1923, 2000)

Between Bachelier and Black–Scholes, there were several researchers who produced formulas similar to that of Black–Scholes, differing from it only by their use of a discount rate that was not riskless. While Bachelier had the Black–Scholes equation with no drift and under an arithmetic Brownian motion, others added the drift, albeit a nonarbitrage derived one, in addition to the geometric motion for the dynamics. Of these equations we can cite Sprenkle (1961), Boness (1964), Samuelson (1965), and Samuelson and Merton (1969). All of their resultant pricing equations involved unknown risk premiums that would have been determined to be zero had they used the put–call replication argument we illustrated above. Furthermore, the put–call parity constraint was already present in the literature (see Stoll, 1969).

8. Conclusion

Dynamic hedging is neither strictly required nor strictly necessary for plausibly valuing options; it is less relied upon in practice than is commonly believed. Much of financial valuation does not require such complexity of exposition, elegant though it may be†. The formulas it leads to can often be obtained much more simply and intuitively by constrained interpolation. Finally, the pricing of contingent claims by interpolation and static replication opens the door to valuing options on assets without necessarily demanding that such assets have finite square variation, and thus sets the grounds for the use of a richer class of distributions with finite first moment.

Acknowledgments

The authors thank Gur Huberman for helpful comments and for alerting us to early work on put/call parity.

References

Arrow, K., The role of securities in the optimal allocation of risk-bearing. *Econometrie*, 1953.

Black, F. and Scholes, M., The pricing of options and corporate liabilities. *J. Polit. Econ.*, 1973, **81**, 637–654.

Boness, A.J., Elements of a theory of stock option value. *J. Polit. Econ.*, 1964, **81**(3), 637–654.

Derman, E., Karasinski, P. and Wecker, J., Understanding guaranteed-exchange-rate options. In *Currency Derivatives*, edited by D. DeRosa, 1998 (Wiley: New York).

Hakkanson, N.H., The fantastic world of finance: progress and the free lunch. *J. Financ. Quant. Anal.*, 1979, **14**, 717–734.

Keynes, J.M., *A Tract on Monetary Reform*, 1923 (2000) (Prometheus Books: Amherst).

MacKenzie, D., An equation and its worlds: bricolage, exemplars, disunity and performativity in financial economics. *Soc. Stud. Sci.*, 2003, **33**(6), 831–868.

Merton, R.C., *Continuous Time Finance*, 1992 (Blackwell: London).

Samuelson, P.A., Rational theory of warrant pricing. *Ind. Manag. Rev.*, 1965, **6**(2), 13–32.

Samuelson, P.A. and Merton, R.C., A complete model of asset prices that maximizes utility. *Ind. Manag. Rev.*, 1969, **10**, 17–46.

Sprenkle, C.M., Warrant prices as indicators of expectations and preferences. *Yale Economic Essays*, 1961, **1**(2), 178–231 [reprinted in *The Random Character of Stock Market Prices*, edited by P.H. Cootner, 1967 (MIT Press: Cambridge, MA)].

Stoll, H.R., The relationship between put and call prices. *J. Finance*, 1969, **24**(5), 801–824.

Taleb, N.N., *Dynamic Hedging: Managing Vanilla and Exotic Options*, 1997 (Wiley: New York).

Taleb, N.N., Replication d'options et structure de marché, Report, Université Paris IX Dauphine, 1998.

†For a sociological study of the Black and Scholes formula, see MacKenzie (2003).

Volume 77, Issue 2,
February 2011

77 (2) 97-240
(2011)

ISSN 0167-2681

JOURNAL OF Economic Behavior & Organization

CONTENTS

www.elsevier.com/locate/jebo

Journal of Economic Behavior & Organization 77 (2011) 97–106

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Option traders use (very) sophisticated heuristics, never the Black–Scholes–Merton formula

Espen Gaarder Haug[b], Nassim Nicholas Taleb[a,*]

[a] NYU Poly, 6 Metrotech Center, Brooklyn, USA
[b] Independent Option Trader, London, United Kingdom

ARTICLE INFO

Article history:
Received 30 November 2009
Received in revised form
14 September 2010
Accepted 15 September 2010
Available online 20 October 2010

JEL classification:
G12
G13

Keywords:
Options
Option hedging
Delta hedging
Put-call parity
Derivatives

ABSTRACT

Option traders use a heuristically derived pricing formula which they adapt by fudging and changing the tails and skewness by varying one parameter, the standard deviation of a Gaussian. Such formula is popularly called "Black–Scholes–Merton" owing to an attributed eponymous discovery (though changing the standard deviation parameter is in contradiction with it). However, we have historical evidence that: (1) the said Black, Scholes and Merton did not invent any formula, just found an argument to make a well known (and used) formula compatible with the economics establishment, by removing the "risk" parameter through "dynamic hedging", (2) option traders use (and evidently have used since 1902) sophisticated heuristics and tricks more compatible with the previous versions of the formula of Louis Bachelier and Edward O. Thorp (that allow a broad choice of probability distributions) and removed the risk parameter using put-call parity, (3) option traders did not use the Black–Scholes–Merton formula or similar formulas after 1973 but continued their bottom-up heuristics more robust to the high impact rare event. The paper draws on historical trading methods and 19th and early 20th century references ignored by the finance literature. It is time to stop using the wrong designation for option pricing.

1. Breaking the chain of transmission

For us practitioners, theories about practice should arise from practice[1] or at least avoid conflict with it. This explains our concern with the "scientific" notion that practice should fit theory. Option hedging, pricing, and trading are neither philosophy nor mathematics, but an extremely rich craft rich with heuristics with traders learning from traders (or traders copying other traders) and tricks developing under evolution pressures, in a bottom-up manner. It is *technë*, not *ëpistemë*. Had it been a science it would not have survived—for the empirical and scientific fitness of the pricing and hedging theories offered are, we will see, at best, defective and unscientific (and, at the worst, the hedging methods create more risks than they reduce). Our approach in this paper is to ferret out historical evidence of *technë* showing how option traders went about their business in the past.

Options, we will show, have been extremely active in the pre-modern finance world. Complicated, tacitly transmitted tricks and heuristically derived methodologies in option trading and risk management of derivatives books have been developed over the past century, and used quite effectively by operators. In parallel, many derivations were produced

* Corresponding author.
E-mail address: info@theblackswan.org (N.N. Taleb).

[1] For us, in this discussion, a practitioner is deemed to be someone involved in repeated decisions about option hedging, not a support quant who writes pricing software or an academic who provides "consulting" advice.

0167-2681/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2010.09.013

by mathematical researchers.[2] The economics literature, however, did not recognize these contributions, substituting the rediscoveries or subsequent reformulations done by (some) economists. There is evidence of an attribution problem with Black–Scholes–Merton option "formula", which was developed, used, and adapted in a robust way by a long tradition of researchers and used heuristically by option market makers and "book runners". Furthermore, in a case of scientific puzzle, the exact formula called "Black–Sholes–Merton" was written down (and used) by Edward Thorp which, paradoxically, while being robust and realistic, has been considered unrigorous. This raises the following: (1) The Black–Scholes–Merton was, according to modern finance, just a neoclassical finance argument, no more than a thought experiment,[3] (2) we are not aware of traders using their argument or their version of the formula.

2. The Black–Scholes–Merton "formula" was an argument

Option traders call the formula they use the "Black–Scholes–Merton" formula without being aware that by some irony, of all the possible options formulas that have been produced in the past century, what is called the Black–Scholes–Merton "formula" (after Black and Scholes, 1973; Merton, 1973) is the one the furthest away from what they are using. In fact of the formulas written down in a long history it is the only formula that is fragile to jumps and tail events.

First, something seems to have been lost in translation: Black and Scholes (1973) and Merton (1973) actually never came up with a *new* option formula, but only an theoretical economic *argument* built on a new way of "deriving", rather re-deriving, an already existing – and well known – formula. The argument, we will see, is extremely fragile to assumptions. The foundations of option hedging and pricing were already far more firmly laid down before them. The Black–Scholes–Merton argument, simply, is that an option can be hedged using a certain methodology called "dynamic hedging" and then turned into a risk-free instrument, as the portfolio would no longer be stochastic. Indeed what Black, Scholes and Merton did was "marketing", finding a way to make a well-known formula palatable to the economics establishment of the time, little else, and in fact distorting its essence.

Such argument requires strange far-fetched assumptions: some liquidity at the level of transactions, knowledge of the probabilities of future events (in a neoclassical Arrow–Debreu style),[4] and, more critically, a certain mathematical structure that requires "thin-tails", or mild randomness, on which, later. The entire argument is indeed, quite strange and rather inapplicable for someone clinically and observation-driven standing outside conventional neoclassical economics. Simply, the dynamic hedging argument is dangerous in practice as it subjects you to blowups; it makes no sense unless you are concerned with neoclassical economic theory. The Black–Scholes–Merton argument and equation flow a top-down general equilibrium theory, built upon the assumptions of operators working *in full knowledge* of the probability distribution of future outcomes—in addition to a collection of assumptions that, we will see, are highly invalid mathematically, the main one being the ability to cut the risks using continuous trading which only works in the very narrowly special case of thin-tailed distributions (or, possibly, jumps of a well-known structure). But it is not just these flaws that make it inapplicable: option traders do not "buy theories", particularly speculative general equilibrium ones, which they find too risky for them and extremely lacking in standards of reliability. A normative theory is, simply, not good for decision-making under uncertainty (particularly if it is in chronic disagreement with empirical evidence). Operators may take decisions based on heuristics under the impression of using speculative theories, but avoid the fragility of theories in running their risks.

Yet professional traders, including, initially, the authors (and, alas, the Swedish Academy of Science) have operated under the illusion that it was the Black–Scholes–Merton "formula" they actually used—we were told so. This myth has been progressively reinforced in the literature and in business schools, as the original sources have been lost or frowned upon as "anecdotal" (Merton, 1992). As Fig. 1 shows, these simple random jumps represent too large a share of the variability of returns to make the Black-Scholes-Merton argument scientifically acceptable – the Swedish Academy does not grant the Nobel in Medicine to works that are grounded in the assumption that men were mice.

This discussion will present our real-world, ecological understanding of option pricing and hedging based on what option traders actually do and did for more than a hundred years.

This is a very general problem. As we said, option traders develop a chain of transmission of *technë*, like many professions. But the problem is that the "chain" is often broken as universities do not store the acquired skills by operators. Effectively plenty of robust heuristically derived implementations have been developed over the years, but the economics establishment has refused to quote them or acknowledge them. This makes traders need to relearn matters periodically. Failure of dynamic hedging in 1987, by such firm as Leland O'Brien Rubinstein, for instance, does not seem to appear in the academic literature

[2] Heuristics as tacit knowledge: Gigerenzer and Todd (2000).

[3] Here we question the notion of confusing thought experiments in a hypothetical world, of no predictive power, with either science or practice. The fact that the Black–Scholes–Merton argument works in a Platonic world and appears to be "elegant" does not mean anything since one can always produce a Platonic world in which a certain equation works, or in which a "rigorous" proof can be provided, thanks to a process called reverse-engineering.

[4] Of all the misplaced assumptions of Black Scholes that cause it to be a mere thought experiment, though an extremely elegant one, a flaw shared with modern portfolio theory, is the certain knowledge of future delivered variance for the random variable (or, equivalently, all the future probabilities). This is what makes it clash with practice—the rectification by the market fattening the tails is a negation of the Black–Scholes thought experiment.

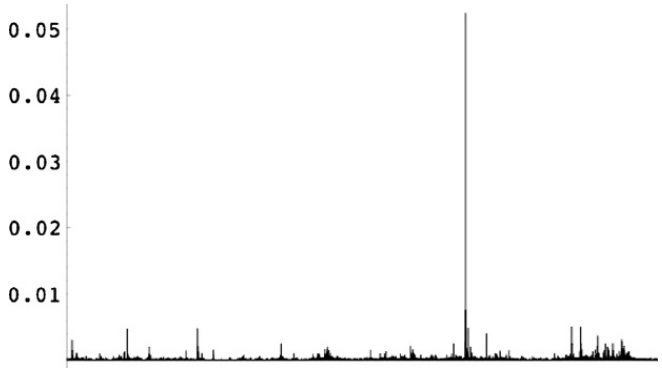


Fig. 1. The typical "risk reduction" performed by the Black–Scholes–Merton argument. These are the variations of a dynamically hedged portfolio. BSM indeed "smoothes" out risks but exposes the operator to massive tail events—reminiscent of such blowups as LTCM. Other option formulas are robust to the rare event and make no such claims.

published after the event[5] (Merton, 1992; Rubinstein, 1998; Ross, 2005); to the contrary dynamic hedging is held to be a standard operation.

There are central elements of the real world that can escape them—academic research without feedback from practice (in a *practical* and applied field) can cause the diversions we witness between laboratory and ecological frameworks. This explains why some many finance academics have had the tendency to make smooth returns, then "blow up" (that, is experience a terminal or near-terminal sharp loss) using their own theories.[6] We started the other way around, first spending years trading options and performing millions of hedges and option trades. We did this in combination with the investigating the forgotten and ignored ancient knowledge in option pricing and trading we will explain some common myths about option pricing and hedging.

There are indeed two myths:

- That we had to wait for the Black–Scholes–Merton options formula to trade the product, price options, and manage option books. In fact the introduction of the Black, Scholes and Merton argument increased our risks and set us back in risk management. More generally, it is a myth that traders rely on theories, even less a general equilibrium theory, to price options.
- That we "use" the Black–Scholes–Merton options "pricing formula". We, simply don't.

In our discussion of these myths we will focus on the bottom-up literature on option theory that has been hidden in the dark recesses of libraries. And that addresses only recorded matters—not the actual practice of option trading that has been lost.

3. Myth 1: people did not properly "price" options before the Black–Scholes–Merton theory

It is assumed that the Black–Scholes–Merton theory is what made it possible for option traders to calculate their delta hedge (against the underlying) and to price options. This argument is highly debatable, both historically and analytically.

Options had been actively trading at least in 1600 as described by De La Vega (1688)—implying some form of *technë*, a heuristic method to price them and deal with their exposure. De La Vega describes option trading in the Netherlands, indicating that operators had some expertise in option pricing and hedging. He diffusely points to put-call parity, and his book was not even meant to teach people about the technicalities in option trading. De Pinto (1771) even more explicitly points out how to convert call options into put options.[7] Our insistence on the use of put-call parity is critical for the following reason: the Black–Scholes–Merton's claim to fame is removing the necessity of a risk-based drift from the underlying security—to make the trade "risk-neutral". But one does not need dynamic hedging for that: simple put-call parity can suffice (Derman and Taleb, 2005), as we will discuss later. And yet it is this central removal of the "risk-premium" that apparently was behind the decision by the Nobel committee to grant Merton and Scholes the (then called) Bank of Sweden Prize in Honor of Alfred Nobel: "Black, Merton and Scholes made a vital contribution by *showing* that it is in fact not necessary to use any risk premium when valuing an option. This does not mean that the risk premium disappears; instead it is already included

[5] For instance how mistakes never resurface into the consciousness, Mark Rubinstein was awarded in 1995 the Financial Engineer of the Year award by the International Association of Financial Engineers. There was no mention of portfolio insurance and the failure of dynamic hedging.

[6] For a standard reaction to a rare event, see the following: "Wednesday is the type of day people will remember in quant-land for a very long time," said Mr. Rothman, a University of Chicago Ph.D. who ran a quantitative fund before joining Lehman Brothers. "Events that models only predicted would happen once in 10,000 years happened every day for three days." One 'Quant' Sees Shakeout For the Ages – '10,000 Years' By Kaja Whitehouse, August 11, 2007; Page B3.

[7] See Poitras (2009).

in the stock price."[8] It is for having removed the effect of the drift on the value of the option, *using a thought experiment*, that their work was originally cited, something that was mechanically present by any form of trading and converting using far simpler techniques.

Options have a much richer history than shown in the conventional literature. Forward contracts seems to date all the way back to Mesopotamian clay tablets dating all the way back to 1750 B.C. Gelderblom and Jonker (2005) show that Amsterdam grain dealers had already used options and forward in 1550 (but Amsterdam is not the earliest, as even more sources document even earlier uses in Europe[9]).

In the late 1800 and the early 1900 there were active option markets in London and New York as well as in Paris and several other European exchanges. Markets it seems, were active and extremely sophisticated option markets in 1870. Kairys and Valerio (1997) discuss the market for equity options in USA in the 1870s, indirectly showing that traders were sophisticated enough to price for tail events.[10] In a recent paper Mixon (2009a,b) looks at option pricing in the past versus the present and concludes that:

> "Traders in the nineteenth century appear to have priced options the same way that twenty-first century traders price options. Empirical regularities relating implied volatility to realized volatility, stock prices, and other implied volatilities (including the volatility skew), are qualitatively the same in both eras."

There was even active option arbitrage trading taking place between some of these markets. There is a long list of missing treatises on option trading: we traced at least 10 German treatises on options written between the late 1800s and the hyperinflation episode.[11]

Mixon (2009a,b) describes a relatively active Foreign Exchange Option Market from 1917 to 1921. The currency option market at that time evolved from one involving relatively large sums of money per transaction to one focused on tiny retail transactions.

A study by Moore and Juh (2006) looks at warrants traded on the Johannesburg Stock Exchange as well as call options written on 112 stocks in the early 20th century. The authors find that the warrant prices were surprisingly accurately priced in the pre Black-Scholes era.

When Cyrus Field finally succeeded in joining Europe and America by cable in 1866, intercontinental arbitrage was made possible. Although American securities had been purchased in considerable volume abroad after 1800, the lack of quick communication placed a definite limit on the amount of active trading in securities between London and New York markets, (see Weinstein, 1931). Furthermore, one extant source, Nelson (1904), speaks volumes: an option trader and arbitrageur, S.A. Nelson published a book "The A B C of Options and Arbitrage" based on his observations around the turn of the twentieth century. The author states that up to 500 messages per hour and typically 2000–3000 messages per day were sent between the London and the New York market through the cable companies. Each message was transmitted over the wire system in less than a minute. In a heuristic method that was repeated in *Dynamic Hedging* by one of the authors, (Taleb, 1997), Nelson, describe in a theory-free way many rigorously clinical aspects of his arbitrage business: the cost of shipping shares, the cost of insuring shares, interest expenses, the possibilities to switch shares directly between someone being long securities in New York and short in London and in this way saving shipping and insurance costs, as well as many more similar tricks.

The formal financial economics canon does not include historical sources from outside economics, a mechanism discussed in Taleb (2007). The put-call parity was according to the formal option literature first fully described by Stoll (1969), but neither he nor others in the field even mention Nelson. Not only was the put-call parity argument fully understood and described in detail by Nelson (1904), but he, in turn, makes frequent reference to Higgins (1902). Just as an example Nelson (1904) referring to Higgins (1902) writes:

> "It may be worthy of remark that 'calls' are more often dealt than 'puts' the reason probably being that the majority of 'punters' in stocks and shares are more inclined to look at the bright side of things, and therefore more often 'see' a rise than a fall in prices.

[8] See http://www.Nobel.se.

[9] See Bell et al. (2007)—we thank Barkley Rosser for ferreting out earlier uses.

[10] The historical description of the market is informative until Kairys and Valerio try to gauge whether options in the 1870s were underpriced or overpriced (using Black–Scholes–Merton style methods). There was one tail-event in this period, the great panic of September 1873. Kairys and Valerio find that holding puts was profitable, but deem that the market panic was just a one-time event: "However, the put contracts benefit from the "financial panic" that hit the market in September, 1873. Viewing this as a "one-time" event, we repeat the analysis for puts excluding any unexpired contracts written before the stock market panic." Using references to the economic literature that also conclude that options in general were overpriced in the 1950s 1960s and 1970s they conclude: "Our analysis shows that option contracts were generally overpriced and were unattractive for retail investors to purchase". They add: "Empirically we find that both put and call options were regularly overpriced relative to a theoretical valuation model." These results are contradicted by the practitioner Nelson (1904): ". . .the majority of the great option dealers who have found by experience that it is the givers, and not the takers, of option money who have gained the advantage in the long run".

[11] Here is a partial list: Bielschowsky, R (1892): Ueber die rechtliche Natur der Prämiengeschäfte, Bresl. Genoss.-Buchdr Granichstaedten-Czerva, R (1917): Die Prämiengeschäfte an der Wiener Börse, Frankfurt am Main Holz, L. (1905) Die Prämiengeschäfte, Thesis (doctoral)–Universität Rostock Kitzing, C. (1925): Prämiengeschäfte : Vorprämien-, Rückprämien-, Stellagen- u. Nochgeschäfte ; Die solidesten Spekulationsgeschäfte mit Versicherg auf Kursverlust, Berlin Leser, E, (1875): Zur Geschichte der Prämiengeschäfte Szkolny, I. (1883): Theorie und praxis der prämiengeschäfte nach einer originalen methode dargestellt., Frankfurt am Main Author Unknown (1925): Das Wesen der Prämiengeschäfte, Berlin : Eugen Bab & Co., Bankgeschäft.

> This special inclination to buy 'calls' and to leave the 'puts' severely alone does not, however, tend to make 'calls' dear and 'puts' cheap, for it can be shown that the adroit dealer in options can convert a 'put' into a 'call,' a 'call' into a 'put', a 'call o' more' into a 'put- and-call,' in fact any option into another, by dealing against it in the stock. We may therefore assume, with tolerable accuracy, that the 'call' of a stock at any moment costs the same as the 'put' of that stock, and half as much as the Put-and-Call."

The Put-and-Call was simply a put plus a call with the same strike and maturity, what we today would call a straddle. Nelson describes the put-call parity over many pages in full detail. Static market neutral delta hedging was also known at that time, in his book Nelson for example writes:

> "Sellers of options in London as a result of long experience, if they sell a Call, straightway buy half the stock against which the Call is sold; or if a Put is sold; they sell half the stock immediately."

We must interpret the value of this statement in the light that standard options in London at that time were issued at-the-money (as explicitly pointed out by Nelson); furthermore, all standard options in London were European style. In London in- or out-of-the-money options were only traded occasionally and were known as "fancy options". It is quite clear from this and the rest of Nelson's book that the option dealers were well aware that the delta for at-the-money options was approximately 50%. As a matter of fact, at-the-money options trading in London at that time were adjusted to be struck to be at-the-money forward, in order to make puts and calls of the same price. We know today that options that are at-the-money forward and do not have very long time to maturity have a delta very close to 50% (naturally minus 50% for puts). The options in London at that time typically had one month to maturity when issued.

Nelson also diffusely points to dynamic delta hedging, and that it worked better in theory than practice (see Haug, 2007). It is clear from all the details described by Nelson that options in the early 1900 traded actively and that option traders at that time in no way felt helpless in either pricing or in hedging them.

Herbert Filer was another option trader that was active from 1919 to the 1960s. Filer (1959) describes what must be considered a reasonable active option market in New York and Europe in the early 1920s and 1930s. Due to World War II, there was no trading on European Exchanges; the London markets did not resume until 1958. Since in the early 1900s, option traders in London were considered to be the most sophisticated, according to Nelson, it could well be that World War II and the subsequent shutdown of option trading for many years was the reason known robust arbitrage principles about options were forgotten and almost lost, to be partly re-discovered by finance professors such as Stoll (1969).

The put-call parity in the older literature seems to serve two main purposes:

1. As a pure arbitrage constrain,
2. but also as a tool to create calls out of puts, puts out of calls and straddles out of calls or puts for the purpose of hedging options with options. In other words more than simply arbitrage constraint, but a very important tool to transfer risk between options.

The original descriptions and uses of the put-call parity concept, unlike later theories, consider that supply and demand for options will affect option prices. Even if the Black, Scholes, Merton model in strict theoretical sense is fully consistent with the arbitrage constrain of the put-call parity, the model is actually not consistent with the original invention and use of the put-call parity.

In 1908, Vinzenz Bronzin published a book deriving several option pricing formulas, and a formula very similar to what today is known as the Black–Scholes–Merton formula, see Hafner and Zimmermann (2007, 2009). Bronzin based his *risk-neutral* option valuation on robust arbitrage principles such as the put-call parity and the link between the forward price and call and put options—in a way that was rediscovered by Derman and Taleb (2005).[12] Indeed, the put-call parity restriction is sufficient to remove the need to incorporate a future return in the underlying security—it forces the lining up of options to the forward price.[13]

Again, in 1910 Henry Deutsch describes put-call parity but in less detail than Higgins and Nelson. In 1961 Reinach again described the put-call parity in quite some detail. Traders at New York stock exchange specializing in using the put-call parity to convert puts into calls or calls into puts was at that time known as Converters, Reinach (1961):

[12] The argument of Derman and Taleb (2005) was present in Taleb (1997) but remained unnoticed.

[13] Ruffino and Treussard (2006) accept that one could have solved the risk-premium by happenstance, not realizing that put-call parity was so extensively used in history. But they find it insufficient. Indeed the argument may not be *sufficient* for someone who subsequently complicated the representation of the world with some implements of modern finance such as "stochastic discount rates"—while simplifying it at the same time to make it limited to the Gaussian and allowing dynamic hedging. They write that "the use of a non-stochastic discount rate common to both the call and the put options is inconsistent with modern equilibrium capital asset pricing theory." Given that we have never seen a practitioner use "stochastic discount rate", we, like our option trading predecessors, feel that put-call parity is sufficient & does the job. The situation is akin to that of scientists lecturing birds on how to fly, and taking credit for their subsequent performance—except that here it would be lecturing them the wrong way.

> "Although I have no figures to substantiate my claim, I estimate that over 60 per cent of all Calls are made possible by the existence of Converters."

In other words the *Converters* (dealers) who basically operated as market makers were able to operate and hedge most of their risk by "statically" hedging options with options. Reinach wrote that he was an option trader (Converter) and gave examples on how he and his colleagues tended to hedge and arbitrage options against options by taking advantage of options embedded in convertible bonds:

> "Writers and traders have figured out other procedures for making profits writing Puts & Calls. Most are too specialized for all but the seasoned professional. One such procedure is the ownership of a convertible bond and then writing of Calls against the stock into which the bonds are convertible. If the stock is called converted and the stock is delivered."

Higgins, Nelson and Reinach all describe the importance of put-call parity and *hedging options with options*. Option traders were in no way helpless in hedging or pricing before the Black–Scholes–Merton formula. As already mentioned static market-neutral delta hedging was described by Higgins and Nelson in 1902 and 1904. Also, Gann (1937) discusses market neutral delta hedging for at-the-money options, but in much less details than Nelson (1904). Gann also indicates some forms of auxiliary dynamic hedging.

Mills (1927) illustrates how jumps and fat tails were present in the literature in the pre-Modern Portfolio Theory days. He writes: "A distribution may depart widely from the Gaussian type because the influence of one or two extreme price changes."

4. Option formulas and delta hedging

Which brings us to option pricing formulas. The first identifiable one was Bachelier (1900). Sprenkle (1961) extended Bacheliers work to assume lognormal rather than normally distributed asset price. It also avoids discounting (to no significant effect since many markets, particularly the U.S., option premia were paid at expiration).

Boness (1964) also assumed a lognormal asset price. He derives a formula for the price of a call option that is actually identical to the Black–Scholes–Merton, 1973 formula. This is among several others also pointed out by Rubinstein (2006):

> "The real significance of the formula to the financial theory of investment lies not in itself, but rather in how it was derived. Ten years earlier the same formula had been derived by Case M. Sprenkle (1961) and A. James Boness (1964)."

Samuelson (1965) and Thorp (1969) published somewhat similar option pricing formulas to Boness and Sprenkle. Thorp (2007) claims that he actually had an identical formula to the Black–Scholes–Merton formula programmed into his computer years before Black, Scholes and Merton published their theory.

It is also worth to mention that McKean (1965) derives a formula for perpetual American put option, but without assuming continuous delta hedging. The formula was later modified by Merton (1973) to assume risk neutrality based on continuous dynamic hedging.

Now, delta hedging. As already mentioned static market-neutral delta hedging was clearly described by Higgins (1902) and Nelson (1904). Thorp and Kassouf (1967) presented market neutral static delta hedging in more details, not only for at-the-money options, but for options *with any delta*. In his 1969 paper Thorp is shortly describing market neutral static delta hedging, also briefly pointed in the direction of some dynamic delta hedging, not as a central pricing device, but a risk-management tool. Filer also points to dynamic hedging of options, but without showing much knowledge about how to calculate the delta. Another "ignored" and "forgotten" text is Bernhard (1970), a book/booklet published in 1970 by Arnold Bernhard & Co. The authors are clearly aware of market neutral static delta hedging or what they name "balanced hedge" for any level in the strike or asset price. This book has multiple examples of how to buy warrants or convertible bonds and construct a market neutral delta hedge by shorting the right amount of common shares. Arnold Bernhard & Co also published deltas for a large number of warrants and convertible bonds that they distributed to investors on Wall Street.

Referring to Thorp and Kassouf (1967), Black, Scholes and Merton took the idea of delta hedging one step further, Black and Scholes (1973):

> "If the hedge is maintained continuously, then the approximations mentioned above become exact, and the return on the hedged position is completely independent of the change in the value of the stock. In fact, the return on the hedged position becomes certain. This was pointed out to us by Robert Merton."

This may be a brilliant mathematical idea, but option pricing is not mathematical theory.

Just after Black, Scholes and Merton published their papers, Thorp (1973) showed how there could not be risk-neutrality once one moved away from continuous delta hedging. And given that continuous delta hedging was obviously impossible in practice, the paper showed how a similar option formula derived under discrete time delta hedging in the limit (of continuous hedging) was equivalent with the Black, Scholes, Merton model. However, his point was that the continuous time delta hedging of the formula was not correct since continuous hedging is impossible and such hedging is very non-robust – see Thorp (2002).

5. Myth 2: option traders today "use" the Black–Scholes–Merton formula

5.1. Traders do not do "valuation"

First, operationally, a price is not quite "valuation". Valuation requires a strong theoretical framework with its corresponding fragility to both assumptions and structure of a model. For traders, a "price" produced to buy an option when one has no knowledge of the probability distribution of the future is not "valuation", but an expedient. Such price could change. Their beliefs do not enter such price. It can also be determined by his inventory.

This distinction is critical: traders are engineers, whether boundedly rational (or even non-interested in any form of probabilistic rationality), they are not privy to informational transparency about the future states of the world and their probabilities. So they do not need a general theory to produce a price—merely the avoidance of Dutch-book style arbitrages against them, and the compatibility with some standard restriction: in addition to put-call parity, a call of a certain strike K cannot trade at a lower price than a call $K+\Delta K$ (avoidance of negative call and put spreads), a call struck at K and a call struck at $K+2\Delta K$ cannot be more expensive than twice the price of a call struck at $K+\Delta K$ (negative butterflies), horizontal calendar spreads cannot be negative (when interest rates are low), and so forth. The degrees of freedom for traders are thus reduced: they need to abide by put-call parity and compatibility with other options in the market.

In that sense, traders do not perform "valuation" with some "pricing kernel" until the expiration of the security, but, rather, produce a price of an option compatible with other instruments in the markets, with a holding time that is stochastic. They do not need top-down "science".

5.2. When do we value?

If you find traders operated solo, in a desert island, having for some to produce an option price and hold it to expiration, in a market in which the forward is absent, then some valuation would be necessary—but then their book would be minuscule. And this thought experiment is a distortion: people would not trade options unless they are in the business of trading options, in which case they would need to have a book with offsetting trades. For without offsetting trades, we doubt traders would be able to produce a position beyond a minimum (and negligible) size as dynamic hedging not possible. (Again we are not aware of many non-blownup option traders and institutions who have managed to operate in the vacuum of the Black–Scholes–Merton argument.) It is to the impossibility of such hedging that we turn next.

5.3. On the mathematical impossibility of dynamic hedging

Finally, we discuss the severe flaw in the dynamic hedging concept. It assumes, nay, requires all moments of the probability distribution to exist.[14]

Assume that the distribution of returns has a scale-free or fractal property that we can simplify as follows: for x large enough, (i.e. "in the tails"), $P[X>n\,x]/P[X>x]$ depends on n, not on x. In financial securities, say, where X is a daily return, there is no reason for $P[X>20\%]/P[X>10\%]$ to be different from $P[X>15\%]/P[X>7.5\%]$. This self-similarity at all scales generates power-law, or Paretian, tails, i.e., above a crossover point, $P[X>x]=Kx^{-\alpha}$. It happens, looking at millions of pieces of data, that such property holds in markets—all markets, baring sample error. For overwhelming empirical evidence, see Mandelbrot (1963), which predates Black–Scholes–Merton (1973) and the jump-diffusion of Merton (1976); see also Stanley et al. (2000), and Gabaix et al. (2003). The argument to assume the scale-free is as follows: the distribution might have thin tails at some point (say above some value of X). But we do not know where such point is—we are epistemologically in the dark as to where to put the boundary, which forces us to use infinity.

Some criticism of these "true fat-tails" accept that such property might apply for daily returns, but, owing to the Central Limit Theorem, the distribution is held to become Gaussian under aggregation for cases in which α is deemed higher than 2. Such argument does not hold owing to the preasymptotics of scalable distributions: Bouchaud and Potters (2003) and Mandelbrot and Taleb (2010) argue that the preasymptotics of fractal distributions are such that the effect of the Central Limit Theorem are exceedingly slow in the tails—in fact irrelevant. Furthermore, there is sampling error as we have less data for longer periods, hence fewer tail episodes, which give an in-sample illusion of thinner tails. In addition, the point that aggregation thins out the tails does not hold for dynamic hedging—in which the operator depends necessarily on high frequency data and their statistical properties. So long as it is scale-free at the time period of dynamic hedge, higher moments become explosive, "infinite" to disallow the formation of a dynamically hedge portfolio. Simply a Taylor expansion is impossible as moments of higher order that 2 matter critically—one of the moments is going to be infinite.

The mechanics of dynamic hedging are as follows: Assume the risk-free interest rate of 0 with no loss of generality. The canonical Black–Scholes–Merton package consists in selling a call and purchasing shares of stock that provide a hedge against instantaneous moves in the security. Thus the portfolio π locally "hedged" against exposure to the first moment of

[14] Merton (1992) seemed to accept the inapplicability of dynamic hedging but he perhaps thought that these ills would be cured thanks to his prediction of the financial world "spiraling towards dynamic completeness". Fifteen years later, we have, if anything, spiraled *away* from it.

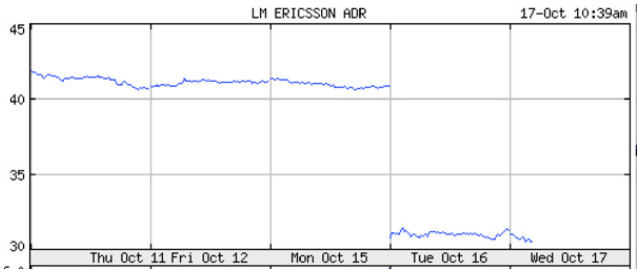


Fig. 2. A 25% gap in ericsson, one of the most liquid stocks in the world. Such move can dominate hundreds of weeks of dynamic hedging.

the distribution is the following:

$$\pi = -C + \frac{\partial C}{\partial S} S$$

where C is the call price, and S the underlying security.

Take the discrete time change in the values of the portfolio

$$\Delta\pi = -\Delta C + \frac{\partial C}{\partial S}\Delta S$$

By expanding around the initial values of S, we have the changes in the portfolio in discrete time. Conventional option theory applies to the Gaussian in which all orders higher than ΔS^2 and disappears rapidly.

$$\Delta\pi = -\frac{\partial C}{\partial S}\Delta t - \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\Delta S^2 + O(\Delta S^3)$$

Taking expectations on both sides, we can see here very strict requirements on moment finiteness: all moments need to converge. If we include another term, of order ΔS^3, such term may be of significance in a probability distribution with significant cubic or quartic terms. Indeed, although the nth derivative with respect to S can decline very sharply, for options that have a strike K away from the center of the distribution, it remains that the delivered higher orders of ΔS are rising disproportionately fast for that to carry a mitigating effect on the hedges.

So here we mean *all* moments—no approximation. The logic of the Black–Scholes–Merton so-called solution thanks to Ito's lemma was that the portfolio collapses into a deterministic payoff. But let us see how quickly or effectively this works in practice.

The actual replication process is as follows: the payoff of a call should be replicated with the following stream of dynamic hedges, the limit of which can be seen here, between t and T

$$\lim_{\Delta t \to 0} \left(\sum_{i=1}^{n=(T/\Delta t)} \frac{\partial C}{\partial S} \Big| S = S_{1+(t=1)\Delta t}, t = t + (i-1)\Delta t, \quad S_{t+i\Delta t} - S_{t+(i-1)\Delta t} \right)$$

Such policy does not match the call value: the difference remains stochastic (while according to Black Scholes it should shrink). Unless one lives in a fantasy world in which such risk reduction is possible.[15]

Further, there is an inconsistency in the works of Merton making us confused as to what theory finds acceptable: in Merton (1976) he agrees that we can use Bachelier-style option derivation in the presence of jumps and discontinuities – no dynamic hedging – but only when the underlying stock price is uncorrelated to the market. This seems to be an admission that dynamic hedging argument applies only to *some* securities: those that do not jump and are correlated to the market (Fig. 2).

5.4. The robustness of the Gaussian

The success of the "formula" last developed by Thorp, and called "Black–Scholes–Merton" was due to a simple attribute of the Gaussian: you can express *any* probability distribution in terms of Gaussian, even if it has fat tails, by varying the standard deviation σ at the level of the density of the random variable. It does not mean that you are using a Gaussian, nor does it mean that the Gaussian is particularly parsimonious (since you have to attach a σ for every level of the price). It simply mean that the Gaussian can express anything you want if you add a function for the parameter σ, making it function of strike price and time to expiration.

[15] We often hear the misplaced comparison to Newtonian mechanics. It supposedly provided a good approximation until we had relativity. The problem with the comparison is that the thin-tailed distributions are *not approximations* for fat-tailed ones: there is a deep qualitative difference.

This "volatility smile", i.e., varying one parameter to produce $\sigma(K)$, or "volatility surface", varying two parameter, $\sigma(S,t)$ is effectively what was done in different ways by Dupire (1994, 2005) and Derman and Kani (1994, 1998), see Gatheral (2006). They assume a volatility process not because there is necessarily such a thing—only as a method of fitting option prices to a Gaussian. Furthermore, although the Gaussian has finite second moment (and finite all higher moments as well), you can express a scalable with infinite variance using Gaussian "volatility surface". One strong constrain on the σ parameter is that it must be the same for a put and call with same strike (if both are European-style), and the drift should be that of the forward.[16]

Indeed, ironically, the volatility smile is inconsistent with the Black–Scholes–Merton theory. This has lead to hundreds if not thousands of papers trying extend (what was perceived to be) the Black–Scholes–Merton model to incorporate stochastic volatility and jump-diffusion. Several of these researchers have been surprised that so few traders actually use stochastic volatility models. It is not a model that says how the volatility smile should look like, or evolves over time; it is a hedging method that is robust and consistent with an arbitrage free volatility surface that evolves over time.

In other words, you can use a volatility surface as a map, not a territory. However, it is foolish to justify Black–Scholes–Merton on grounds of its use: we repeat that the Gaussian bans the use of probability distributions that are not Gaussian–whereas non-dynamic hedging derivations (Bachelier, Thorp) are not grounded in the Gaussian.

5.5. Order flow and options

It is clear that option traders are not necessarily interested in probability distribution at expiration time—given that this is abstract, even metaphysical for them. In addition to the put-call parity constrains, we can hedge away inventory risk in options with other options. One very important implication of this method is that if you hedge options with options then option pricing will be largely demand and supply based.[17] This in strong contrast to the Black–Scholes–Merton (1973) theory in which demand and supply for options simply should not affect the price of options. If someone wants to buy more options the market makers can simply manufacture them by dynamic delta hedging that will be a perfect substitute for the option itself.

This raises a critical point: option traders do not "estimate" the odds of rare events by pricing out-of-the-money options. They just respond to supply and demand. The notion of "implied probability distribution" is merely a Dutch-book compatibility type of proposition.

5.6. Bachelier–Thorp

The argument often casually propounded attributing the success of option volume to the quality of the Black Scholes formula is rather weak. It is particularly weakened by the fact that options had been so successful at different time periods and places.

Furthermore, there is evidence that while both the Chicago Board Options Exchange and the Black–Scholes–Merton formula came about in 1973, the model was "rarely used by traders" before the 1980s (O'Connell, 2001). When one of the authors (Taleb) became a pit trader in 1992, almost two decades after Black–Scholes–Merton, he was surprised to find that many traders still priced options heuristically "sheets free", "pricing off the butterfly", and "off the conversion", without recourse to any formula.

Even a book written in 1975 by a finance academic appears to credit Thorp and Kassouf (1967)—rather than Black and Scholes (1973), although the latter was present in its bibliography, Auster (1975).

Sidney Fried wrote on warrant hedges before 1950, but it was not until 1967 that the book Beat the Market by Edward O. Thorp and Sheen T. Kassouf rigorously, but simply, explained the "short warrant/long common" hedge to a wide audience.

We conclude with the following remark. Sadly, all the equations, from the first (Bachelier), to the last pre-Black–Scholes–Merton (Thorp) accommodate a scale-free distribution. The notion of explicitly removing the expectation from the forward was present in Keynes (1924) and later by Blau (1944)—and long a Call short a put of the same strike equals a forward. These simple and effective arbitrage relationships appeared to be well known heuristics in 1904.

One could easily attribute the explosion in option volume to the computer age and the ease of processing transactions, added to the long stretch of peaceful economic growth and absence of hyperinflation. From the evidence (once one removes the propaganda), the development of scholastic finance appears to be an epiphenomenon rather than a cause of option trading. Once again, lecturing birds how to fly does not allow one to take subsequent credit.

This is why we call the equation Bachelier–Thorp. We have been using it all along and gave it the wrong name, after the wrong method and with attribution to the wrong persons. It does not mean that dynamic hedging is out of the question; it is just not a central part of the pricing paradigm. It led to the writing down of a certain stochastic process that may have its uses, some day, should markets "spiral towards dynamic completeness". But not in the present.

[16] See Breeden and Litzenberger (1978), Gatheral (2006). See also Bouchaud and Potters (2001) for hedging errors in the real world.

[17] See <fn0085>Gârleanu et al. (2009).

Acknowledgements

We thank Russ Arbuthnot, John (Barkley) Rosser, and others for useful comments.

References

Auster, R., 1975. Option Writing and Hedging Strategies. Exposition Press, New York.
Bachelier, L., 1900. Theory of speculation. In: Cootner, P. (Ed.), 1964. The Random Character of Stock Market Prices. MIT Press, Cambridge.
Bell, A.R., Brooks, C., Dryburgh, P.R., 2007. The English Wool Market c. 1230–1323. Cambridge University Press, Cambridge.
Bernhard, A., 1970. More Profit and Less Risk: Convertible Securities and Warrants. Written and Edited by the Publisher and Editors of The Value Line Convertible Survey. Arnold Bernhard & Co., Inc.
Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–654.
Blau, G., 1944. Some aspects of the theory of futures trading. The Review of Economic Studies 12, 1–30.
Boness, A., 1964. Elements of a theory of stock-option value. Journal of Political Economy 72, 163–175.
Bouchaud, J.-P., Potters, M., 2003. Theory of Financial Risks and Derivatives Pricing, From Statistical Physics to Risk Management, 2nd ed. Cambridge University Press.
Bouchaud, J.-P., Potters, M., 2001. Welcome to a non-Black-Scholes world. Quantitative Finance 1, 482–483.
Breeden, D., Litzenberger, R., 1978. Prices of state-contingent claims implicit in option prices. Journal of Business 51, 621–651.
Bronzin, V., 1908. Theorie der Prämiengeschäfte. Verlag Franz Deticke, Leipzig und Wien.
De La Vega, J., 1688. Confusión de Confusiones. Re-printed In: Fridson, M., S., (Eds.), Extraordinary Popular Delusions and the Madness of Crowds & Confusión de Confusiones, 1996. New York: Wiley Publishing.
De Pinto, I., 1771. An Essay on Circulation of Currency and Credit in Four Parts and a Letter on the Jealousy of Commerce, translated with annotations by S. Baggs 1774. Reprinted by Gregg International Publishers 1969, London.
Derman, E., Kani, I., 1994. Riding on a smile. Risk 7, 32–39.
Derman, E., Kani, I., 1998. Stochastic implied trees: arbitrage pricing with stochastic term and strike structure of volatility. International Journal of Theoretical and Applied Finance 1, 61–110.
Derman, E., Taleb, N., 2005. The illusion of dynamic delta replication. Quantitative Finance 5, 323–326.
Dupire, B., 1994. Pricing with a smile. Risk 7, 18–20.
Dupire, B., 2005. Volatility derivatives modeling presentation. NYU. a copy of the presentation exist online: Available from: http://www.math.nyu.edu/carrp/mfseminar/bruno.ppt.
Filer, H., 1959. Understanding Put and Call Options. Popular Library, New York.
Gabaix, X., Gopikrishnan, P., Plerou, V., Stanley, H.E., 2003. A theory of power-law distributions in financial market fluctuations. Nature 423, 267–270.
Gann, W.D., 1937. How to Make Profits in Puts and Calls. Lambert Gann Publishing Co., WA.
Gârleanu, N., Pedersen, L.H., Poteshman, A.M., 2009. Demand-based option pricing. Review of Financial Studies 22, 4259–4299.
Gatheral, J., 2006. The Volatility Surface. John Wiley & Sons, New York.
Gelderblom, O., Jonker, J., 2005. Amsterdam as the cradle of modern futures and options trading. In: Goetzmann, N., Rouwenhorst, K.G. (Eds.), The Origins of Value: The Financial Innovations that Created Modern Capital Markets. Oxford University Press, USA, pp. 1550–1650.
Gigerenzer, G., Todd, P.M., The ABC Research Group, 2000. Simple Heuristics That Make Us Smart. Oxford University Press, Oxford.
Haug, E.G., 2007. Derivatives Models on Models. John Wiley & Sons, New York.
Hafner, W., Zimmermann, H., 2007. Amazing discovery: Vincenz Bronzin's option pricing models. Journal of Banking and Finance 31, 531–546.
Hafner, W., Zimmermann, H., 2009. Vinzenz Bronzin's Option Pricing Models. Exposition and Appraisal. Springer Verlag.
Higgins, L.R., 1902. The Put-and-Call. E. Wilson, London.
Kairys, J.P., Valerio, N., 1997. The market for equity options in the 1870s. Journal of Finance 52, 1707–1723.
Keynes, J.M., 1924. A Tract on Monetary Reform. Re-printed 2000. Prometheus Books, Amherst New York.
Mandelbrot, B., 1963. The variation of certain speculative prices. The Journal of Business 36, 394–419.
Mandelbrot, B., Taleb, N., 2010. Mild vs. wild randomness: focusing on risks that matter. In: Diebold, F., Doherty, N., Herring, R. (Eds.), The Known, the Unknown and the Unknowable in Financial Institutions. Princeton University Press, Princeton, N.J.
McKean, H.P., 1965. A free boundary problem for the heat equation arising from a problem in mathematical economics. Industrial Management Review 6 (2), 32–39.
Merton, R.C., 1973. Theory of rational option pricing. Bell Journal of Economics and Management Science 4, 141–183.
Merton, R.C., 1976. Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics 3, 125–144.
Merton, R.C., 1992. Continuous-Time Finance, Revised ed. Blackwell.
Mills, F., 1927. The Behaviour of Prices. National Bureau of Economic Research, New York.
Mixon, S., 2009a. Option markets and implied volatility: past versus present. Journal of Financial Economics 94, 171–191.
Mixon, S., 2009. The foreign exchange option market, 1917–1921. Working Paper, SSRN.
Moore, L., Juh, S., 2006. Derivative Pricing 60 Years before Black–Scholes: Evidence from the Johannesburg Stock Exchange. The Journal of Finance LXI (6), 3069–3098.
Nelson, S.A., 1904. The A B C of Options and Arbitrage. The Wall Street Library, New York.
O'Connell, M.P., 2001. The Business of Options. John Wiley & Sons, New York.
Poitras, G., 2009. The early history of option contracts. In: Hafner, W., Zimmermann, H. (Eds.), Vinzenz Bronzin's Option Pricing Models. Exposition and Appraisal. Springer Verlag.
Reinach, A.M., 1961. The Nature of Puts & Calls. The Book-mailer, New York.
Ross, S., 2005. Neoclassical Finance. Princeton University Press, Princeton.
Rubinstein, M., 1998. Derivatives. Available from: http://www.in-the-money.com.
Rubinstein, M., 2006. A History of The Theory of Investments. John Wiley & Sons, New York.
Ruffino, D., Treussard, J., 2006. Derman and Taleb's 'The ilusion of dynamic replication': a comment. Quantitative Finance 6, 365–367.
Samuelson, P., 1965. Rational theory of warrant pricing. Industrial Management Review 6, 13–31.
Sprenkle, C., 1961. Warrant prices as indicators of expectations and preferences. Yale Economics Essays 1, 178–231.
Stanley, H.E., Amaral, L.A.N., Gopikrishnan, P., Plerou, V., 2000. Scale invariance and universality of economic fluctuations. Physica A 283, 31–41.
Stoll, H., 1969. The relationship between put and call prices. Journal of Finance 24, 801–824.
Taleb, N., 1997. Dynamic Hedging. John Wiley & Sons, New York.
Taleb, N., 2007. The Black Swan. Random House, New York.
Thorp, E.O., 1973. A corrected derivation of the Black–Scholes option model. In: Presented at the CRSP proceedings in 1976.
Thorp, E.O., 1969. Optimal gambling systems for favorable games. Review of the International Statistics Institute 37, 273–293.
Thorp, E.O., Kassouf, S.T., 1967. Beat the Market. Random House, New York.
Thorp, E.O., 2002. What I knew and when I knew it–Part 1, Part 2, Part 3. Wilmott Magazine, Sep-02, Dec-02, Jan-03.
Thorp, E.O., 2007. Edward Thorp on gambling and trading. In: Haug, E.G. (Ed.), 2007. Derivatives Models on Models. John Wiley & Sons, New York.
Weinstein, M.H., 1931. Arbitrage to Securities. Harper Brothers, New York.

Getty Images

THE SIX MISTAKES EXECUTIVES MAKE IN RISK MANAGEMENT

Black Swan events are almost impossible to predict. Instead of perpetuating the illusion that we can anticipate the future, risk management should try to reduce the impact of the threats we don't understand. | **by Nassim N. Taleb, Daniel G. Goldstein, and Mark W. Spitznagel**

WE DON'T LIVE in the world for which conventional risk-management textbooks prepare us. No forecasting model predicted the impact of the current economic crisis, and its consequences continue to take establishment economists and business academics by surprise. Moreover, as we all know, the crisis has been compounded by the banks' so-called risk-management models, which increased their exposure to risk instead of limiting it and rendered the global economic system more fragile than ever.

Low-probability, high-impact events that are almost impossible to forecast –

we call them Black Swan events – are increasingly dominating the environment. Because of the internet and globalization, the world has become a complex system, made up of a tangled web of relationships and other interdependent factors. Complexity not only increases the incidence of Black Swan events but also makes forecasting even ordinary events impossible. All we can predict is that companies that ignore Black Swan events will go under.

Instead of trying to anticipate low-probability, high-impact events, we should reduce our vulnerability to them. Risk management, we believe, should be about lessening the impact of what we don't understand – not a futile attempt to develop sophisticated techniques and stories that perpetuate our illusions of being able to understand and predict the social and economic environment.

To change the way we think about risk, we must avoid making six mistakes.

We think we can manage risk by predicting extreme events.

This is the worst error we make, for a couple of reasons. One, we have an abysmal record of predicting Black Swan events. Two, by focusing our attention on a few extreme scenarios, we neglect other possibilities. In the process, we become more vulnerable.

It's more effective to focus on the consequences – that is, to evaluate the possible impact of extreme events. Realizing this, energy companies have finally shifted from predicting when accidents in nuclear plants might happen to preparing for the eventualities. In the same way, try to gauge how your company will be affected, compared with competitors, by dramatic changes in the environment. Will a small but unexpected fall in demand or supply affect your company a great deal? If so, it won't be able to withstand sharp drops in orders, sudden rises in inventory, and so on.

In our private lives, we sometimes act in ways that allow us to absorb the impact of Black Swan events. We don't try to calculate the odds that events will occur; we only worry about whether we can handle the consequences if they do. In addition, we readily buy insurance for health care, cars, houses, and so on. Does anyone buy a house and then check the cost of insuring it? You make your decision after taking into account the insurance costs. Yet in business we treat insurance as though it's an option. It isn't; companies must be prepared to tackle consequences and buy insurance to hedge their risks.

We are convinced that studying the past will help us manage risk.

Risk managers mistakenly use hindsight as foresight. Alas, our research shows that past events don't bear any relation to future shocks. World War I, the attacks of September 11, 2001 – major events like those didn't have predecessors. The same is true of price changes. Until the late 1980s, the worst decline in stock prices in a single day had been around 10%. Yet prices tumbled by 23% on October 19, 1987. Why then would anyone have expected a meltdown after that to be only as little as 23%? History fools many.

You often hear risk managers – particularly those employed in the financial services industry – use the excuse "This is unprecedented." They assume that if they try hard enough, they can find precedents for anything and predict everything. But Black Swan events don't have precedents. In addition, today's world doesn't resemble the past; both interdependencies and nonlinearities have increased. Some policies have no effect for much of the time and then cause a large reaction.

People don't take into account the types of randomness inherent in many economic variables. There are two kinds, with socioeconomic randomness being less structured and tractable than the randomness you encounter in statistics textbooks and casinos. It causes winner-take-all effects that have severe consequences. Less than 0.25% of all the companies listed in the world represent around half the market capitalization, less than 0.2% of books account for approximately half their sales, less than 0.1% of drugs generate a little more than half the pharmaceutical industry's sales – and less than 0.1% of risky events will cause at least half your losses.

Because of socioeconomic randomness, there's no such thing as a "typical" failure or a "typical" success. There are typical heights and weights, but there's no such thing as a typical victory or catastrophe. We

Because of socioeconomic randomness, there's no such thing as a **"typical" failure or a "typical" success.**

have to predict both an event and its magnitude, which is tough because impacts aren't typical in complex systems. For instance, when we studied the pharmaceuticals industry, we found that most sales forecasts don't correlate with new drug sales. Even when companies had predicted success, they underestimated drugs' sales by 22 times! Predicting major changes is almost impossible.

3 We don't listen to advice about what we shouldn't do.

Recommendations of the "don't" kind are usually more robust than "dos." For instance, telling someone not to smoke outweighs any other health-related advice you can provide. "The harmful effects of smoking are roughly equivalent to the combined good ones of every medical intervention developed since World War II. Getting rid of smoking provides more benefit than being able to cure people of every possible type of cancer," points out genetics researcher Druin Burch in *Taking the Medicine.* In the same vein, had banks in the U.S. heeded the advice not to accumulate large exposures to low-probability, high-impact events, they wouldn't be nearly insolvent today, although they would have made lower profits in the past.

Psychologists distinguish between acts of commission and those of omission. Although their impact is the same in economic terms – a dollar not lost is a dollar earned – risk managers don't treat them equally. They place a greater emphasis on earning profits than they do on avoiding losses. However, a company can be successful by preventing losses while its rivals go bust – and it can then take market share from them. In chess, grand masters focus on avoiding errors; rookies try to win. Similarly, risk managers don't like not to invest and thereby conserve value. But consider where you would be today if your investment portfolio had remained intact over the past two years, when everyone else's fell by 40%. Not losing almost half your retirement is undoubtedly a victory.

No one should have a piece of the upside **without a share of the downside.**

Positive advice is the province of the charlatan. The business sections in bookstores are full of success stories; there are far fewer tomes about failure. Such disparagement of negative advice makes companies treat risk management as distinct from profit making and as an afterthought. Instead, corporations should integrate risk-management activities into profit centers and treat them as profit-generating activities, particularly if the companies are susceptible to Black Swan events.

4 We assume that risk can be measured by standard deviation.

Standard deviation – used extensively in finance as a measure of investment risk – shouldn't be used in risk management. The standard deviation corresponds to the square root of average *squared* variations – not average variations. The use of squares and square roots makes the measure complicated. It only means that, in a world of tame randomness, around two-thirds of changes should fall within certain limits (the –1 and +1 standard deviations) and that variations in excess of seven standard deviations are practically impossible. However, this is inapplicable in real life, where movements can exceed 10, 20, or sometimes even 30 standard deviations. Risk managers should avoid using methods and measures connected to standard deviation, such as regression models, R-squares, and betas.

Standard deviation is poorly understood. Even quantitative analysts don't seem to get their heads around the concept. In experiments we conducted in 2007, we gave a group of quants information about the average absolute movement of a stock (the mean absolute deviation), and they promptly confused it with the standard deviation when asked to perform some computations. When experts are confused, it's unlikely that other people will get it right. In any case, anyone looking for a single number to represent risk is inviting disaster.

5 We don't appreciate that what's mathematically equivalent isn't psychologically so.

In 1965, physicist Richard Feynman wrote in *The Character of Physical Law* that two mathematically equivalent formulations can be unequal in the sense that they present themselves to the human mind in different ways. Similarly, our research shows that the way a risk is framed influences peo-

ple's understanding of it. If you tell investors that, on average, they will lose all their money only every 30 years, they are more likely to invest than if you tell them they have a 3.3% chance of losing a certain amount each year.

The same is true of airplane rides. We asked participants in an experiment: "You are on vacation in a foreign country and are considering flying a local airline to see a special island. Safety statistics show that, on average, there has been one crash every 1,000 years on this airline. It is unlikely you'll visit this part of the world again. Would you take the flight?" All the respondents said they would.

We then changed the second sentence so it read: "Safety statistics show that, on average, one in 1,000 flights on this airline has crashed." Only 70% of the sample said they would take the flight. In both cases, the chance of a crash is 1 in 1,000; the latter formulation simply sounds more risky.

Providing a best-case scenario usually increases the appetite for risk. Always look for the different ways in which risk can be presented to ensure that you aren't being taken in by the framing or the math.

We are taught that efficiency and maximizing shareholder value don't tolerate redundancy.

Most executives don't realize that optimization makes companies vulnerable to changes in the environment. Biological systems cope with change; Mother Nature is the best risk manager of all. That's partly because she loves redundancy. Evolution has given us spare parts – we have two lungs and two kidneys, for instance – that allow us to survive.

In companies, redundancy consists of apparent inefficiency: idle capacities, unused parts, and money that isn't put to work. The opposite is leverage, which we are taught is good. It isn't; debt makes companies – and the economic system – fragile. If you are highly leveraged, you could go under if your company misses a sales forecast, interest rates change, or other risks crop up. If you aren't carrying debt on your books, you can cope better with changes.

Overspecialization hampers companies' evolution. David Ricardo's theory of comparative advantage recommended that for optimal efficiency, one country should specialize in making wine, another in manufacturing clothes, and so on. Arguments like this ignore unexpected changes. What will happen if the price of wine collapses? In the 1800s many cultures in Arizona and New Mexico vanished because they depended on a few crops that couldn't survive changes in the environment.

• • •

One of the myths about capitalism is that it is about incentives. It is also about disincentives. No one should have a piece of the upside without a share of the downside. However, the very nature of compensation adds to risk. If you give someone a bonus without clawback provisions, he or she will have an incentive to hide risk by engaging in transactions that have a high probability of generating small profits and a small probability of blowups. Executives can thus collect bonuses for several years. If blowups eventually take place, the managers may have to apologize but won't have to return past bonuses. This applies to corporations, too. That's why many CEOs become rich while shareholders stay poor. Society and shareholders should have the legal power to get back the bonuses of those who fail us. That would make the world a better place.

Moreover, we shouldn't offer bonuses to those who manage risky establishments such as nuclear plants and banks. The chances are that they will cut corners in order to maximize profits. Society gives its greatest risk-management task to the military, but soldiers don't get bonuses.

Remember that the biggest risk lies within us: We overestimate our abilities and underestimate what can go wrong. The ancients considered hubris the greatest defect, and the gods punished it mercilessly. Look at the number of heroes who faced fatal retribution for their hubris: Achilles and Agamemnon died as a price of their arrogance; Xerxes failed because of his conceit when he attacked Greece; and many generals throughout history have died for not recognizing their limits. Any corporation that doesn't recognize its Achilles' heel is fated to die because of it.

Nassim N. Taleb *is the Distinguished Professor of Risk Engineering at New York University's Polytechnic Institute and a principal of Universa Investments, a firm in Santa Monica, California. He is the author of several books, including* The Black Swan: The Impact of the Highly Improbable *(Random House, 2007).* **Daniel G. Goldstein** *is an assistant professor of marketing at London Business School and a principal research scientist at Yahoo.* **Mark W. Spitznagel** *is a principal of Universa Investments.*

Reprint R0910G To order, see page 143.

The Journal of Behavioral Finance
2004, Vol. 5, No. 1, 2–7

PRACTITIONER'S PERSPECTIVE

Bleed or Blowup? Why Do We Prefer Asymmetric Payoffs ?

In some strategies and life situations, it is said, one gambles dollars to win a succession of pennies. In others one risks a succession of pennies to win dollars. While one would think that the second category would be more appealing to investors and economic agents, we have an overwhelming evidence of the popularity of the first. A popular illustration of such asymmetry in returns is evident in the story of the Long Term Capital Management hedge fund. The fund derived steady returns over a dozen quarters then lost all of them in addition to almost all its capital in a single observation (see Lowenstein, 2000)–only for the main principals to restart a new, albeit milder, version of the strategy. Is there a systematic bias in favor of such return profiles?[1]

Negative (or "left") skewness can be presented by considering a stream of gambles that differ from most symmetric lotteries generally presented in the literature (where the agent usually has a 50% probability of realizing a given gain, *G*, and a 50% probability of realizing a loss, *L*). The asymmetric case we consider has, for a given expectation, both probabilities markedly diverging from 50%. A considerably negatively skewed bet can present more than 99% probability of making *G* and less 1% probability of losing *L. W*hile such skewness may sound extreme we will see that there is an abundance of instruments in the financial markets that actually deliver such payoffs (one may even say that *almost all* derivative products offer asymmetric properties). More technically, the mathematical representation of negative skewness defines it as a negative third central moment, the product of the probabilities by the cube of the payoffs deducted from the mean.

Would an economic agent facing a stream of stochastic monetary payoffs prefer negative skewness? Given a profile of monetary gambles, would he prefer to "bleed" (i.e. undergo small but frequent losses) or "blowup,"[2] i.e., take severe hits concentrated in small periods of time? Statistical properties of popular classes of investments, earnings management on the part of corporations (where corporations manage their earnings to moderately beat estimates most of the time and take hits on occasion [3]), and mechanisms like covered call writing (where investors clip their upside gains for a small fee) shows a strong evidence for the predilection for negative skewness on the part of investors. Indeed such preference is mostly expressed in the growth of classes of financial securities like hedge funds that, according to the empirical literature (see Fung and Hsieh, 1997, Kat, 2002), seem to be severely plagued by such properties, even possibly designed to cater to the investors' biases.

We divide biases into two categories, namely, a) cognitive, as agents may not understand the true implications of skewness, or why the expectation of a payoff is not necessarily better even if it generates steady returns; b) behavioral, as they may prefer a set of payoffs to another. The aim of this short discussion is to make the connect this preference for skewness with research that has been done in the behavioral literature–and describe further experiments would be needed to confirm it. It is organized as follows. Skewness is discussed, along with its prevalence in the growing new investments classes. We examine three major angles in the behavioral research: a) the belief in the "law of small numbers" and aspects of the overconfidence literature, b) prospect theory, and c) the promising field of hedonic psychology, an offshoot of prospect theory. The aim if this note is not the display of the evidence but hints and direction for confirmatory research.

Skewed Payoffs and Financial Instruments

How are these payoffs constructed? Instruments abound. Consider the following examples, which we divide into direct (i.e. strategy analyzed on their own merit) and comparative (strategies analyzed in comparison with a benchmark).

Examples of Directly Negatively Skewed Bets

Loans and credit-related instruments. You lend to an entity at a rate higher than the risk-free one prevailing in the economy. You have a high probability to earn the entire interest amount, except, of course in the event of default where you may lose (depending on the recovery rate) approximately half of your investment. The lower the risk of default, the more asymmetric the payoff. The same applies to investments in high yielding currencies that are pegged to a more stable one (say the

Argentine peso to the dollar) but occasionally experience a sharp devaluation.

Derivative instruments. A trader sells a contingent claim. If the option is out-of-the-money the payoff stream for such strategy is frequent profits, infrequent large losses, in proportion to how far out-of-the-money the option is. It is easy to see in the volumes that most traded options are out-of-the-money.[4] Note that a "market neutral" or "hedged" (i.e., made insensitive to the direction of the market)[5] such strategy does not significantly mitigate such asymmetry, since the mitigation of such risk of large losses implies continuous adjustment of the position, a strategy that fails with discontinuous jumps in the price of the underlying security. A seller of an out-of-the money option can make a profit as frequently as he wishes, possibly 99% of the time by, say selling on a monthly basis options estimated by the market to expire worthless 99% of the time.

Arbitrage. There are classes of arbitrage operations such as: 1) "merger arbitrage" in which the operator engages in betting that the merger will take place at a given probability and loses if the merger is cancelled (the opposite is called a "Chinese"). These trades generally have the long odds against the merger. 2) "Convergence trading" where a high yielding security is owned and an equivalent one is shorted thinking that they converge to each others, which tends to happen except in rare circumstances. The hedge fund boom has resulted in a proliferation of packaged instruments of some opacity that engage in a variety of the above strategies–ones that do let themselves be revealed through naive statistical observation.

Example of Comparatively Skewed Bets

Covered calls writing. Investors have long engaged in the "covered write" strategies by selling an option against their portfolio, thereby increasing the probability of a profit in return for a reduction of the upside potential. There is an abundant empirical literature on covered writes (see Board, Sutcliffe and Patrinos, 2000, for a review, and Whaley, 2002 for a recent utility-based explanation) where investors find gains in utility from capping payoffs as the marginal utility of gains decreases at a higher asset price. Indeed the fact that individual investors sell options at cheaper than their actuarial value can only be explained by the utility effect. For a mutual fund manager, doing such "covered writing" against his portfolio increases the probability of beating the index in the short run, but subjects him to long term underperformance as he will give back such outperformance during large rallies.

Properties of a Left-skewed Payoff

In brief, a negatively skewed stream offer the following attributes:

Property 1: Camouflage of the Mean and Variance

The true mean of the payoff is different from the median, in proportion to the skewness of the bet. A typical return will, say, be higher than the expected return. It is consequently easier for the observer of the process to be fooled by the true mean particularly if he observes the returns without a clear idea about the nature of the underlying (probability) generator. But things are worse for the variance as most of the time it we be lower than the true one (intuitively if a shock happens 1% of the time then the observed variance over a time window will decrease between realizations then sharply jump after the shock).

Property 2: Sufficiency of Sample Size

It takes a considerably longer sample to observe the properties under a skewed process than otherwise. For example, consider a bet with 99% probability of making G and 1% probability of losing L. In this example, the properties will not reveal themselves 99% of the time–and when they do, it is always a little late as the decision has already been made. Contrast that with a symmetric bet where the properties converge rather rapidly at the square root of the number of observations.

Property 3: The Smooth Ride Effect.

As mentioned above, the observed variance of the process is generally lower than the true variance most of the time. This means, simply, that the more skewness, the more the process will generate steady returns with smooth ride attributes, concentrating the variance in a brief period, the brevity of which is proportional to the variance. In another word, an investor has, without a decrease in risk, a more comfortable ride most of the time, with an occasional crash.

Overconfidence and Belief in the Law of Small Numbers

The first hint of an explanation for the neglect of the small risks of large losses comes from the early literature on behavior under uncertainty. Tversky and Kahneman (1971) wrote "We submit that people view a sample randomly drawn from a population as highly representative, that is, similar to a population in all essential characteristics." The consequence is the induc-

tive fallacy: overconfidence in the ability to infer general properties from observed facts, "undue confidence in early trends" and the stability of observed patterns and deriving conclusions with more confidence attached to them than can be warranted by the data. Worst, the agent finds causal explanations or perhaps distributional attributes that confirm his undue generalization.[6]

It is easy to see that the "small numbers" gets exacerbated with skewness since the observed mean will usually be different from the true mean and the observed variance will usually be lower than the true one. Now consider that it is a fact that in life, unlike a laboratory or a casino, we do not observe the probability distribution from which random variables are drawn. We only see the realizations of these random processes. It would be nice if we could, but it remains that we do not measure probabilities as we would measure the temperature or the height of a person. This means that when we compute probabilities from past data we are making assumptions about the skewness of the generator of the random series–all data is conditional upon a generator. In short, with skewed packages, Property 1 comes into play *and* we tend to believe what we see.

The literature on "small numbers" implies that agents have a compressed, narrower distribution in their minds than warranted from the data. The literature on overconfidence studies the bias from another angle by examining the wedge between the perception of unlikely events and their actual occurrence as well as the failure to calibrate from past errors. Since Alpert and Raiffa (1982) studies have documented how agents underestimate the extreme values of a distribution in a surprising manner; violations are far more excessive than one would expect: events that are estimated to occur less than 2% of the time will take place up to 49%. There has been since a long literature on overconfidence (in the sense of agents discounting the probability of adverse events while engaging in a variety of projects), see Kahneman and Lovallo (1993), Hilton (2003).

"Every Day is a New Day": The Implications of Prospect Theory

Prospect theory derives its name from the way agents face prospects or lotteries (Kahneman and Tversky, 1979). Its central idea is that economic agents reset their "utility" function to ignore, to some extent, accumulated performance and focus on the changes in wealth in their decision making under uncertainty. One may accumulate large quantities of wealth, but habituation makes him reset to the old Wall Street adage "every day is a new trading day," which means that he will look at gains and losses from the particular strategy, not the absolute levels of wealth and make decisions accordingly. The reference point is the individual's point of comparison, the "status quo" against which alternative scenarios are compared. Moreover prospect theory differs from "utility theory" *per se* in the separation of decision probability from the "value function." Decision probability, or weighting function, has the property of exaggerating small probabilities and underestimating large ones.

It is noteworthy that prospect theory was empirically derived from one-shot experiments with agents subjected to questions in which the odds were supplied.[7] It is the value function of the prospect theory that we examine next, rather than probabilities used in the decision-making. The normative neoclassical utility theory stipulates an increased sensitivity to losses and a decreased one to gains (investors would prefer negative skewness only for their increase but not decrease, in wealth). On the other hand, the value function of prospect theory documents a decreased sensitivity to both gains and losses, hence a marked overall preference for negative skewness. At the core, the difference is simply related to the fact that operators are more concerned with the utility of changes in wealth rather than those of the accumulated wealth itself, creating a preference for a given path dependence in the sequences of payoffs.

To see how the empirically derived version of utility theory presents asymmetric higher order properties, consider the following proposed representation of Tversky and Kahneman (1992). One has a value function, $V+(x)$, for x positive or 0, and $V-(x)$ for x strictly negative.

$$V+(x)= x^{\alpha}$$

$$V-(x)= (-\lambda)(-x)^{\alpha}$$

From this it is easy to see that, with $\alpha<1$, that V is concave in the profit domain and convex to in the loss domain. Looking at the second derivatives, one observes:

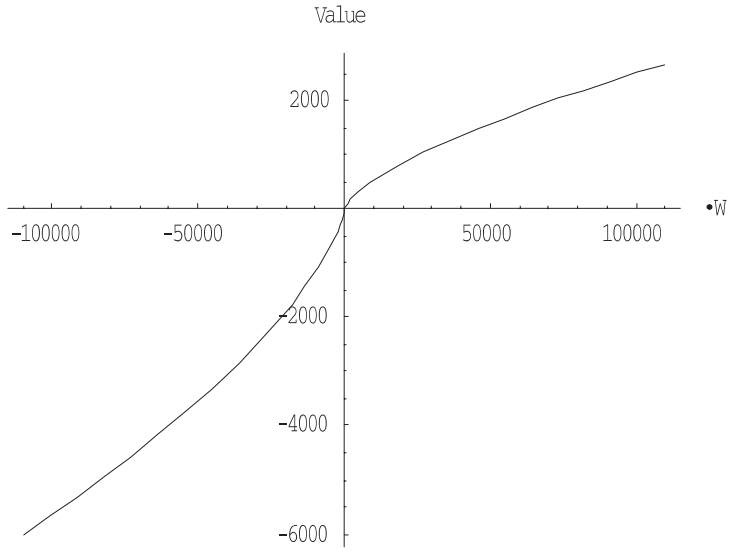
$V+''(x)= \alpha\ (\alpha-1)\ x^{\alpha-2}$; it is negative: a large profit has an incrementally smaller impact on the "utility" of the individual.

$V-''(x)= -\lambda\ \alpha\ (\alpha-1)\ (-x)^{\alpha-2}$; it is positive, larger losses have a numbing effect.

Hence, the value of a large loss is higher than the sum of the value of losses. In other words the agent's utility resides in incurring a sharp hit than the same amount in piecemeal tranches. A loss of 100 (blowup) is better (from the value function standpoint) than 100 times a loss of 1 (bleed).[8]

Note that, by comparison, the attributes of the conventional Von Neuman-Morgenstern utility of wealth (instead of payoffs). It results in asymmetry in preferences: U(W) is concave for all levels of wealth W which makes the investor prefer the frequent small

FIGURE 1
The Kahneman-Tversky Value Function



Note: The estimation in Tversky and Kahneman (1992) for =2.25; = .65: losses are 2.25 times worse than profits. Note that a loss of $100,000 has for value –56,517, a loss of $1000 has for value –982, 1/57th of the pain. The relationship reverses for the gain domain.

losses to the large occasional one, as well as the frequent small profits to the large infrequent ones. In the domain of gains or increase in wealth there is an convergence between the two methods of viewing utility.

So far the value function seems to confirm the preference for skewness hypothesis. Prospect theory has been subjected to all manner of experiments and the concavity in the domain of losses has shown to be robust. However we are confronted here with a modicum of ambiguity–the overestimation of the odds in the probability weighting function seems incompatible with the statement in the previous section concerning the underestimation of the outliers.

We can safely ignore at the probability weighting function, as we are looking at the results of risk taking in a framework of purely inductive inference, where the probabilities and the risks are not supplied, only discovered by agents and therefore subjected to cognitive biases. Recent research (Barron and Erev, 2003) shows experimental evidence that agents underweight small probabilities when they engage in sequential experiments in which *they derive the probabilities themselves*. Whether this comes from biases in our inductive inference machinery or the fact that we do not handle abstract probabilities properly (the "risk as feeling" theories) is to be ascertained. Note that the intuition of the problem is presented in an early paper by Slovic, Fischhoff, Lichtenstein, Corrigan, and Combs (1977), with the explanatory title of "preference for insuring against probable small losses"; they attributed their results consistent with the neglect of large infrequent losses to the "the sequential nature of the problem."[9]

Hedonic Adaptation and Quality of Life Perspective

The central idea behind recent research on well-being in hedonic psychology is the existence of a set-point of happiness, to which the agent tends to revert after some circumstantial departure. A positive or negative change in material conditions brings some changes in the individual well-being, but soon the process of habituation causes the reversion to the old level of life satisfaction. It is the equivalent of the utility curve resetting at the origin in the prospect theory case and the new changes in conditions mattering more–as if the accumulated changes did not bring a permanent change in one's utility.

The idea has been called the "hedonic treadmill" after the seminal paper, Brickman and Campbell (1971). Studies document that paraplegics after suffering from the onset of the impairment converge soon after to the general level of happiness of the population. Lottery winners also do not seem to hold on to significant permanent gains in their happiness and well-being. Academics granted tenure are no happier a few months later than they were before. The same applies to general societies experiencing abundance. Such mechanisms of adaptation are the backbone of the research on happiness and economics–an emerging branch in research, see Layard (2003), Frey and Stutzer (2002). Indeed utility was associated by Bentham to a measure of individual and communal happiness; it seems that economics has made a return to it.[10]

Seen in the context of skewness, the notion of habituation implies the following: *concavity* of good events;

convexity of bad ones. Indeed the value function of "hedonic enjoyment" is deemed to have the same higher order properties as prospect theory (Kahneman, 1999). Again consider if the value function in the positive domain, V+[x] is concave, then the implication is that it is better to receive a steady flow of "good" events than the same quantity in one block–whether they are monetary gains or flows of enjoyment.

Consider two economic agents operating with mirror portfolios and strategies. The first, whom we call Nero, loses $1000 for 99 consecutive weeks (he "bleeds"), then makes $99,000 on the 100th. The second, whom we call Carlos, has the exact opposite payoff (he "blows up" on the 100th week). According to such aspect of the hedonic literature, Carlos' well-being and quality of life should be superior to those of Nero. The arithmetic sum of the pleasure/pain should swing squarely in his favor: consider that Carlos will experience 99 pleasurable weeks, will go to work every Monday with the expectation of more good news, and that the pain experienced from the loss will be short lived since he will recover from it soon after. As to Nero, the exhilaration of the gain will not compensate the lengthy bleed period. As such this theory provides the explanation that, everything else being held equal, for a given mathematical expectation of a payoff, a negatively skewed one provides higher quality of life. A few question remains, however, to answer before the above argument can be accepted.

1. It seems that to analyze the summation of utilities through time might not be straightforward as a measure of total utility–and it would be normative to assume that the agent *should* be subjected to one instead of another. Indeed research (see Kahneman, 1999) from such experiments as those of subjects undergoing colonoscopy, show that those do not base their decisions on past linear summation of utility, but to more complicated rules that tend to favor the peak and ending part of the sequence ("peak-end rule"). This gave rise to four possible utilities:

   a. experienced utility–the summation of the value function over the periods considered.
   b. remembered utility–the agent's recollection of the total experienced utility, often at odds with the previous one.
   c. predicted utility–the utility that the agent believes will result from the action.
   d. decision utility–the utility used in the decision process.

We assumed that the experienced utility (here the sum of the value function over time) was the one that mattered. Further experiments are needed to confirm such a point. Would it be the case that Nero, in spite of his negative experienced utility, would have a higher remembered utility from the episode?

2. It seems that such treadmill effects are selective and domain specific: there are things that lead to permanent happiness, or to an injection of utility that carries permanent effects. In all of these situations we do not revert to the origin or the set point. In some cases, repetition or duration of a constant stimulus even results in an increasing hedonic response–a process the literature calls sensitization, the exact opposite of the treadmill effect. The well-being literature (Frederick and Loewenstein (1998)) shows evidence that there are:

   a. Some things to which we adapt rapidly: (imprisonment, increases in wealth, and disabilities like paralysis),
   b. Condition to which we adapt slowly (the death of a loved one), and
   c. Things to which we do not seem to adapt (noise, debilitating diseases, foods, or an annoying roommate).

Now the question: do people adapt to "bleeding"? In other words do people increase in sensitivity to the pain of the "Chinese torture" treatment of slow losses? On this experiments should be done.

Conclusion

This discussion has explored skewness from the utility standpoint in addition to the perception of the probability of large adverse shocks. Prospect theory provides hints on this, but further research is needed to examine how agents react in a multi-period framework. This discussion also found evidence in the literature for the undervaluation of the probability of large adverse shocks when risks are neither salient nor directly observable. This may explain the appetite for negatively skewed payoffs. Finally, more research is needed for determining the relationship between utility of *streams* of payoffs and decision making.

Notes

1. We ignore in this discussion economic arguments justifying skewness, the most significant of which is the "moral hazard" argument. This argument stipulates that agents risking other people's capital would have the incentive to camouflage the properties by showing a steady income. Intuitively, hedge funds are paid on an annual basis while disasters happen every four or five years, for example. The fund manager does not repay his incentive fee.
2. See Gladwell (2002) for a popular presentation of the difference between the two classes of strategies.
3. DeGeorge and Zeckhauser (1999) show the skewness in the distribution of the difference between announced and expected corporate earnings. For an illustration of the custom by a master of the practice see the memoirs of Jack Welsh (Welsh, 2001)

who explains explicitly (and candidly) how he managed to use accounting methods to the smooth the earnings of the conglomerate GE.

4. See Wilmott(1998) and Taleb (1997) for a discussion of dynamic hedging properties for an option seller.
5. This is called "delta hedging" where the operator buys and sells the underlying security to respond to the changes in sensitivity of the option to the underlying security. As the underlying price rises the option trader may be insufficiently covered and would need to buy more of the asset. Likewise the operator needs to sell in response to the fall in the asset price. It is key here that volatility causes losses for such an agent –particularly discontinuities and jumps in the underlying security.
6. See Rabin(2000) for a modern treatment of the "small number" problem.
7. There have been few studies of sequential behavior –see Thaler and Johnson (1990) study of the sequential behavior of agents to see how they are affected by gains and losses.
8.
9. There is a relevant recent piece of research in the recent literature on the affect heuristic (the tendency to determine the probability of an event by the emotional response that it causes). Hsee and Rottenstreich (2004) show that agents, when subjected to valuation "by feeling" (as opposed to valuation by calculating – a process that is not subjected to the affect heuristic) tend to be sensitive to the presence or absence of a given stimulus rather than its magnitude. This implies that a loss is a loss first, with further implications later. The same with profits. This explains the concavity/convexity of the value function. The agent would prefer the number of losses to be low and the number of gains to be high, rather than optimizing the total performance
10. Bentham's definition (Behnthm, 1789): "By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered: if that party be the community in general, then the happiness of the community: if a particular individual, then the happiness of that individual."

References

Alpert, M. and H. Raiffa, H. "A Progress Report on the Training of Probability Assessors" in D. Kahneman, P. Slovic & A. Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press, 1982, pp.294–305.

Barron, G. and I Erev. "Small Feedback-based Decisions and Their Limited Correspondence to Description-based Decisions." *Journal of Behavioral Decision Making,* 16, (2003), pp. 215–233.

Bentham, J. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1789,1996.

Board, J., C.Sutcliffe and E. Patrinos. "Performance of Covered Calls." *European Journal of Finance*, 6, (2002), pp. 1–17.

Brickman, P. and D. T. Campbell. "Hedonic Relativism And Planning The Good Society" in M.H. Apley, ed., *Adaptation-Level Theory: A Symposium*. New York: Academic Press, 1971, pp. 287–302.

DeGeorge, J.P. and R. Zeckhauser. "Earnings Management to Exceed Thresholds." *Journal of Business*, 72, (1999).

Frederick, S., and G. Loewenstein. "Hedonic Adaptation" in *Well Being: Foundations of Hedonic Psychology*, ed., D. Kahneman, E. Diener and N. Schwartz. New York: Russell Sage Foundation, 1998.

Frey, B.S. and A. Stutzer. *Happiness and Economics: How the Economy and Institutions Affect Human Well-bein*g. New Jersey: Princeton University Press, 2002.

Fung, W. and D. Hsieh. "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds." *The Review of Financial Studies,* 10, (1997), pp. 275–302.

Gladwell, M. "Blowing Up." *New Yorker*, April 22, 2002.

Hilton, D. "Psychology and the Financial Markets: Applications To Understanding and Remedying Irrational Decision-Making" in I. Brocas and J.D. Carillo eds., *The Psychology of Economic Decisions: Vol 1: Rationality and Well-Being*. Oxford: Oxford University Press, 2003.

Hsee, C. K. and Y. R. Rottenstreich. "Music, Pandas and Muggers: On the Affective Psychology of Value." Forthcoming, *Journal of Experimental Psychology.*

Kahneman, D. and D. Lovallo. "Timid Choices and Bold Forecasts: A Cognitive Perspective On Risk-Taking," *Management Science*, 39, (1993), pp. 17–31.

Kahneman, D. "Objective Happiness." In *Well Being: Foundations of Hedonic Psychology*, ed., D. Kahneman, E. Diener and N. Schwartz. New York: Russell Sage Foundation, 1999.

Kahneman, D. and A. Tversky. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica*, (1979).

Kat, H. "CTAs and Hedge Funds: A Marriage Made in Heaven." Cass Business School, London City University, Manuscript, 2002.

Layard, R. *Happiness: Has Social Science a Clue?* Robins Memorial Lecture, London School of Economics, 2003.

Lowenstein, R. W*hen Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House, 2000.

Rabin, M. "Inference by Believers in the Law of Small Numbers." Economics Department, University of California, Berkeley, manuscript, 2000.

Slovic, P., B. Fischhoff, S. Lichtenstein, B. Corrigan and B. Combs. "Preference for Insuring Against Probable Small Losses: Implications For The Theory And Practice Of Insurance." *Journal of Risk and Insurance*, 44, (1977), pp. 237–258. Reprinted in P. Slovic, ed., *The Perception of Risk.* London: Earthscan.

Taleb, N. *Dynamic Hedging: Managing Vanilla and Exotic Options*. New York: Wiley, 1997.

Thaler, R., and E. Johnson. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes in Risky Choice." *Management Science*, June (1990).

Tversky, A. and D. Kahneman. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 5, (1990), pp. 297!323.

Tversky, A. and D. Kahneman. "Belief in the Law of Small Numbers." *Psychology Bulletin,* 76(2), (1971), pp. 105–10.

Welsh, J. *Jack*: *Straight From the Gut*. New York: Warner Books, 2001.

Wilmott, P. *Derivatives*. Chichester: Wiley, 1998.

Nassim Nicholas Taleb
New York University

Randomness and the Prediction of Action

A. PILPEL & N.N. TALEB

In *Blackwell's Companion to the Philosophy of Action*

INTRODUCTION

The philosophical motivation behind the interest in action (unless otherwise noted, by 'action' here we mean – as is usually the case – intentional human action by a single individual) is, traditionally, twofold. First, it is metaphysical: e.g., how various theories of action relate to the mind-body problem, to the age-old problem of free will (e.g., is there such a thing as a free action in a deterministic world), etc. Second, it is ethical and legal: whatever view action one accepts relates directly to the issue of what legal and moral responsibility, if any, an agent has for her (free or otherwise) actions. Both of these issues are dealt with elsewhere in this collection.

In this part (part III) of the book, the contributors concentrate, instead, on the scientific theories, and their prediction and explanation of action. It is mostly concerned with epistemology, rather than with metaphysics or ethics: when, and under what circumstances, do we know enough about the person's psychology (or folk psychology or evolutionary psychology), neuroscience, ethnology, social position, etc., to predict (or explain) her actions. Of course, the science of action affects the philosophy of action and vice versa: e.g., developments in neuroscience can challenge free will, while on the other hand Librtarianism (in the Incompatibilist sense, not the political sense) might come up with convincing arguments why no scientific advance will ever eradicate it.

The particular scientific theories and their relationship to action have been dealt with in other entries in the third section. In this entry, we wish – fittingly, perhaps, for the last entry in the last section of the collection that deals with general metaphysical and scientific theories of action (leaving only the fourth section which deals with the specific theories of individual philosophers) – to take a somewhat different track. Instead of arguing for a particular metaphysical or scientific view of action, we want to make a general point about predicting action, one that does not depend on accepting any particular metaphysical or scientific theory about it.

OUR ARGUMENT

We argue that X's action in situation T is hard (in a sense to be made clear presently) for Y to predict, precisely in those situations where, for whatever reason, predicting X's action is, for Y, a case of prediction under uncertainty; while it is comparatively easy when, as sometimes occurs, predicting X's actions at T is, for Y, a case of prediction under known and computable probabilistic structure (such as that discussed by von Neumann and Morgenstern [1944]). This is the case whether X (or Y) are individuals or groups (such as, say, economic agents involved in transactions, or voters in an election), or indeed when X = Y (as in the case of an agent considering her own future actions.) To avoid a possible misunderstanding, we are not dealing with why prediction X's actions is often a case of prediction under uncertainty for Y. The libertarian and the (hard-determinist) neuroscientist will have very different answers. We are only saying that it is in those cases that predicting X's action is (especially) difficult for Y.

We argue, further, that this distinction between tractable probabilistic structure and unstable, more complicated uncertainty is of great practical importance. It helps explains why in some cases the rational choice and economics literature considers that it is easier to predict the action of entire groups than of individuals, or why others may know our actions better than we know ourselves. It also explains why so many confident predictions fail: very often, the predictor falls into we call the 'ludic fallacy' – i.e., the creation of a crisp structure of games from which we can produce analytical responses that only hold in this artificial construct, and break down outside of it. It leads to the mistake of confusing a situation that is one of predicting under unstructured uncertainty as if it is one of predicting under structured probability (as in games of chance, hence the fallacy's name).

CLASSES OF UNCERTAINTY

In rational choice theory, there are three types of decision making. The first is decision making under certainty, when Y knows the (single) state of the world, S1, and therefore what outcome oi1 will be the certain result of each of his possible choices.

The second is decision making under known and computable probabilistic structure, sometimes called "risk" following Knight (1921), when Y knows not only the possible states of the world, S1, S2, … Sj, … Sm, but also the probability each of them will occur. This is, for example, the case in games of chance (Y knows what numbers are on the roulette wheel and also what the probability is of each one coming up.)

The third is that of true uncertainty, when there isn't even any reasonable probability value to give to the possible outcomes, or, as Keynes put it, where 'there is no scientific basis to form any calculable probability whatever': giving, as an example, the 'price of copper and the rate of interest twenty years hence' [Keynes 1937]. In particular, games are a case where both sides must make decisions under uncertainty, since it is precisely the fact that the opponent is a human being that makes free choices and can act as she wishes (choosing any of the possible strategies within the context of the game) that makes it impossible to assign probabilities to her choice, as if she were a slot machine (see Von Neumann and Morgenstern [1944] or any of the numerous treatised on game theory, such as the classic Luce and Raiffa [1958]).

The distinction between the three cases is not absolute: choice under certainty can be seen as a limiting case of choice under known probabilistic structure, with the probability p=1 for one outcome and p=0 for all the rest, although, as Levi [1980] points out, the analogy is not exact since having probability 1 is not logically the same as being certain. (There is a probability of 1 that randomly choosing a number between 0 and 1, 0.5 will not be chosen, but it is it not certain that it won't be.) Uncertainty, too, admits of degrees: one can be certain that the probability of an event is between 0.2 and 0.8, while being uncertain about what it is, as opposed to complete uncertainty, where no probability value at all can be assigned – or, equivalently, the agent is in complete uncertainty, assigning no narrower range of probabilities to the event than [0,1] (see Levi [1980]).

PREDICTING OTHER PEOPLE'S ACTION

There is a well-known tension between free choice (or, more generally, free action, or free will) and prediction. This goes back to St. Augustine [Augustine 1988] who wondered how free will is possible in a world where God has foreknowledge of all events, and is much older than that. This was discussed elsewhere in this book.

The tension applies to probability as well: one is no more freely choosing what to have for breakfast tomorrow if an all-knowing being realized one has exactly a 87.5% of having fried eggs and a 12.5% of having cornflakes than that being decided one has a 100% probability of having fried eggs. Indeed, some modern thinkers the "standard" probability-centered view of economics as misguided due to its lack of concern for people's freedom to choose [Shackle 1979]. As said above, it is for this reason that outcomes are considered cases of decision under true uncertainty, not computable probabilistic structure.

Typically, when Y attempts to predict X's actions and X is an individual agent, then Y is in a situation where (as in games) the prediction is impossible since one is in a situation of uncertainty: Y has no way to assign X's possible choices (the possible future actions he will decide to take) any probability, since X – from Y's point of view – has freedom of choice in his actions. This situation is very common – it is probably the typical case when it comes to predicting individual actions – and is what was above called the "hard" case for prediction of X's actions.

The reason that the situation is typically that of uncertainty and not of risk is, to describe things using the terminology of randomness, that in order to have a good idea of what X's probabilities of actions are Y needs to know what X's "generator" – that is, the generating function that determines the mean, variance, and higher-level moments of his actions – what makes him tick, what is it in X that makes X have a certain probability to choose one way and another probability to choose another way.

But, with human beings (for whatever reason, free will or neurochemical complexity or…), the generator is hidden, and there is no reason to believe it is of a "good" type. X's generating process might not even have truly quantifiable properties, including such metrics used in statistical methods such as the mean – let alone a variance; these might change over time; and so on. But typically, for Y to reduce the situation to that of risk – as in games of chance – X's generator must be known, of a "good" type (has a finite mean), and be stable. Usually, when Y attempts to predict X's actions, none of these conditions hold [see Taleb 2007a, 2007b].

Nevertheless, Y can try to reduce this uncertainty: Y can learn about X's psychological state, his genetics, his social position, and so on, in an attempt to predict X's actions. This is sometimes successful. It is not at all rare for X's spouse, or psychologist, or co-workers, to assign quite reasonable probabilities (indeed, sometimes even certainties) to X's future actions – to "know X better than X knows themselves". This means that Y has managed to reduce the situation to that of risk (or certainty) instead of uncertainty – that Y is able, due to his knowledge of X, to assign (reliable)

probabilities to X's future actions. This is the "easy" case.

It should be noted that, from Y's point of view, X is not a free agent in the indeterminist sense of the term – X has no choice but to act as he does (either perform a certain action for sure, or to choose between several actions with a given probability for each). At most X can be a free actor in the compatilibist / "soft" determinist sense of (roughly) doing what one wants to do even if one could not help but want to do it.

The disagreement between different views about free will and determinism can be described as a disagreement about whether, and if so under what conditions, can Y ever actually make the prediction of X's actions a case of prediction under risk instead of under uncertainty: determinists say it could in principle, indeed sometimes in practice, be done, indeterminists deny it since they believe free will does not allow it, and in particular say that alleged counterexamples such as one's spouse knowing what one will do tomorrow before oneself do not work for various reasons. (For a more detailed discussion of the free will / determinism issue, see elsewhere in this book).

PREDICTING ONE'S OWN ACTIONS

It is important to consider what happens when X = Y – when an agent attempts to predict their own future actions. As Levi [1990], Shackle [1979], and many others noted, inasmuch as an agent succeeds in this task, the agent is no longer acting freely (making a genuine choice) in the future: if I determine now that I will have eggs for breakfast tomorrow, I can predict my future actions, but I will no longer be making a choice tomorrow about what to have for breakfast. My actions tomorrow have been determined by the time of prediction. Predicting one's own future actions is sometimes possible, but at a price.

PREDICTING GROUP ACTION

There is another way to reduce the uncertainty in prediction of action that, it seems, is inherent to the human condition where people make choices. It is for Y to predict, not what an individual X is doing, but what a group of individuals is doing. Here, the law of large numbers often comes to Y's aid. It is sometimes possible to be able to predict the behavior of large groups of people – that of, say, the Republicans, or the investors in the market – despite the fact that each individual is unpredictable. If X is a group or organization, its potential actions might be (from Y's point of view) describable with reliable probabilities despite the fact that each individual in the group cannot be.

However, one must be very careful. For the law of large numbers to be applicable the random generators of individuals in the group must be independent of each other. If independence does not hold – if agents do not make decisions in isolation but while considering the actions of other agents, if Republicans do not decide independently on each issue but also take into account what other Republicans are doing – then convergence to commonly tractable properties will not take place, making the law of large numbers of little applicability and use .

This behavior of the aggregate is qualitatively central. There is no exact definition on what constitutes a "complex system", except for a consensus across the interdisciplinary literature that the degree of interdependence or "connectedness" of the elements is an essential determining property. The difference is crucial: in an ordinary system, the agents might not be predictable, but the various idiosyncrasies will tend to compensate each other, and the aggregate will appear more stable than any of its components, hence more predictable. However this cancelling-out effect is lost when the agents start acting in lock-steps, with contagions and feed-back loops causing exacerbation of the properties. In such situations we have, in effect, a "group mind" with its own single generator – a single individual (from the point of view of predicting its actions) – and, what's more, an "individual" more prone, if anything, to have a "bad" sort of generator, one that makes predicting its actions difficult (a case of prediction under uncertainty) than an actual person, due to the disproportionate effect extreme individuals tend to have on a group's behavior, as can be seen in panics and bubbles (e.g., Mackay [1995]).

THE DANGER OF PREDICTION

To summarize, predicting X's action is (relatively) easy when one has reason to believe X's actions can be described under risk – with reliable probability functions. This is not impossible. But there are two risk involved here that are often involved.

The first is what Taleb and Pilpel (2007) called the nonobservability of the generators of the random process, also described as the inverse problem. Upon observing a series of points, an infinity of generators, from four qualitatively different classes, can be ascribed to the data. In a way similar to Goodman's riddle of induction, [Goodman, x] , the empirical data can justifiably lead to two completely opposite extrapolations. Indeed agents fall prey to choosing from the data what confirms (does not disconfirm) their theories.

This is related to the ludic fallacy, coming from the assumption by agents that outcomes resemble the structures of games of chance. The conventional economics and rational choice literature has traditionally assumed that people confront clear choices, in one period, with clear answers (even in some situations of unknown probabilities). The error is to believe that the passage from the "ludic", casino-like analysis can be generalized outside of it. Indeed people tend to overestimate their knowledge of the world – here, they tend to overestimate, often ludicrously so, the amount of knowledge they have about the 'random generator' of X, whose actions they're trying to predict. They tend to treat both other individuals and other groups as if those groups and individuals are as simple as games of chance – as if the behavior of Republicans or of the market or of individual strangers could, with a little shoe-horning perhaps, be described in terms of probability functions no more complex, and as (or more) reliable, than that of a roulette wheel.

So the "forcing" of a situation of prediction under uncertainty of X's future actions into one under known probabilistic structure commits two main errors. First of all, as we said, it assigns probabilities to X's known actions when there is no justification. What's more, and worse, it tends to ignore unknown and unimagined actions X could take: the very fact of analyzing X's possible actions in terms of risk means to have a set of possible outcomes (X's actions) among which the probability is distributed. But there is an unknown risk taken that some actions have been forgotten or ignored – typically, the most extreme ones! So we are not just dealing with the underestimation of the magnitude of possible outcomes, but with the possible sources of randomness. And such sources of randomness about other's actions have a disproportionately high effect in real-life situations .

The second problem is that of high impact uncertainty. or consequential low probability events. Sometimes we may be able to predict an agent's action in ordinary circumstances where such prediction does not carry serious consequences, yet fail in those situations where prediction matters. We may be able to predict what a criminal can eat for breakfast, but miss out on whether or when he may commit a crime. We may be able to predict what the pilot would do on the weekend, but not if he will crash the plane. The point is serious as 1) these less ordinary, low probability events have a structure that is less computable than ordinary events, and 2) they represent the bulk of what is meaningful to predict (Taleb and Pilpel, 2007). Indeed the role of these high-impact outliers is dominant in history, economic life, and politics.

REFERENCES

Augustine of Hippo (au.), Dyson, R. W. (trs.) (1988). The City of God against the Pagans. Cambridge: Cambridge University Press.

Goodman, x

Keynes, J. M. (1937), 'The General Theory of Employment', Quarterly Journal of Economics, February 1937.

Knight, F. H. (1921), Risk, Uncertainty, and Profit. Boston: Hughton Mifflin.

Levi, I. (1980). The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance. Cambridge: MIT Press.

Levi, I. (1990). Hard Choices: Decision Making Under Unresolved Conflict. Cambridge: Cambridge University Press.

Luce, R. D. and Raiffa, H. (1958). Games and Decisions. New York: Wiley.

Mackay, C. (au.) and Tobias, A. (forward) (1995). Extraordinary Popular Delusions and the Madness of Crowds. New York: Three Rivers Press.

Shackle, G. L. S. (1979). Imagination and the Nature of Choice. Edinburgh: Edinburgh University Press.

Taleb, N. N. (2007a). The Black Swan: The Impact of the Highly Improbably. New York: Random House and London: Penguin.

Taleb, N. N. (2007b)." Black Swans and the Domains of Statistics", *The American Statistician*, August 2007, Vol. 61, No. 3.

Taleb, N.N. and Pilpel, A. 2007, "Epistemology and Risk Management", *Risk and Regulation*, *Risk and Regulation*, June

Von Neumann, J. and Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton: Princeton University Press.

On the Unfortunate Problem of the Nonobservability of the Probability Distribution

Nassim Nicholas Taleb and Avital Pilpel[1]

SECOND DRAFT OCTOBER 2004

Cannot be quoted without permission[2]

[1] We thank participants at the American Association of Artificial Intelligence Symposium on Chance Discovery in Cape Cod in November 2002, Stanford University Mathematics Seminar in March 2003, the Italian Institute of Risk Studies in April 2003, and the ICBI Derivatives conference in Barcelona in May 2003. We thank Benoit Mandelbrot and Didier Sornette for helpful comments.

[2] This paper was never submitted to any journal. It is not intended for submission.

A severe problem with risk bearing is when one does not have the faintest idea about the risks incurred. A more severe problem is when one does not have the faintest idea about the risks incurred yet thinks he has a precise idea of them. Simply, one needs a probability distribution to be able to compute the risks and assess the likelihood of some events.

These probability distributions are not directly observable, which makes any risk calculation suspicious since it hinges on knowledge about these distributions. Do we have enough data? If the distribution is, say, the traditional bell-shaped Gaussian, then yes, we may say that we have sufficient data. But if the distribution is not from such well-bred family, then we do not have enough data. But how do we know which distribution we have on our hands? Well, *from the data itself*. If one needs a probability distribution to gauge knowledge about the future behavior of the distribution from its past results, and if, at the same time, one needs the past to derive a probability distribution in the first place, then we are facing a severe regress loop–a problem of self reference akin to that of Epimenides the Cretan saying whether the Cretans are liars or not liars. And this self-reference problem is only the beginning.

What is a probability distribution? Mathematically, it is a function with various properties over a domain of "possible outcomes", X, which assigns values to (some) subsets of X. A probability distribution describes a general property of a system: a die is a *fair die* if the probability distribution assigned to it gives the values… It is not that different, essentially, than describing mathematically other properties of the system (such as describing its mass by assigning it a numerical value of two kilograms).

The probability function is derived from specific instances from the system's past: the tosses of the die in the past might justify the conclusion that, in fact, the die has the property of being fair, and thus correctly described by the probability function above. Typically with time series one uses the past for sample, and generates attributes of the future based on what was observed in the past. Very elegant perhaps, very rapid shortcut maybe, but certainly dependent on the following: that the properties of the future resemble those of the past, that the observed properties in the past are sufficient, and that one has an idea on how large a sample of the past one needs to observe to infer properties about the future.

But there are worst news. Some distributions change all the time, so no matter how large the data, definite attributes about the risks of a given event cannot be inferred. Either the properties are slippery, or they are unstable, or they become unstable because we tend to act upon them and cause some changes in them.

Then what is all such fuss about "scientific risk management" in the social sciences with plenty of equations, plenty of data, and neither any adequate empirical validity (these methods regularly fail to estimate the risks that matter) or any intellectual one (the argument above). Are we missing something?

An example. Consider the statement "it is a ten sigma event"[3], which is frequently heard in connection with an operator in a stochastic environment who, facing an unforeseen adverse rare event, rationalizes it by attributing the event to a realization of a random process whose moments are well known by him, rather than considering the possibility that he used the wrong probability distribution.

Risk management in the social sciences (particularly Economics) is plagued by the following central problem: *one does not observe probability distributions, merely the outcome of random generators*. Much of the sophistication in the growing science of risk measurement (since Markowitz 1952) has gone into the mathematical and econometric details of the process, rather than realizing that the hazards of using the wrong probability distribution will carry more effect than those that can be displayed by the distribution itself. This recalls the story of the drunkard looking for his lost keys at night under the lantern, because "that is where the light is". One example is the blowup of the hedge fund Long Term Capital Management in Greenwich, Connecticut[4]. The partners explained the blowup as the result of "ten sigma event", which should take place once per lifetime of the universe. Perhaps it would be more convincing to consider that, rather, they used the wrong distribution.

It is important to focus on catastrophic events for this discussion, because they are the ones that cause the more effect –so no matter how low their probability (assuming it is as low as operators seem to believe) the effect on the expectation will be high. We shall call such catastrophic events *black swan events*. Karl Popper remarked[5] that when it comes to generalizations like "all swans are white", it is enough for *one* black swan to exist for this conclusion to be false. Furthermore, before you find the black swan, *no amount of information* about white swans – whether you observed one, 100, or 1,000,000 of them – could help you to determine whether or not the generalization "all swans are white" is true or not.

We claims that risk bearing agents are in the same situations. Not only can they not tell before the fact whether a catastrophic event will happen, but *no amount of information about the past behavior of the market* will allow them to limit their ignorance--say, by assigning meaningful probabilities to the "black swan" event. The only thing they can honestly say about catastrophic events before the fact is: "it might happen". And, if it *does* indeed happen, then it can *completely destroy our previous conclusions about the expectation operator*, just like finding a black swan does to the hypothesis "all swans are white". But by then, it's too late.

Obviously, mathematical statistics is unequipped to answer questions about whether or not such catastrophic events will happen: it *assumes* the outcomes of the process we observe is

---

[3] That is, an event which is ten standard deviations away from the mean given a Normal distribution. Its probability is about once in several times the life of the universe.

[4] Lowenstein 2000.

[5] We use "remarked" not "noticed"—Aristotle already "noticed" this fact; it's what he did with the fact that's important.

governed by a probability distribution of a certain sort (usually, a Gaussian curve.) It tells us nothing about why to prefer this type of "well behaved" distributions to those who have "catastrophic" distributions, or what to do if we suspect the probability distributions might change on us unexpectedly.

This leads us to consider epistemology. By epistemology we mean the problem of the theory of knowledge, although in a more applied sense than problems currently dealt with in the discipline: *what can we know about the future, given the past*? We claim that there are good philosophical and scientific reasons to believe that, in economics and the social sciences, one *cannot* exclude the possibility of future "black swan events".

THREE TYPES OF DECISION MAKING AND THE PROBLEM OF RISK MANAGEMENT

Suppose one wants to know whether or not P is the case for some proposition P – "The current president of the United States is George W. Bush, Jr."; "The next coin I will toss will land 'heads'"; "There are advanced life forms on a planet orbiting the star Tau Ceti".

In the first case, one can become *certain* of the truth-value of the proposition if one has the right data: who is the president of the United States. If one has to choose one's actions based on the truth (or falsity) of this proposition – whether it is appropriate, for example, to greet Mr. Bush as "Mr. President" – one is in a state of *decision making under certainty*. In the second case, one cannot find out the truth-value of the proposition, but one can find out the *probability* of it being true. There is – in practice - no way to tell whether or not the coin will land "heads" or "tails" on its next toss, but under certain conditions one can conclude that p('heads') = p('tails') = 0.5. If one has to choose one's actions based on the truth (or falsity) of this proposition – for example, whether or not to accept a bet with 1:3 odds that the coin will land "heads" – one is in a state of *decision making under risk*.

In the third case, not only can one not find out the truth-value of the proposition, but one cannot give it any meaningful probability. It is not only that one doesn't know whether advanced life exist on Tau Ceti; one does not have any information that would enable one to even estimate its probability. If one must make a decision based on whether or not such life exists, it is a case of *decision making under uncertainty*.[6] See Knight, 1921, and Keynes, 1937, for the difference between risk and uncertaintly as first defined.

More relevant to economics, is the case when one needs to make decisions based whether or not future social, economical, or political events occur – say, whether or not a war breaks out.

---

[6] For the first distinction between risk and uncertainty see Knight (1921) for what became known as "Knightian risk" and "Knightian uncertainty". In this framework , the distinction is irrelevant, actually misleading, since, outside of laboratory experiments, the operator does not know beforehand if he is in a situation of "Kightian risk".

Many thinkers believed that, where the future depends not only on the physical universe but on human actions, there are no laws – even probabilistic laws – that determine the outcome; one is always "under uncertainty"[7]. As Keynes(1937) says:

> By "uncertain" knowledge… I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty… The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention… About these matters, there is no scientific basis on which to form any calculable probability whatever. We simply do not know![8]

Certainty, risk, and uncertainty differ not merely in the probabilities (or range of probabilities) one assigns to P, but in the strategies one must use to make a decision under these different conditions. Traditionally, in the "certainty" case, one chooses the outcome with the highest utility. In the "risk" case, one chooses the outcome with the highest *expected* utility[9]. In the (completely) "uncertain" case, many strategies have been proposed. The most famous one is the minmax strategy (von Neumann and Morgenstern, 1944; Wald, 1950), but others exist as well, such as Savage's "minmax of Regret" or "Horowitz's alpha". These strategies require bounded distributions. In the event of the distributions being unbounded the literature provides no meaningful answer.

THE CENTRAL PROBLEM OF RISK BEARING

Using a decision-making strategy relevant to decision under risk in situations that are best described as cases of uncertainty, will lead to grief. If a shadowy man in a street corner offers me to play a game of three-card Monte, I will quickly lose everything if I consider the game a

---

[7] Queasiness about the issue of uncertainty, especially in the case of such future events, had lead Ramsey (1931), DeFinetti(1937), and Savage(1954) to develop a "personalistic" or "subjective" view of probability, independent of any objective chance or lack thereof.

[8] "We simply do not know" is not necessarily a pessimistic claim. Indeed, Shackle(1955) bases his entire economic theory on this "essential unknowledge" – that is, uncertainty - of the future. It is this "unknowledge" that allows for effective human choice and free will: for the human ability to *create* a specific future out of "unknowledge" by its efforts.

[9] These ideas seem almost tautological today, but this of course is not so. It took von Neumann and Morgenstern(1944), with their rigorous mathematical treatment, to convince the world that one can assign a meaningful expected-utility function to the different options when one makes choices under risk or uncertainty, and that maximizing this expected utility (as opposed to some other parameter) is the "rational" thing to do. The idea of "expected utility" *per se* is already in Bernoulli(1738) and Cramer(1728), but for a variety of reasons its importance was not clearly recognized at the time.

risk situation with p(winning) = 1/3. I should also consider the possibility that the game is rigged and my actual chances of winning are closer to zero. Being uncertain where in the range [0, 1/3] does my real chance of winning lies should lead one to the (minmax) uncertainty strategy, and reject the bet.

We claim that the practice of risk management (defined as the monitoring of the possibility and magnitude of adverse outcomes) subjects agents to just such mistakes. We argue below that, for various reasons, risk managers cannot rule out "catastrophic events". We then show that this ever-present possibility of black swan events means that, in most situations, Risk managers are *essentially uncertain* of the future in the Knightian sense: where no meaningful probability can be assigned to possible future results.

Worse, it means that in many cases no known lower (or upper) bound can even be assigned to the range of outcomes;

worst of all., it means that, while it is often the case that sampling or other actions can reduce the uncertainty in many situations, risk managers often face situations where no amount of information will help narrow this uncertainty.

The general problem of risk management is that, due to essential properties of the generators risk managers are dealing with, they are dealing with a situation of essential uncertainty, and not of risk.

To put the same point slightly more formally: risk managers look at collection of state spaces[10] that have a cumulative probability that exceeds a given arbitrary number. That implies that a *generator* of a certain *general type* (e.g., known probability distribution: Normal, Binomial, Poisson, etc. or mere histogram of frequencies) determines the occurrences. This generator has specific *parameters* (e.g. a specific mean, standard deviation, and higher-level moments) that – together with the information about its general type – determine the values of its distribution. Once the risk manager settles on the distribution, he can calculate the "risk" – e.g., the probability - of certain states of the world in which he is interested.

In almost all important cases, whether in the "hard" or "soft" sciences, the generator is hidden,. There is *no* independent way to find out the parameters – e.g. the mean, standard deviation, etc. - of the generator except for trying to *infer it from the past behavior* of the generator. On the other hand, in order to give any estimate of these parameters in the first place, one must first *assume* that the generator in question *is* of a certain general type: that it is a Normal generator, or a Poisson generator, etc. The agent needs to provide a joint estimation of the generator and the parameters.

Under some circumstances, one is justified in assuming that the generator is of a certain general type and that the estimation of parameters from past behavior is reliable. This is the

[10] By "state-space" is meant the foundational Arrow-Debreu state-space framework in neoclassical economics.

situation, for example, in the case of a repeated coin toss as one can observe the nature of the generator and assess the boundedness of its payoffs.

Under other circumstances, one might be justified in assuming that the generator is of a certain general type, but *not* be justified in using the past data to tell us anything reliable about the moments of the generator, no matter how much data one has.

Under even more troubling circumstances, one might have no justification not only for guessing the generator's parameters, but also in guessing what *general type* of generator one is dealing with. In that case, naturally, it is meaningless to assign any values to the parameters of the generator, since we don't know what parameters to look for in the first place.

We claim that most situations risk managers deal with are just such "bad" cases where one cannot figure out the general type of generator solely from the data, or at least give worthwhile estimate of its parameters. This means that any relation between the risks they calculate for "black swan" events, and the *actual* risks of such events, may be purely coincidental. We are in uncertainty: we cannot tell not only whether or not $\underline{X}$ will happen, but not even give any reliable estimate of what $p(\underline{X})$ is. The cardinal sin risk managers commit is to "force" the square peg of uncertainty into the round hole of risk, by becoming convinced without justification both of the generator type and of the generator parameters.

In the remainder of this paper we present the problem in the "Gedanken" format. Then we examine the optimal policy (if one exists) in the presence of uncertainty attending the generator.

Four "Gedanken" Monte Carlo Experiments

Let us introduce an invisible generator of a stochastic process. Associated with a probability space it produces observable outcomes. What can these outcomes reveal to us about the generator – and, in turn, about the future outcomes? What – if anything – do they tell us about its mean, variance, and higher order moments, or how likely the results in the future are to match the past?

The answer depends, of course, on the properties of the generator. As said above, Mother Nature failed to endow us with the ability to observe the generator--doubly so in the case of social science generators (particularly in economics).

Let us consider four cases. In all of these cases we observe mere *realizations* while the generator is operated from behind a veil. Assume that the draws are generated by a Monte Carlo generation by a person who refuses to reveal the program, but would offer samples of the series.

Table 1: The Four *Gedanken* Experiments

Gedanken	Probability Space	Selected Process	Effect	Comments
1	Bounded	Bernouilli	Fast convergence	"Easiest" case
2	Unbounded	Gaussian (General)	Semi-fast convergence	"Easy" case
3	Unbounded	Gaussian (mixed)	Slow convergence (sometimes *too* slow)	Problems with solutions
4	Unbounded	Stable Pareto-Lévy-Mandelbrot (with $\alpha<1$)[11]	No convergence	No known solutions

THE "REGULAR" CASE, TYPE I: DICE AND CARDS

The simplest kind of random process (or "chance setups" as they are sometimes called is when all possible realizations of the process are bounded. A trivial case is the one of tossing a die. The probability space only allows discrete outcomes between 1 and 6, inclusive.

The effect of having the wrong moments of the distribution is benign. First, note that the generator is *bounded*: the outcome cannot be more than 6 or less than 1. One cannot be off by more than a finite amount in estimating the mean, and similarly by some finite amount when estimating the other moments (although, to be sure, it might become a relatively large amount for higher-level moments) (note the difference between unbounded and infinite. As long as the moments exist, one *must* be off by only a finite amount, no matter what one guesses. The point is that the finite amount is *unbounded* by anything *a priori*. Give examples in literature—original one, preferably, E.).

Second, the bounded-ness of the generator means that there are *no extreme events*, There are no rare, low-probability events that any short run of the generator is unlikely to cover, but yet have a significant effect on the value of the true moments. There are certainly *no "black swan" events*—no outcomes whose result could destroy our previous estimates of the generator's moments no matter how much previous data we have. That is, $E(\underline{X}_n)$ (the observed mean) is likely to be close to $E(\underline{X})$ (the "real") mean since there is no rare, 1-in-

[11] More technically, Taleb shows in Taleb (2006) that the $\alpha<1$ is not necessarily the cutting point. Entire classes of scalable distributions converge to the Gaussian too slowly to be of any significance –by a misunderstanding of Central Limit Theorem and its speed of reaching the limit.

1,000,000 chance of the die landing "1,000,000" – which would raise E($\underline{X}$) from the E($\underline{X}$) of a "regular" die almost by 1 but will not be in the observed outcomes $x_1, x_2, \ldots x_n$ unless one is extremely lucky. (Make the example more mathematical.)

THE "REGULAR" CASE, TYPE II: NORMAL DISTRIBUTION

A more complicated case is the situation where the probability space is unbounded. Consider the normal distribution with density function f_2. In this case, there is a certain >0 probability for the outcome to be arbitrarily high or low; for it to be >M or <m for arbitrary M, m∈R.

However, as M increase s and m decreases, the probability of the outcome to be >M or <m becomes very small very quickly.

Although the outcomes are unbounded the epistemic value of the parameters identification is simplified by the "compactness" arument used in economics by Samuelson[12] .

A compact distribution, short for "distribution with compact support", has the following mathematical property: the moments M[n] become exponentially smaller in relation to the second moment[13] [add references to Samuelson.].

But there is another twist to the Gaussian distribution. It has the beautiful property that it can be *entirely characterized by its first two moments*[14]. All moments M[n] from n ={3,4,…,∞} are merely a multiple of M[1] and M[2].

Thus, knowledge of the mean and variance of the distribution would be sufficient to derive higher moments. We will return to this point a little later. (Note tangentially that the Gaussian distribution would be the maximum entropy distribution conditional on the knowledge of the mean and the variance.)

From this point on—consider Levi more? Also induction more? These are the things that we need to add…

---

[12] See Samuelson (1952).

[13] A Noncentral moment is defined as $M[n] \equiv \int_{\Omega} x^n \phi(x)\, dx$.

[14] Take a particle W in a two dimensional space W(t). It moves in random increments ΔW over laps of time Δt. At times t+Δt, we have W(t+Δt)= W + ΔW + ½ ΔW^2 + 1/6 ΔW^3 + 1/24 ΔW^4 + … Now taking expectations on both sides: E[W(t+Δt)] = W +M[1]+ M[2]/2 +M[3]]/6+ M[4]//24, etc. Since odd moments are 0 and even moments are a multiple of the second moment, by stopping the Taylor expansion at M[2] one is capturing most of the information available by the system.

Another intuition: as the Gaussian density function for a random variable x is written as a scaling of $e^{-\frac{(x-m)^2}{2\sigma^2}}$, we can see that the density wanes very rapidly as x increases, as we can see in the tapering of the tail of the Gaussian. The interesting implication is as follows: Using basic Bayes' Theorem, we can compute the conditional probability that, given that (x-m) exceeds a given 2 σ , that it falls under 3 σ and 4 σ becomes $\frac{\int_2^3 \phi\left(\frac{x-m}{\sigma}\right)dx}{1-\int_{-\infty}^2 \phi\left(\frac{x-m}{\sigma}\right)dx}$ =94% and $\frac{\int_2^4 \phi\left(\frac{x-m}{\sigma}\right)dx}{1-\int_{-\infty}^2 \phi\left(\frac{x-m}{\sigma}\right)dx}$ = 99.8% respectively.

THE "SEMI-PESSIMISTIC" CASE, TYPE III: "WEIRD" DISTRIBUTION WITH EXISTING MOMENTS.

Consider another case of unbounded distribution: this time, a linear combination of a "regular" distribution with a "weird" one, with very small probabilities of a very large outcome.

For the sake of concreteness assume that one is sampling from two Gaussian distributions. We have π_1 probability of sampling from a normal N_1 with mean μ_1 and standard deviation σ_1 and $\pi_2 = 1- \pi_1$ probability of sampling from a normal N_2 with mean μ_2 and standard deviation σ_2.

Assume that N_1 is the "normal" regime as π_1 is high and N_2 the "rare" regime where π_2 is low. Assume further that $|\mu_1|<<|\mu_2|$, and $|\sigma_1|<<|\sigma_2|$. (add graph.) The density function f_3 of this distribution is a linear combination of the density functions N1 and N2. (in the same graph, here:.)

Its moment-generating function, M_3, is also the weighted average of the moment generating functions M_1 and M_2, of the "regular" and "weird" normal distributions, respectively, according to the well-known theorem in Feller (1971)[15]. This in turn means that the moments

[15] Note that the process known as a "jump process", i.e., diffusion + Poisson is a special case of a mixture.The mean $m= \pi_1 \mu_1 + \pi_2 \mu_2$ and the standard deviation

themselves (μ_3, σ_3, …) are a linear combination of the moments of the two normal distributions.

While the properties of this generator and the outcomes expected of it are much less "stable" (in a sense to be explained later) than either of the previous cases, it is at least the case that the mean, variance, and higher moments exist for this generator, Moreover, this distribution over time settles to a Gaussian distribution, albeit at an unknown rate.

This, however, is not much a of a consolation when σ_2 or μ_2 are very large compared to σ_1 and μ_1, as assumed here. It takes a sample size in inverse proportion to π_2 to begin to reach the true moments: When π_2 is very small, say 1/1000, it takes at least 1000 observations to start seeing the contribution of σ_2 and m_2 to the total moments.

THE "PESSIMISTIC CASE: NO FIXED GENERATOR

Consider now a case where the generator itself is not fixed, but changes continuously over time in an unpredictable way; where the outcome x_1 is the result of a generator G_1 at time t_1, outcome x_2 that of generator G_2 at later time t_2, and so on. In this case, there is of course no single density function, moment-generating function, or moment can be assigned to the changing generator.

Equivalently, we can say that the outcome behaves as if it is produced by a generator which has no moments – no definite mean, infinite variance, and so on. One such generator is the one with moment-generating function M_4 and density function f_4 – the Pareto-Lévy-Mandelbrot distribution[16] with parametrization $\alpha<1$ providing all infinite moments, which is a case of the stable distribution "L" Stable (for Lévy-stable).

THE DIFFERENCES BETWEEN THE GENERATORS

Suppose now that we observe the outcomes x_1, x_2, x_3… x_n of generators of type (1)-(4) above, from the bound dice-throwing to the Pareto-Lévy-Mandelbrot distribution. What could we infer from that data in each case? To figure this out, there are two steps: first, we need to do is figure out the *mathematical* relation between the observed moments ($E(X_n)$, $Var(X_n)$, etc.) and the actual moments of the generator. Then, we need to see what

---

$$\sigma = \sqrt{\pi_1(m_1^2 + \sigma_1^2) + \pi_2(m_2^2 + \sigma_2^2) - (\pi_1 m_1 + \pi_2 m_2)^2}\ .$$

[16] See Samorodnitsky and Taqqu(1994). It is interesting that the Pareto-Levy-Mandelbrot distribution is only known by its characteristic function, not its density which cannot be expressed in closed form mathematically, but only as a numerical inversion of the Fourier transform.

epistemology tells us about the significance of these relations to out ability to *know* the actual moments.

THE FIRST AND SECOND CASES.

In the first and second case, the moments of the generator (e.g., $E_1(X)$, $Var_1(X)$, $E_2(X)$, $Var_2(X)$, and higher-level moments) can be quickly inferred from the observation of the actual outcomes.

For example, the observed first moment – the observed mean $E(X_n) = (x_1+x_2+\ldots+x_n)/n$ – quickly converges to the actual mean $E_1(X)$ or $E_2(X)$ as n increases. Same with the observed variance of the sample $\{x_1 \ldots x_n\}$, $Var(X_n)$, converging to $Var_1(X)$ or $Var_2(X)$. The same is also true with higher-level moments.

Let us illustrate this point—the fast convergence of the observed moments to the actual moments—by considering the first moment, or the mean. In the first case (the dice), the outcomes are bounded, so that we know that $\min(X)<x<\max(X)$ for sure. In the second case (the Normal distribution) the outcomes are not bounded, but their probability decreases drastically as they vary from the mean.

That is, $p_i(x)*x \rightarrow 0$ quickly as x increases to extreme values both in the case of the first and the second generator (that is, for i=1,2). In the first case this is due to the fact that $p_1(x)=0$ for $x<\min(X)$ or $x>\max(X)$; in the second, because $p_2(x)$ decreases much faster than the deviation of x from the mean.

This means that the effect of extreme values on the mean of the generator, $E_i(X) = \Sigma_x x*p_i(x)$, is negligible in both the bounded case (i=1) and the Normal case (i=2). That is, $\Sigma_x x*p_i(x) \sim \Sigma_{x \text{ not an extreme value}} x*p_i(x)$ for both generators.

Consider now the data we actually observe. Even if the low-probability extreme values of the generator (if such exist) are *not* observed at all in the outcomes x_1, $x_2 \ldots$ x_n, the "experimental" $E(X_n) = (x_1+x_2+\ldots x_n)/n$ is *still* converging towards $\Sigma_{x \text{ not an extreme value}} x*p_i(x)$. This, as we said, will not differ much from the actual $E_1(X)$ or $E_2(X)$. One does not, in other words, need to wait until a rare extreme event occurs, even if the possibility of such events exists, in order to get a reasonable estimate of the real $E_1(X)$ or $E_2(X)$ from the experimental $E(X_n)$.

For similar reasons, $Var(X_n)$ will converge quickly to $Var_1(X)$ or $Var_2(X)$, and the same for higher-level moments, even if x_1, x_2, … x_n does not include any of the extreme values that could occur – if any.

THE "SEMI-PESSIMISTIC" CASE

Suppose now that the generator which generated our data – outcomes $x_1, x_2, \ldots x_n$ – is of the third type, the "semi-pessimistic" case of a linear combination between a Normal and Poisson distribution.

In this case, the extreme values of the generator are *not* negligible for the calculations of the generator's moment. That is since, while $p_3(x) \rightarrow 0$ as x deviates greatly from the mean, it does not do so "fast enough" to make extreme values negligible. That is, $p_3(x)*x$ does not $\rightarrow 0$ as x becomes extreme.

In such situations, $E_3(X) = \Sigma_x p_3(x)*x \neq \Sigma_{x \text{ not extreme value}} p_3(x)*x$. Therefore, as long as the rare extreme events do not occur, the "experimental" $E(X_n)$ is converging towards $\Sigma_{x \text{ not extreme value}} p_3(x)*x$ - which might be very different from $E_3(X) = \Sigma_x p_3(x)*x$.

In other words, the rare, extreme events need to *actually occur* before $E(X_n)$ will be close to $E_3(X)$ (if then). And similarly for $Var(X_n)$ vs. $Var_3(X)$ and the higher-level moments.

This is seen by the fact that in such generators, the conversion is much slower. (add formula for the convergence in the first moment and second moment).

Furthermore, until extreme "black swan" results actually occur, the observed outcomes of the second (Normal) generator would be *indistinguishable* from the results of the third (Normal + Poisson) generator. We shall consider the implications of this later.

THE "PESSIMISTIC" CASE

In the "pessimistic" case, things can be intractable. It is not that it is takes time for the experimental moments $E(X_n)$, $Var(X_n)$, etc. to converge to the "true" $E_4(X)$, $Var_4(X)$, etc. In this case, these moments simply do not exist. This means, of course, that no amount of observation whatsoever will give us $E(X_n)$, $Var(X_n)$, or higher-level moments that are close to the "true" values of the moments, since no true values exist.

THE PROBLEM OF INDUCTIVE INFERENCE AND ITS RELATION TO THE MATHEMATICAL RELATIONS DISCUSSED ABOVE

So far, we have just *described four generators* and saw the mathematical relation they imply between the value of the estimated moments and the actual moments (if they exist).

We now need to see how these properties affect the original question we considered: namely, *under what circumstances can we use the data of the previous outcomes of the generator to establish the type of the generator and its parameters*, and thus be able *to predict the risk of future outcomes*.

It should be emphasized that while these two problems – the *mathematical relation* between the generator's true moments and the observed moments, on the one hand, and the ability to *predict the future outcomes* of the generator are closely related, they are by no means identical. The first one is a purely mathematical problem. The second is an epistemological problem.

One can never conclude much about the future *solely* from a small specific set of outcomes, our "experimental data". In the modern literature[17], a corpus of knowledge, suggesting availability of background information is always imperative.

For example, one cannot tell, from a million observations of a coin toss *alone*, that the coin has a certain probability of landing "heads" on the next toss. There is nothing "in the data" itself that excludes, for example, the possibility that the coin will land neither "heads" nor "tails" the next time, but will explode like a nuclear bomb. Despite the close *mathematical* relation between the observed and actual moments, unless we have the right "background information", we will not be able to make any *epistemological* conclusion from the data to the future behavior of the generator. The reason such outcomes as "will explode like a nuclear bomb" are excluded is that, in most case, we *have* the right kind of "background information" to exclude it – e.g., our knowledge of physics.

On the other hand, even if the generator is of the "pessimistic" Pareto-Lévy-Mandelbrot type above, the lack of *mathematical* relation between the observed moments and the real moments might not – in theory! – exclude one from making an *epistemological* conclusion about the future outcomes of the generator. If by some miracle, for example, we have an access to an angel that whispers in our ear the next outcome of the generator before it occurs, then part of our "background information" simply *includes* the generator's outcome, and we could tell what the outcomes would be.

However, such cases are usually of a fantastic nature—in most cases we deal with, as seen below, the mathematical information is necessary, but *not* sufficient, to reach the epistemological conclusions we are interested in.

THE IMPLIED BACKGROUND INFORMATION AND OUR CLAIMS

As we said we are interested here in the epistemological problem given a *specific type of background information*, which is the situation in practice when risk managers need to "show their stuff". We assume that the background information is such that:

1. Outcomes are created by some random generator;
2. That this random generator will continue to produce them in the future;

[17] See for example Levi(1980), Kyburg(1974).

3. One does not have any independent way to estimate either the type of generator or its parameters except from the data of the previous outcomes, and that furthermore

4. The generator can be any one of the four different types of exclusive and exhaustive generators discussed above.

The first three assumptions are not controversial (where is this information coming from? Why is it agreed? Add references? E.). The fourth one is

Our epistemological question is: *if* the background information is as above, *what if anything* can we conclude about the moments of the generator (and, hence, about its future behavior) from 1). the observed past behavior of the generator, and 2). This background information? Our practical question is: *when is it* the case that, indeed, the generator can be of all four types, or at least of the "pessimistic" type, type 3 or 4?

We claim that:

1) If the generator *can be* or type 3 or 4 ("semi-pessimistic" or "pessimistic"), that is enough to *invalidate* our ability to conclude much from its past behavior to its future behavior; in particular, it makes it impossible for us to assign *any specific probability* to future outcomes, which makes the situation one of uncertainty, as claimed in the introduction above.

2) It is precisely in situations dealt with by risk managers where the generator *can be* of type 3 or 4.

THE PROBLEM OF INDUCTIVE INFERENCE: THE FIRST PART

Let us begin, then, with the first part of the problem, the "if-then" part: namely, under what circumstances we can (or cannot) say something about the moments of the the generator *if* we know (or do not know) the background something about what the generator is, or what type it could be.

There are two possibilities. It might be that certain information about the moments is a *deductive consequence* of what I already know about it. For example, if I know that a generator's outcomes are bound between the values *a* and *b*, I know that the first moment is also so bound. This is not a matter of choice or decision: to be logically consistent, I *must* accept all such consequences the background information implies about the moments[18].

[18] See also the distinction between "doxastic commitment" and "doxastic performance" in the section about induction and deduction, below.

More complicated is the case of *induction*. Even when (as we always assume) all the deductive consequences of the background information are known, it might be that no specific value for the moments emerges. In that case, we are not forced to settle on a specific value for them. Nevertheless, we might conclude that under the circumstances, we are *inductively justified* in assigning the mean of the generator a certain value (say, "3.5" in the "fair die" case), and similarly for higher moments.

We discuss induction more specifically below, in a separate part. But before we begin this section, a short summary is necessary.

As Peirce showed, this is really a epistemic *decision problem*. I am given background information about the generator ("it looks like a die of some sort is tossed") and the previous outcomes ("the outcomes were 4, 4, 3, 2, 1"). I need to decide whether adding a new conclusion about the generator's moments to my beliefs based on this data is justified (say, "the die is a *fair* die", or more formally "the die's first moment is 3.5").

To solve the decision problem, as in all decisions problems, one needs to consider the *goal* (or goals) one tries to achieve, and the *options* one can choose from. To choose correctly means to choose the option that best achieves one's goals. Decision-making goals can be anything from winning a nuclear war to choosing a good restaurant The goals of inductive inference is (as James showed, below) to *seek new information* while at the same time *avoiding error*. Similarly, the available options can be anything from launching a Trident II missile to driving to the restaurant.. In inductive inference, the options are *adding new claims* to one's beliefs—in this case, claims about the value of a random generator's moments.

These two goals are in tension: the more information I accept, the more likely it is that one will mistakenly include error. The question is, what new claims give me the most information for the least risk if I add them. The result of the inductive inference—the solution of the decision problem—is *adding to one's beliefs the claim that best balances these goals*. Adding this claim is the inductive inference justified under the circumstances.

Note that the null claim—"add no new information"—is always available. If the optimal option is the null option, it means that the justified inductive inference is *no* inference. In our case it would mean that we are not justified in concluding anything about the generator's moments from our background information and past outcomes. As we shall see, this is often the case.

Note, further, that mere high probability, e.g. low risk of error, is *not* itself good enough for acceptance. Consider a lottery with a million tickets: the probability or each ticket winning is 1/1,000,000; but if we accept that this low probability, in itself, is enough to conclude that ticket *n* will *not* win, we reach the absurd conclusion that *no* ticket will win.

In what follows, we need to formalize and quantify the decision situation faced by the agent. For this we use the system developed by Levi. Other formalizations of epistemic decision-making in inquiry exist; in fact, one of the authors (Pilpel) is investigating the differences

between these systems. But in the cases of risk management described below, all of them will recommend the same (pessimistic) conclusion.

TYPE #1 AND #2 GENERATORS

Suppose that an angel came to us and told us the following: "the phenomena which you measured so far, with results $x_1, x_2, \ldots x_n$, is produced by a generator which is bound (type 1 above) between *a* and *b*, or which gives a normal distribution (type 2 above). However, I will not tell you what the mean, variance, or higher moments are; this you need to figure out from that data." Could we do it?

TYPE 1 GENERATORS: FORMAL TREATMENT

To answer, let us put things more formally, using Levi's notation (Levi, 1980, and also below). To simplify things, let us fix *a* and *b* as 1 and 6, and first consider a bounded generator (Type 1) with a finite number of outcomes—say a tossed die with outcomes {1, 2, 3, 4, 5, 6}. John, at time t_0, has to make a decision about the properties of this random generator. What can we say about this situation, epistemically?

THE CORPUS OF KNOWLEDGE: BACKGROUND INFORMATION AND EXPERIMENTAL DATA

First of all, John has a *corpus of knowledge* (or belief), $K_{John,t0}$. It includes the following information:

1) *Background information* John knows about the generator. As the angel said to John, $K_{John,t0}$ includes:

   a) The outcomes of the dice throws are governed by a random generator defined by a probability function X:{1,2,3,4,5,6}→[0,1].

   b) The outcomes of the generators are always one of the set {1,2,3,4,5,6}.

   c) The generator's mean (E(X)), variance (Var(X)), and higher-level moments are fixed, both in the past and in the future.

   d) John knows the laws of statistics, methods of statistical inference, and so on, e.g., the central limit theorem, etc.

2) John's corpus of knowledge $K_{John,t0}$ also includes the *outcomes of the previous trials* up to time t_0:

1) The first toss of the die had outcome $x_1 \in \{1,2,3,4,5,6\}$.

2) The second toss of the die had outcome $x_2 \in \{1,2,3,4,5,6\}$.

3) …

4) ….

n) The nth toss of the die (the last one before time t_0) was $x_n \in \{1,2,3,4,5,6\}$.

We also assume something else of significant importance: that n is large enough for us to use the *normal approximation* for $E(X_n)$. We shall see the importance of this later.

3) The result of 1-n above and John's knowledge of statistics is that, of course, John has *estimates* of the first, second, and higher moments in his corpus:

a) The estimated first moment of X given the first n tosses ($E(X_n)$) is $(\sum_i x_i)/n$. Note that this itself is a random variable, dependant on both the properties of X and on n.

b) The estimated second moment given the first n tosses (Estimated variance, or $Var(X_n)$) is the square of the sample's standard error, or $(\sum_i (x_i - E(X_n))^2)/(n-1)$.

c) … and so on for higher-level moments.

4) Finally, John's corpus of belief includes (by definition, as seen below) *all the deductive consequences* of the above information. In particular, that $1 \leq E(X) \leq 6$, $0 \leq Var(X) \leq 25$ $(=(6-1)^2)$ (actually less, but we can afford to be generous here), etc.

5) However, John's corpus does *not* limit where E(X) can be *deductively* any more than that. It is not logically follow from the outcomes and the background information that E(X) is more specific than [1,6].

John is engaged, at time t_0, in solving a decision problem: given the information above in $K_{John,t0}$, can he give a reliable estimate of the moments of the generator X—and thus, of its future behavior? To simplify, once more, we shall consider only the case of John estimating the first moment, E(X).

THE DECISION PROBLEM: THE OPTIONS

To repeat, giving a reliable estimate of E(X) is another name for saying that John is *justified to infer* that E(X) is of a certain value—that it is a legitimate inductive inference. This is a

decision problem; we need to first consider *what options* for inductive inference exist—that is, between what estimates of E(X) John *can* choose; then, to decide which one (if any) of those John *should* choose.

What are the options available? This depends both on what is deductively excluded by $K_{John,t0}$ and the goals that interest John. In this case, we know that:

1) $K_{John,t0}$ $1 \leq E(X) \leq 6$. Whatever value John chooses as his estimate of E(X), it must be between 1 and 6 on pain of logical inconsistency.

2) From the statistics in $K_{John,t0}$ one knows that the estimate "$E(X)=E(X_n)$" is the only one that is free from an built-in bias.

Now, we *can* limit the number of options John considers accepting or rejecting to a finite number (even to two). For a fixed ε_0, we can consider the two options as whether $|E(X)-E(X_n)|<\varepsilon_0$ or not (H_0). On this view, there are four options altogether: to accept that E(X) is at most ε_0 from the observed $E(X_n)$, to accept that E(X) is ε_0 or more from the observed $E(X_n)$, to accept both (which means that John decides to add information to $K_{John,t0}$ that makes his beliefs inconsistent, by adding $H_0 \wedge \sim H_0$) and to accept neither (that is, to add nothing to $K_{John,t0}$, by "adding" the tautology $H_0 \vee \sim H_0$)

However, there is no need to *a priori* limit the number of possible options. There is a natural set of potential basic options, mutually exclusive and exhaustive (as they must be—see Levi, 1980), that are the most specific possible: namely the set $\{U_x =_{def}$ "$E(X)=x$"$\mid 1 \leq x \leq 6\}$.

In this case, John has a total number of $2^{\aleph}$ options: those that are defined by any sort of (measurable) subset of [1,6]. For example, John might decide that the strongest claim that he accepts is that E(X) is between ½ and 1 or between 4 and 5; that is, John accepts the infinite disjunction $(\bigvee_{0.5<j<1} U_j) \vee (\bigvee_{4<j<5} U_j)$ as true, but does not accept anything more specific. Note that the previous "basic" option H_0 is a non-basic one, the disjunction $\bigvee_{E(Xn)-\varepsilon 0<j<E(Xn)+\varepsilon 0} U_j$.

In particular, John still has the weakest option—accept only the disjunction $\bigvee_{1 \leq j \leq 6} U_j$, that is, that $1 \leq E(X) \leq 6$, which is already in $K_{John,t0}$ and therefore a null addition; and there is a single strongest option—accepting the disjunction $\bigvee$, that is, to accept that *none* of the basic hypotheses U_j are true. This means to accept that $E(X) \notin [1,6]$, in contradiction with information already in $K_{John,t0}$ that it is; that is, the strongest option is to add a contradiction.

THE DECISION PROBLEM: RISK OF ERROR

The next issue to consider in the decision problem is the risk of error by accepting any of the options, and, in particular, the basic options. The *risk of error*, from the agent's point of view, is the *probability that it is wrong*.

Since we are dealing with the infinite case, we must deal not with probability itself (for every basic option, $p(U_j) = p(E(X)=$ *exactly* j) is 0), but with the *density function*, *f*, which in turn determines the probability for any measurable set. Can John estimate this density function? The laws of statistics tell us that John can do so.

The calculations themselves can be found in any statistics textbook. Here is a short sketch: for a "large enough" n (n>30), the random variable $X^*_n =_{def} (\sum_{j=1 \text{ to } n} x_i)/n$ behaves roughly like a normal variable (due to the central limit theorem) with mean E(X) and standard deviation of $\sigma_X/\sqrt{n}$. We do not know what σ_X itself is, of course (the generator's moments are hidden from us) but, since σ_X is bounded from above—if by nothing else, then by $sqr((6-1)^2)=5$, in this case—there is a known upper limit to the standard deviation of X^*_n is for every n. So, for every n, the laws of statistics tell John that he can assume that X_n's density function is roughly that of a normal random variable with a mean E(X) and (maximal) standard deviation of (in this case) $5/(\sqrt{n})$.

(For smaller n, one needs to use Gossett's "Student-t" distribution, but we can assume n is large enough. Also that the normal approximation of X^*_n is unbounded—it can go to, say, -1000 or +1,000,000—while the "real" X^*_n is the observed average of n die tosses, and must be bound between 1 and 6; but, again, for a "large enough" n the tails will be so close to 0 as to make no difference. Finally, one can estimate σ_X by using $s = sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]$, the sample's standard error, which would usually be much smaller than 5; but we can afford to take the "worse case scenario" here.)

What, then, are the *allowable* probability functions, $Q_{John,t0}$, John can consider (for a given n) as possibly representing the actual density function of the probability of the real E(X) being at a certain point around the observed $E(X_n)$? It is a family of normal distributions with mean $E(X_n)$ and maximal variance $5/\sqrt{n}$. So, the density functions are:

Allowable density functions for John at time t_0: $Q_{John,t0} = \{f_v \equiv N(E(X_n),v) | 0<v<5/\sqrt{n}\}$.

Note that the agent can use the laws of statistics to reach conclusions about the probabilities partially because the original random variable X describing the generator does not change wth time. Therefore, the risk of error John takes (given a fixed density function *f* and n) if John accepts the infinite disjunction $(\bigvee_{0.5<j<1} U_j) \vee (\bigvee_{4<j<5} U_j)$ as true (that is, adds it to $K_{John,t0}$) but does not accept anything more specific, is $1-(\int_{[0.5,1]} f(x)dx+\int_{[3,4]} f(x)dx)$, that is, 1-the probability of it being the case that E(X) *is* in that range.

THE DECISION PROBLEM: INFORMATIONAL VALUE

Now we come to *informational value*. What informational value should be assigned to every U_i?

According to it (Levi 1980) the informational value of an hypothesis, Cont(H), is inversely correlated with *a* probability function, M(H): the higher the "probability" of an hypothesis,

the less information it carries. This must be so, if we want certain basic properties of information value to hold: say, that the informational value of a tautology is the minimal possible one, or that the Cont(A∨B) ≤ Cont(A), Cont(B) ≤ Cont(A∧B).

Note that:

1) M is *not* the same as the probability function that the agent assigns to the hypothesis being true, unlike what Popper (1950) and others believed. On the view advocated by Peirce, James, and Levi, informational value is not merely a way to say that something is improbable; probability and informational value are distinct characteristics.

2) The inverse proportion between M(H) and Cont(H) can take several forms: say, $Cont(H)=_{def}1/M(H)$, $Cont(H)=_{def}-(\log(M(H)))$, etc. Levi prefers the simple Cont(H)=1-M(H), for reasons not crucial to this discussion (for the record, in this way his version of information content mimics in certain respects Savage's "degrees of surprise", see Savage (1953), Levi (1980).)

There is here a natural suggestion: that every U_i have an *equal* informational value: it is precisely as informative, or as specific, to say that E(X)_is 0.453 as it is to say that it is 0.991. That means that the M-function, as well, must be "the same" for every U_i. Since we are dealing with the infinite case, any M-function would give probability 0 to every U_i, so we need to look at the density function: we wish the density function *m* of the M-function to be the constant one. In this case, we have *m*≡0.2 over [1,6].

On this view, the informational value of every hypotheses H is 1-M(H), that is, 1-0.2*(H's measure). For example, if John accepts the infinite disjunction $H=(\bigvee_{0.5<j<1}U_j)\vee(\bigvee_{4<j<5}U_j)$ as true (that is, adds it to $K_{John,t0}$) but does not accept anything more specific, John gains informational value of $Cont(H) = 1-M(H) = 1-(\int_{[0.5,1]}0.2dx+\int_{[3,4]}0.2dx)$.

To illustrate, here is a graph of the *m*-function and a few of the potential density functions:

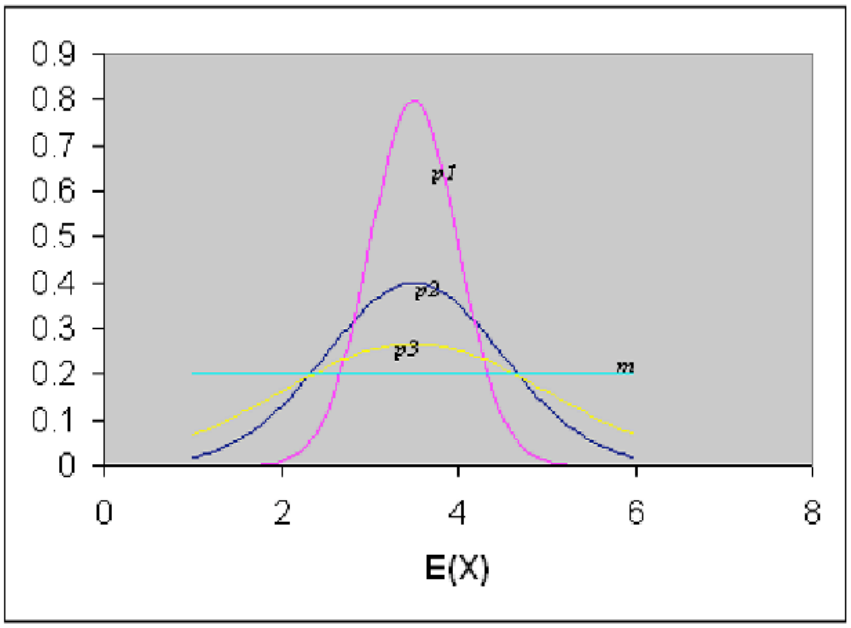


Figure 1: The agent's m-- and p-functions

THE DECISION PROBLEM: THE OPTIMAL INDUCTIVE STRATEGY

THE DECISION PROBLEM: STAGE 1: THE FORMULAS

As always, we follow Levi(1980). Levi recommends to accept an hypothesis where the information value (defined by Cont(H), etc.) is big enough to justify the risk of error (defined by p(H), etc.)

How does one determine what is a "small enough" risk of error or a "large enough" informational value? Levi (1980) concludes that the way to go is as follows:

Rejection Rule: if U_i is a basic option, $p(U_i)$ is the credal probability (e.g., the probability the agent assigns to U_i being true) of U_i, and $M(U_i)$ the probability function determining its informational value $Cont(U_i) =_{def} 1-M(U_i)$, the agent should *reject* U_i (e.g., *add* $\sim U_i$ to their corpus of knowledge) if and only if $p(U_i)<qM(U_i)$, where $0<q<1$ is the agent's "boldness index".

Let us consider this for a moment. To *accept* hypothesis U_i is the same thing as *rejecting* $\sim U_i$; and $Cont(U_i)=1-M(U_i)=M(\sim U_i)$. For a fixed p function and fixed q, the higher the informational value $Cont(U_i)$, the higher $M(\sim U_i)$, and the more likely that $\sim U_i$ will be rejected—that is, U_i accepted. That is, the higher the informational value of U_i, then—*ceteris paribus*—the more likely it is to be accepted.

Similarly, On the other hand, for a fixed Cont(U_i) and q, the higher $p(U_i)$, the lower $p(\sim U_i) = 1-p(U_i)$. This means that it is more likely that $p(\sim U_i)$ will be lower than qM($\sim U_i$); that is, $\sim U_i$ will be rejected, or U_i accepted. The more probable U_i, the more likely it is (*ceteris paribus*) to be accepted.

Now, what is q? This depends on the agent and the situation. For a fixed *p* and M functions, the higher q is, the more options are rejected, and the smaller (and more specific) the number of remaining options. The agent is therefore bolder in accepting the risk of error for information. The lower q is, the less options are rejected, and the larger (and less specific) the number of remaining options.

There is no *a priori* reason to fix q at a specific number. However, as Levi shows, q should never be 0 (let alone below), since this would mean the agent might hesitate and not accept options even if they carry *no* risk of error (e.g., they have probability =1). And q should never be 1 (or above), since that would mean the agent might accept to their beliefs options that carry a risk of error *for sure* (e.g., have probability = 0).

In the infinite case, as in here, one cannot use the probability functions themselves, since for every basic option U_j, $p(U_j) = M(U_j) = 0$, and therefore for every q the inequality does not hold (it is 0<0). The natural extrapolation (see also Levi, 1980, esp. 5.10, 5.11) is, in this case, to consider the density functions: to reject U_j for $1\leq j\leq 6$ if and only if $f(j)<qm(j)$, that is, if and only if $f(j)<0.2q$.

This means that, for a specific q and *f*, there is a "cutoff point ε_0, where $f(E(X_n)-\varepsilon_0) = f(E(X_n)+\varepsilon_0) = 0.2q$; John should *rejects the tails re the value of f is below 0.2q, that is, the agent adds the information that the value of E(X) is between* $E(X_n)-\varepsilon_0$ *to* $E(X_n)+\varepsilon_0$ *to* $K_{John,t0}$.

THE DECISION PROBLEM, STAGE 2: E-ADMISSIBLITY AND SUSPENDING JUDGMENT

Things, however are not that simple, for two reasons: first, John has more than one possible density function, and they do not always give the same recommendation. Second, once it *is* decided by John what he should add to his belief given a specific density function, the question is: which one of those to actually recommend?

An option that is recommended by *a* specific probability function the agent considers legitimate is called by Levi an *E-admissible* option. In this case, the set of E-admissible options for John are:

{Add to $K_{John,t0}$ that $(E(X_n)-\varepsilon(f)\leq E(X)\leq E(X_n)+\varepsilon(f)|$ for every $f\in Q_{John,t0}$, $\varepsilon(f)=_{def}$ distance from $E(X_n)$ where $f(\varepsilon(f))=0.2q$}

It is easy to see that this set is a set of stronger and stronger options, depending on what the variance of the allowable density function is, since the set of density functions is the normal density functions with mean $E(X_n)$ and standard deviation from 0 to $5/\sqrt{n}$, as said above. This

means that if *f* is a "spread out" function (with a relatively high variance), ε(*f*) is relatively large and John only accepts, given that *f*, that the true value of E(X) is between relatively far apart $E(X_n)$-ε(*f*) and $E(X_n)$+ε(*f*). if *f* is a "concentrated" function (with a low variance), ε(*f*) is correspondingly smaller and John accepts a stronger claim—that the real value of E(X) is within a narrower range.

So much for the E-admissible options. Which one to choose? Levi suggests (Levi, 1980) a *rule for ties*:

Rule for ties: If an agent has two E-admissible options E_1 and E_2, and it is reasonable to *suspend judgment* between them (accept $E_1 \vee E_2$)—that is, in particular, that $E_1 \vee E_2$ is itself E-admissible—then one should choose the E-admissible $E_1 \vee E_2$ over either the E-admissible E_1 or the E-admissible E_2.

In this case, all the possible options are arranged by logical strength from the weakest (accept only that E(X) is between $E(X_n)$-ε to $E(X_n)$+ε when ε is when the density function $N(E(X_n,5/\sqrt{n})=0.2q)$ to the strongest (accept that $E(X_n)=E(X)$ exactly; that is, to consider the limit case where the normal distribution has variance 0). Of any two options, one implies the other, so that their disjunction is simply the weaker option. The rule for tie tells us to take the total disjunction—in this case, the *weakest* possibility. So, in sum, John accepts that:

***John's Acceptance, stage 1*:** Adds to $K_{John,t0}$ the fact that $E(X_n)$ is between $E(X_n)$-ε and $E(X_n)$+ε when ε is when the density function $N(E(X_n,5/\sqrt{n})=0.2q$.

THE DECISION PROBLEM, STAGE 3: ITERATION

However, we are still not done. Now that John accepted certain claims to be true, says Levi, John needs to *iterate* the inductive inference. John's new K, $K_{John,t1}$, is the deductive closure of $K_{John,t0}$ and the disjunction $\vee_{x|E(Xn)-\varepsilon 0 \leq x \leq E(Xn)+\varepsilon 0}$("E(X)=*x*"). Or, in Levi's symbolism, John *expanded* his corpus to a larger one, holding more beliefs. In Levi's symbolism, if $H_1 =_{def} \vee_{x|E(Xn)-\varepsilon 0 \leq x \leq E(Xn)+\varepsilon 0}$("E(X)=*x*"):

$K_{John,t1} = K_{John,t0}{}^{+}{}_{H1}$.

John's probability functions, in $Q_{John,t0}$, also change: also change: they are now the set of *conditional* probabilities, given that John added the disjunction that E(X) is between $E(X_n)$-ε and $E(X_n)$+ε to his beliefs. (Levi calls this the *conditionalization commitment*. See Levi, 1980.) That is, John's new probability functions at time t_1, is:

$Q_{John,t1}$ = {p| p(x) = q(x|H1), for every $q \in Q_{John,t0}$}

The informational value function also changes. It becomes determined by the conditional, new M-function, which is 0 outside [$E(X_n)-\varepsilon_0$, $E(X_n)+\varepsilon_0$] and $1/2\varepsilon_0$ inside this interval.

John now has a *stage 2 decision problem*: which, if any, of the options U_x= "E(X)=x", for $x \in [E(X_n)-\varepsilon_0, E(X_n)+\varepsilon_0]$, with these new probability and content functions, should he reject?

As before, one does the calculations and sees that one should reject just those U_x's where the weakest conditional density function, $N(E(X_n), 5/\sqrt{n})$ given that x is between $E(X_n)-\varepsilon$ and $E(X_n)+\varepsilon$, is below qm(x)—that is, $q(1/2\varepsilon_0)$.

Possibly some more hypotheses will get rejected. If there are some, then John needs to yet again add more information to his beliefs—add to $K_{John,t1}$ the fact that E(X) is *not* farther away from $E(X_n)$ thant some ε', $(0<\varepsilon'<\varepsilon)$. Then, John needs once more iterate—conditionalize $Q_{John,t2}$ based on $Q_{John,t1}$ given the new rejections, make M defined by the new $m \equiv 1/2\varepsilon'$, and so on.

This process continues indefinitely. John solves a series of decision problems given $K_{John,t0}$, $K_{John,t1}$, $K_{John,t2}$, … each saying that E(X) is at most ε, ε', ε'', ε''' … away from $E(X_n)$, with the conditional $Q_{John,t0}$, $Q_{John,t1}$, $Q_{John,t2}$, …, each based on the previous one and the new information added, with the new m density function being 1/5 (the original one), 1/2ε, 1/2ε', 1/2ε'', 1/2ε''' …, etc.

It can be shown that eventuall—and perhaps even the first time—John will reach a certain $K_{John,t*}$, $Q_{John,t*}$, with the strongest claim in $K_{John,t*}$ being that E(X) is at most $0<\varepsilon^*$ away from $E(X_n)$, m being $1/2\varepsilon^*$, where *the recommendation is not to reject any more hypotheses*. John, as it were, rejected all the he could reasonably reject.

JOHN'S FINAL DECISION

The final recommendation—the strongest—is:

***John's Acceptance, stage 1*:** Adds to $K_{John,t0}$ the fact that $E(X_n)$ is between $E(X_n)-\varepsilon^*$ and $E(X_n)+\varepsilon^*$ when $0<\varepsilon^* \leq \varepsilon$, ε being the value where John's original density function, the (unconditional) $N(E(X_n),5/\sqrt{n})=0.2q$.

DISCUSSION

The result of the inductive decision problem is "John's acceptance", above. That is, induction recommends that John, in this situation, and for a given n and q, accept that E(X) is in the range described by "John's Final Decision".

In practice, this means two things:

1) Unless q is very small, then for any n that is not too small (say, n≈30 or so, or higher, as we assume) the range that John accepts as possible value for E(X) is relatively small, even if one uses the maximum possible estimation of X^*_n's standard deviation, that is, $5/\sqrt{n}$.

2) If one uses the standard estimation of X^*_n's standard deviation (the standard error), then ε, even after only one iteration, will be even smaller, since the weakest (most spread out) density function John considers in the first case will be $N(E(X_n), s/\sqrt{n})$, with s the standard error, not $N(EX_n)$, $5/\sqrt{n}$), and $s<5$—and thus $N(E(X_n), s/\sqrt{n})$ would reach 0.2q faster (closer to $E(X_n)$.

3) Successive iterations of the decision problem might lead the agent to reject even more hypotheses, eventually settling on the claim that E(X) is in $[E(X_n)-\varepsilon^*, E(X_n)+\varepsilon^*]$, with $0<\varepsilon^*\leq\varepsilon$.

(2) and (3), in this case, are almost unnecessary, however. For a reasonably large n—one large enough to use the normal approximation for X^*_n—even doing *only one iteration* of the decision process and using the *maximal possible size* of X^*_n's standard deviation would usually significantly limit what is accepted.

In short, So when one has a type 1 generator, John *can* tell, pretty quickly, quite a bit about the value of the generator's moment, E(X). John is *justified in inductively accepting* that it is within a range, ε, that is small to begin with in most circumstances (as 1 above says) and gets smaller quickly as the number of observations increases.

Note, also, an important point. First, obviously information about the previous outcomes of the generator is essential for the agent to reach the conclusion. But the law of statistics could only be used *because* we have background information about the type of generator we have here—a "well-behaved" one.

TYPE 2 GENERATORS: NORMAL DISTRIBUTION

BACKGROUND INFORMATION

Type 2 generators are similar to type 1 generators, as we shall see, with a few difference. Again, to fix the discussion, let us presume that the generator is normal, with (actual) moments E(X), Var(X), etc. As before, let us assume that John is trying to estimate the first moment, or what E(X) is.

The background information is very similar to the one with the case of the bounded distribution, of course with the change that John knows that the generator is normal, not bounded. In particular, John knows that E(X) and Var(X) are fixed and will remain so in the future, and the laws of statistics. John also knows, due to these laws, that the (same) estimates ($E(X_n)$, $Var(X_n)$) are the only ones of the generator's moments that do not have a built-in bias.

When it comes to the data, John knows what the past outcomes (x_1, … x_n) of the generator were. As before, let us consider the first moment E(X) and John's estimation of it.

DIFFERENCES FROM BOUNDED DISTRIBUTION—AND WHY IT DOESN'T MATTER IN THIS CASE

There are two things that can ruin it for John. In the bounded case, there were *no extreme events*, first, and *σ_X was bounded from above by a known quantity*. In the normal case, it *could* be that an extreme event would be observed in x_1, … x_n, and significantly "throw off" $E(X_n)$. Or, if σ_X is extremely large, it might take a very large n to get $E(X_n)$ to converge to E(X). In both cases, even for a large n, $E(X_n)$ could still be significantly different from E(X).

Consider, however, what we are trying to achieve in the first place. We are *not* claiming that *all* "well behaved" generators—e.g., all normal distributions—can be easily "worked on" in practice, no matter what their properties or what the outcomes in the past happened to be. If the normal distribution has a very large variance, it will indeed take a lot of time for that pattern to emerge. If an extreme 10-σ event dis in fact occur, the estimate $E(X_n)$ will be off from E(X) for a while.

But such occurrences are *observable*: John will see them occurring in the outcomes, and know to be care in reaching conclusions about the future. Our problem is not with the "bad" generators (large σ_X) or "bad" outcomes (10-σ events) that wear their "badness" on their sleeves, that is, in the outcomes already observed. We are concerned here with exactly the opposite: what we can say about a generator when it is assumed that the outcomes *do* look good—that is, when σ_X is small and no extreme events occured in the past.

So we can assume that the outcomes *do* "look good": that σ_X is relatively small and that no 10-σ events occurred. We want to know: *given these outcomes, what can the agent deduce, if anything, about the qualities of the generator*? In this case, quite a lot.

As above, the random variable X^*_n behaves normally, with random variable $\bar{X}^*_n =_{def} (\sum_{j=1 \text{ to } n} x_i)/n$ behaves roughly like a normal variable (due to the central limit theorem) with mean E(X) and standard deviation of $\sigma_X/\sqrt{n}$. We do not know what σ_X itself is, of course (the generator's moments are hidden from us) but we can estimate it, since one can estimate σ_X by using $s = sqr[(\sum_{j=1 \text{ to } n}(x_j - E(X_n))^2/(n-1)]$, the sample's standard error. This means that we can assume (presuming, again, that n is "large enough" as above) that the density function of X^*_n

is $N(E(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})$. This is a Normal distribution that, as n increases, becomes more and more "centralized" since its $\sigma \rightarrow 0$ as fast as $1/\sqrt{n}$.

In this case, then, John has *one* probability function that determines how probable it is that the actual E(X) is within a certain range of the observed $E(X_n)$:

John's density function, f: $N(E_n(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})$.

(Of course, we could have used this technique to minimize the set of allowable probability functions in the bounded case, as well. But we deliberately did not, to show that even if we *do* allow many probability functions that create a "worst-case scenario" in the bounded situation, John can *still* tell us much about the generator's moments. It also gave us a way to illustrate the rule for ties and E-admissability and to the iteration process, which will be important later on.)

INFORMATIONAL VALUE: SOME COMPLICATIONS

Assigning an M-function (and therefore an informational value function) is a bit more complicated this time. M's density function cannot be the constant function *m* whose integral over the possible range—$(-\infty, +\infty)$—is equally to 1, since there is no such function (the integral is 0 for $m \equiv 0$ and diverges otherwise). There is, simply put, no way for an agent to assign "equal informational value" to "E(X)=*x*" for every $x \in$ and still have the informational value be based on a probability function.

What, then, should M be? There are several possibilities. The one we use—due to our concern with "extreme events"—is as follows. Consider some large L_0, and the range $[E(X_n)-L_0, E(X_n)+L_0]$. There is an infinite number of hypotheses of the value of E(X) within this range (namely, $U_x =_{def}$ "E(X)=*x*" for every $x \in [E(X_n)-L_0, E(X_n)+L_0]$, and two additional hypothesis: $U^- =_{def}$ "$E(X)<E(X_n)-L_0$", and $U^+ =_{def}$ "$E(X_n)+L_0<E(X)$". The M-function that determines the content function will give both of these hytpotheses some the hypothesis U^* some probability, p^- and p^+; we can assume they are the same, p_0.

We can be careful and assume that, first, L_0 is large (relative to the standard error of the sample, s)—say, 10s in length; the reason is that we want these hypotheses to represent extreme possible values of E(X). We also assume that U^- and U^+ are very informative—that is, that p_0 is very small. Within $[E(X_n)-L_0, E(X_n)+L_0]$, we assume that M is determined by the usual, fixed density function *m*; only this time its integral of the $2L_0$ integral isn't 1, but $1-sp_0$. So John's M-function is defined as:

$M(U^+) = M(U^-) = p_0$; $m \equiv (1-2p_0)/2L_0$ over the range $[E(X_n)-L_0, E(X_n)+L_0]$.

THE DECISION PROBLEM

As usual, the agent should *reject an hypothesis U if and only if p(U)<qM(U)*—or, in the case of point hypotheses, use the density functions of *p* and M, repsectively: *reject the hypothesis U if and only if f(U)<qm(U).* On this view, we have:

1) Reject U^- if and only if $p(U^-)<qM(U^-)$: reject U^- if and only if $\int_{-\infty \text{ to } E(Xn)-L0}[N(E_n(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})]dz < qp_0$.

2) Reject U^+ if and only if $p(U^+)<qM(U^+)$: reject U^+ if and only if $\int_{E(Xn)+L0 \text{ to } +\infty}[N(E_n(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})]dz < qp_0$

3) Reject U_x for $x \in [E(X_n)-L_0, E(X_n)+L_0]$ if and only if $f(U_x)<qm(U_x)$, that is, if and only if the value of the normal curve, $N(E_n(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})]<q(1-2p_0)/2L_0$.

Let us consider the possibilities. Suppose as above that L_0 is large and that p_0 is small. Nevertheless, unless p_0 or q are very small indeed, the $\int_{-\infty \text{ to } E(Xn)-L0}[N(E_n(X), sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]/\sqrt{n})]dz$ is going to be far smaller than p_0q, since it is the "tail end" of a normal distribution that is many standard deviations away from the mean. So both U^- and U^+ will be rejected.

Now consider the middle case (3). What we have here is precisely the same situation as in the "bounded" case—with the small difference that the *m*-function is somewhat smaller than the *m*-function in the bounded case over the same range, since $m \equiv (1-sp_0)/2L_0$ and not simply $1/2L_0$, for *m* must account for the possibility of U^- and U^+.

We know how to solve this problem. In fact, it is even easier, since we have a fixed probability function and not a set of such functions. Following the exact same steps as in the bounded case, we get that, after the first iteration:

John's first step: John should reject U^-, reject U^+, and those hypotheses "E(X)=*x*" in the range $[E(X_n)-L_0, E(X_n)+L_0]$ such that $f(x)<qm(x)$, or $N(E_n(X), s/\sqrt{n})<q(1-2p_0)/2L_0$, when s is the standard error, that is, $sqr[(\sum_{j=1 \text{ to } n}(x_j-E(X_n))^2/(n-1)]$. In other words, John should *accept into* $K_{John, t0}$ *the claim that* $E(X) \in [E(X_n)-\varepsilon, E(X_n)+\varepsilon]$, *when* ε *is where the density function* $N(E_n(X), s/\sqrt{n})=q(1-2p_0)/2L_0$.

As before, even in this first step, if n is large enough for X^*_n to use the normal approximation in the first place, ε will be small. And, in addition, for the same reasons as above, it might be that further iterations will allow John to reject even more hypotheses, and accept:

John's Final Inductive Conclusion: John should accept that $E(X) \in [E(X_n)-\varepsilon^*, E(X_n)+\varepsilon^*]$, *when* $0<\varepsilon^* \leq \varepsilon$, *ε being the value where the (original) density function of the probabilities,* $N(E_n(X), s/\sqrt{n}) = q(1-2p_0)/2L_0$.

We see, then, that *even if the generator is unbounded, John can usually justifiably conclude that its E(X) is within a narrow range, as long as the number of observations is large enough to apply the usual laws of statistics (e.g., the assumption that* X^*_n *is normal).* The mere fact that the generator's moment E(X) *could* be any value, including a very large one, does *not* require John to take that possibility seriously. And the same, as before, holds *mutatis mutandis* for higher-level moments of the generator.

TYPE #3 GENERATORS – PART 1.

The problem is that in most cases, the agent does *not* know that the generator is of type I or type II. The agent so *assumes*, but for no better reason than the fact that it is easy to reach seemingly "exact" results with such an assumption.

Suppose, for example, that so far the daily change in a stock's price have been limited to the range between 0 and 10 points. Is there any reason to suspect that it will not move 1000 points one way or the other in the future? If we *knew* the generator that was producing the stock's movements was normal, perhaps. But often we do not know it.

Suppose that an angel told us: "the phenomena you are observing is generated by a generator of type #3. It is a combination of a "regular" Normal distribution and a Normal distribution that gives us very large results with very low probabilities. I will not tell you what the mean, variance, or other moments of this generator are, however. You will have to figure them out from the data." What could we say about the mean, variance, and higher moments of this generator by looking at the data? Very little indeed – at least as long as no catastrophic "black swan" event *had* in fact occurred.

The reason is that in the case of such a distribution, most of the value of the moments comes from the rare and improbable "black swan" events that are due to the extreme Normal distribution, and not the regular and non-catastrophic events that are due to the Normal distribution. As long as no such catastrophic events occurs, we only know a "negative" point: that the observed moments $E(X_n)$, $Var(X_n)$, etc. are <u>not</u> close to the actual moments E(X), Var(X), etc. But that is all we know, *no matter how much (non-catastrophic) data we have*. We cannot say anything about what the size of the difference is until we actually observe such catastrophic events.

Let us put this in more formal epistemic form. Again, let us presume that John wishes to evaluate what E(X) is. And, once more, consider what John knows.

From the background information, John knows what the outcome of the generator so far has been. John also knows the laws of statistics. Furthermore, John knows that the generator is of the form X = (1-*p*)X'+*p*X", where E(X')<<E(X") and *p*<<1. But John does *not* know what *p* is, or what E(X'), E(X") are.

In addition, John knows that *no extreme events occurred*. That is, John knows that all the outcomes so far have been from X'. Can John estimate E(X)?

The answer is negative. To estimate E(X), the agent needs to do two things: 1) estimate *p*, given that no events from X" occurred, and 2) estimate E(X"). While *p can* be estimated, in fact, the fact that we have no information about about E(X") precludes more deliberate information.

How does John estimate *p*? Let us ignore the values of the outcomes and consider a simplification: the outcome is either due to generator X' (with probability 1-*p*) or due to generator X" (with probability *p*). To help out John, and simplify the calculation we will assume that he *knows* (by psychic means, perhaps) whether an outcome is from X' or X". The question is: what is *p*?

STEP 1: EVALUATING *p*

John, here, has an obvious set of options (*p* from 0 to 1), with an obvious M-function (namely, *m*≡1). John has a set of outcomes of length *n* which we know produces the *p* event exactly 0 times. For every *p*, this means that the probability of this occurring is $(1-p)^n$.

Now, when do we reject an hypothesis? We reject the hypothesis U_x ("*p = x"*) if and only if $q(U_x)<qM(U_x)$, or, in this case, $(1-x)^n<q$; that is, John will fail to reject only such x's such that $(1-x)^n \geq q$, or that $1-x \geq q^{1/n}$, or $-x \geq q^{1/n}-1$, or $x \leq 1-q^{1/n}$. That is, John accepts that the real *p* is in the range $(0,1-q^{1/n}]$; as n increases, and $q^{1/n} \rightarrow 1$ (since 0<q<1), this range becomes smaller and smaller.

STEP 2: EVALUATING E(X")

So far so good. However, *John has no information at all about E(X"), and therefore cannot limit E(X) in any way, even with this information about p.*

The problem is this. Consider evaluating *E(X")* given the outcomes, $E(X_n)$—or, more precisely, $E(X_n')$. First, what are the options John has? John is interested in is as before. John is interested in *whether or not the real E(X) (=(1-p)E(X')+pE(X)) is close, or not close, to the observed* $E(X_n)$ $(=E(X'_n))$. This means that John can use the same options as before: namely,

for a given L_0 which is large in relation to the standard error of the sample, John has U^- = "$E(X'')<E(X_n)-L_0$"; U^+ = "$E(X'')>E(X_n)+L_0$", and U_x = "$E(X'') = x$" for $E(X_n)-L_0 \leq x \leq E(X_n)+L_0$.

John is interested in is as before. John is interested in *whether or not the real E(X") is close, or not close, to the observed E(X_n).* This means that John can use the same options as before: namely, for a given L_0 which is large in relation to the standard error of the sample, John has U^- = "$E(X)<E(X_n)-L_0$"; U^+ = have $M(U^+) = M(U^-) = p_0$; $m \equiv (1-2p_0)/2L_0$ over the range $[E(X_n)-L_0, E(X_n)+L_0]$.

Consider, however, what the allowable probability functions about E(X) being in any range are. But John has *no* data at all—no observations—about X", only about X'. So there is no way to evaluate E(X"). To put it differently, since there are no observations, *any* probability density function from $-\infty$ to $+\infty$ is in John's $Q_{John,t0}$. This, of course, is always the case when one has literally no observations of the parameter.

Consider now the situation. For any given density function *f*, U_x in $[E(X_n)-L_0, E(X_n)+L_0]$ will be rejected if and only if the density function $f(x)<q(1-2p_0)/2L_0$; for U^- and U^+, if and only if $\int_{-\infty \text{ to } E(Xn)-L0} f(z)dz < qp_0$ or $\int_{E(Xn)+L0 \text{ to } +\infty} f(z)dz < qp_0$, respectively.

But since *all* probability functions, all *f*'s, that is, are allowed, for *every one* of the hypotheses, U^- and U^+ included, there are some probability functions that recommend rejecting it and some that recommend accepting it. In particular, there is always some probability functions (for example, *f*≡the M-function itself!) that will recommend rejecting *no* hypothesis.

What to do? We can use Levi's rule of ties. Since *every* possible strategy from rejecting no hypothesis to rejecting all but one (it is impossible to reject all of them, as seen above, since that means adding an inconsistency to $K_{John,t0}$, which is never recommended, see Levi, 1980 about "deliberate inductive inference", Ch. 5), that is, they are all E-admissible, the rule of ties recommends using the disjunction of all of them—the hypothesis "reject nothing"—as long as it is "reasonable" (e.g., itself at least E-admissible.) This is the case, as we've just seen.

Finally, there is the case of iteration. But in this case, since nothing is rejected, there is no iteration—the first action ("add nothing") is the final one that is recommended to John. There is no reason to conditionalize the probability functions or M, since nothing is added to $K_{John,t0}$ in the first place.

So the recommended strategy is:

John's Recommended Inductive Inference for E(X"): Remain in complete suspense about E(X"); accept nothing stronger than "$E(X'') \in$ ".

STEP 3: EVALUATING E(X) = (1-p)E(X')+pE(X")

Now John is finally ready to evalute E(X) itself. Could, perhaps, the fact that at least *p* can be bounded by the agent be of use? The answer is negative. For if there is no information at all about E(X"), then there is similarly no information about (1-*p*)E(X')+*p*E(X").

The reason is that the evaluating of E(X") is undounded—it can be anything as far as John is concerned—so that the fact that it is multiplied by a small *p* is of no consequence. John cannot exclude the possibility that E(X")=1,000,000*p*, or $10^{100}p$, for that matter.

To put it somewhat more formally, consider any probability function *g* which supposedly gives us the definition of how E(X) = (1-*p*)E(X')+*p*E(X") is distributed around . It is easy to find some other probability function, *g*", such that if *g*" is the distribution of E(X") in , then *g* is that of E(X). The fact that *p* is small doesn't mean that E(X) must be small; if *g* (say) says that the likelihood of the average of E(X) is distributed around 1,000,000, just choose a *g*" where the likelihood is that E(X") is distributed around 1,000,000/*p*.

So John's possible functions for the likelihood of E(X) being anywhere in is still *all of the possib le probability functions*. And for the same reasons as above:

***John's Recommended Inductive Inference for E(X)*:** Remain in complete suspense about E(X); accept nothing stronger than "E(X)∈ ".

In conclusion: even if we *know* that a certain generator is a type 3 distribution, before a catastrophic event occurs we cannot say anything about the difference between the observed $E(X_n)$ and E(X), the observed $Var(X_n)$ and Var(X), or any other observed moment and the "real" one. Before such an event occurs, extrapolating from past data to future behavior of such a system is *worthless*.

Here we see that the mathematical information is *necessary* for reaching the epistemological conclusion. To conclude that the future is like the past we must know that the mathematical equality $E(X)\sim E(X_n)$ (and the same with other moments) will hold. If we know that this mathematical relations does not hold, then naturally we cannot make any epistemological conclusion about the future based on the past in that case.

TYPE #4 GENERATORS – PART 1

Things are even worse with type 4 generators, for obvious reasons. If an angel tells us that a certain generator is a type 4 one (Pareto-Lévy-Mandelbrot), we know that no relation between the observed moments $E(X_n)$, $Var(X_n)$, etc. and the "real" moments of the generator exist – for the very good reason that there are *no such moments*.

TYPE #3 AND #4 GENERATORS – PART 2

But things are even worse than that. We have just seen that if we know that the generator is of type 1 or type 2, we can rely on the observed moments to be close to the "real" moments. We also showed that if we know that the generator is of type 3 or type 4, the observed moments (at least before a catastrophic "black swan" event occurs) are worthless in finding the values of the real moments.

But all these scenarios assume that we know what type the generator is. Suppose we *don't* know what it is, and want to see if the data helps us figure this out? In that case, the mathematical equality between the observed and actual moments, *even if it holds* (even if the generator, that is, is in fact of type #1 or #2), might not be enough to reach any epistemological conclusions about the similarity of the past to the future. The mathematical equality is necessary, but not sufficient.

Consider the following situation. Suppose an angel tells you that a certain generator is either type 2 (Normal) *or* type 3 distribution (a mixed combination of Normal and Poisson). Consider the data $x_1, x_2, \ldots x_n$. As long as no catastrophic "Poisson event" had actually occurred, the data would be *indistinguishable* between type 2 and type 3 generators, since all the outcomes of the type 3 generator would still be due to the "Normal" part of its distribution. We will not be able to tell due to anything in the data whether it is one or the other.

More generally, suppose that an angel tells us that a certain outcome *might* be due to a generator of type 3 or 4, as well as a type 1 or 2 generators. Does any amount of data tell us anything about whether or not this is true, before a "black swan" event happens? No, since until a low-probability catastrophe actually occurs, *if* the generator is in fact of type 3 or 4, the data would look *indistinguishable* from that of a generator of type 1 or 2, as we've just seen.

So if we *don't know* that the generator is *not* type 3 or 4, then our data is *just as worthless* in assessing the future behavior of the generator as if we knew that it is type 3 or 4. This is not because $E(X_n)$, $Var(X_n)$ and so on *must* be far from the "real" E(X), Var(X), etc. (if they exist), but because we can never tell from the data *whether* they are or not before a catastrophe happens. And if we don't know the moments, *ipso facto* we don't know anything about the probabilities of the generator's outcomes, which depend for their calculation on these moments. We cannot tell anything about the risk of any future outcome. We are in a situation of *decision making under uncertainty*.

In summary, for the epistemic inductive inference from the past outcomes to the future ones to be worthless, we need not know that the generator is of the "dangerous" type: it need not be the case that $E(X) \neq E(X_n)$ (or the same for the other moments). It is enough *not* to know that it is not of that type. In such a situation, a "black swan" could surprise it at any moment – and we wouldn't be able to tell whether it would happen or not until after the fact. The mathematical equality $E(X)=E(X_n)$ is of no use to us if we cannot *know in advance* that it holds before a catastrophic event occurs.

COULD SUCH GENERATORS EXIST?

This entire discussion would have remained completely theoretical if it was not the case that the situations risk managers deal with *could* involve the "bad" types of the generators – that is, unless epistemic assumption #4 above holds.

We have seen above that many economists dismiss the possibility of assumption #4. we claim that, unfortunately, in economical situations generators of this type *can* occur. Physical systems (as Mandelbrot says—add references) must be of the "benign" type – type 1 or 2, or, more specifically, of type 1 (a "bounded" generator). The laws of physics bound their values – specifically, the amount of energy in the system, the entropy of the system, and other such physical characteristics cannot move beyond a certain range (add other references except for Mandelbrot, e.g., his sources.).

Social systems, as well, are bounded. If nothing else, there is a lower bound for the "worse possible outcome" – namely, death. This is not because nothing can be worse from the individual's point of view than his or her own death, but because one can (almost?) always avoid such circumstances by choosing suicide instead. (Is this the case??? Perhaps erase this??? What about "infinite badness" like Hobbes believed???)

In physical and social systems, therefore, it is often the case that we can tell in advance, due to external, purely deductive reasons, that the "generator" must be bounded and therefore (relatively) benign; we can therefore use the past data for inductive inference about the future, as we seen above.

In many *financial* systems, however, this is not the case (references?). There are potential events in many such systems that would cause losses (or gains) that are, in theory, *unbounded*. To convince oneself of this, one need only look at a simple "option": the possibility exists of losing an infinite amount of money combined with the fact that such probability may remain unknown by us. (References.)

This is not to say, of course, that death is somehow "better" than losing a lot of money, or that gaining or losing an infinite (or very, very, large) amount of money is physically possible. The point is, rather, that in the case of a physical system one knows that one can describe the system with a bounded (or, at worse, a compact-supported) generator, while if we look at a financial system this cannot be promised. (Remove this paragraph, perhaps? Or give more references?)

THE RECOMMENDED STRATEGY IN SUCH SITUATIONS, AND "LONG-TERM CAPITAL" REVISITED

The conclusion of this epistemological excursion is as follows: in such situations, we are in an essentially "uncertain" situation.

If we must make decisions in such a situation, our best bet is to use a strategy suited to "uncertainty". Minmax (or similar strategies) will not work, because of unboundedness. (references to the strategies of uncertainty—perhaps again?) "Forcing" oneself to use a specific probability value will lead to grief: it is useless to protect oneself against the risk of a certain outcomes when you really have no reason to give it *any* specific probability.

Note in particular that the well-known device of taking "safety margins" will not work. Suppose that one is willing to take a one-in-a-million risk of bankruptcy, but – in order to "hedge" one's bets – only makes trades that (according to his or her calculations) have a one-in-a-trillion chance of going so badly as to lead into bankruptcy. Will taking such a ludicrous "safety margin" – a factor of 1,000,000 – help the risk manager avoid bankruptcy in such situations?

The answer is no. Taking such "safety measures" is a reasonable device if one *knows* that the generator if of one of the "benign" types, e.g. type 1 or 2, and therefore one *knows* that one *is* justified in making assumptions about the probabilities of events happening in the future using the observed parameters as approximations for the actual parameters of the generator, but might not be completely sure about the *exact values* the parameters should have. In other work, this would work in cases where one knows one can safely describe the situation as one of decision making under risk, although one is not sure exactly what risk.

In a situation where the generator might be of type 3 or 4, however, one doesn't simply have a *vague* idea of what the risk is; one has *no* idea what it is, and cannot assign *any* value to it. Taking only "trillion-to-1" bets against bankruptcy is worthless in such a situation since the assessment of the risk of a certain trade *as* trillion-to-1 is worthless in the first place. There is no 'there' there: the calculated "million to one safety margin" doesn't correspond to anything in reality. (Add something or is this ehough?)

We have no real base to give credence to this estimation; the relaxing number "a trillion to 1" has only psychological significance in such a situation – as the occurrence of the "impossible" 10-σ event in the case of "Long Term Capital" shows. It is not as if a 10-σ event actually occurred. Rather, the belief that it *is* a 10-σ event was based on the *unjustified conclusion* that the generator involved is of the benign type in the first place.

Therefore, the risk managers did not consider the possibility of the generator being of the third or fourth type, where events that *would be* 10-σ events *if* the generator were of the benign type, actually occur far more frequently.

Our only recourse in such situations is Popper's solution: to wait for the "black swan", and make sure that we are not destroyed by it. (Add more about Popper here—the falsification

requirement. I am not sure that this is really our "only recourse". Again, look at strategies under uncertainty for detail—P.)

SUMMARY

In this chapter, we have tried to show the essential problem of risk management is forcing situations of decision making under uncertainty into the straightjacket of decision making under risk.

We showed this in a few steps: :

First, we showed that certain random generators have a "bad" relation between their observed moments and their actual moments. This is a purely mathematical issue.

Second, we have shown if one's background information satisfies certain conditions, then *if* such generators are not ruled out, the mere possibility that they are the generator one is dealing with sabotages any attempt to assign specific values to the "real" moments of the generator, due to the "black swan" problem – the possibility of rare extreme events which have a large influence on the moments. This is an epistemological issue.

Third, this forces us This forces us to conclude we are in a situation of *decision making under uncertainty*. This is a decision-theoretic matter.

Fourth, we showed that, in fact, the situation risk managers deal with are precisely those where such generators cannot be ruled out. This is a scientific issue: it has to do with the different nature of physical and economic systems.

Fifth closely related to the third issue, we showed that common "avoidance" procedures – taking only what seems like "very low" risks – will not work, since the implicitly assume the situation is one of decision making under risk in the first place. Even "usually" procedures for decisiom making under uncertainty – minmax, minamx regret, etc. – will not work, since the "bad" generators are not bound.

Finally, we show that in such situation, the only thing we can do is protect ourselves against the black swan –and recognize that we may not know much about it. This is the (type of strategy) strategy, which is applicable to this sigtuation.

REFERENCES

Add References, of course.

Bernoulli, D. (1738), Specimen Theoriae novae de Mensura Sortis, in Commentarii Academiae Scientiarum Imperialis Petropolitanae 5 , St. Petersburg, 175-192.

Cramer, G. (1738), letter to Nikolaus Bernoulli, published by Bernoulli.

De Finetti, Bruno (1937), La Prevision: ses lois logiques, ses sources subjectives. In Annales de l'institut Henri Poincare, 7, 1-68.

Feller, W. (1971), *An Introduction to Probability Theory and Its Application* (Vol. II; 2nd ed.), New York: Wiley.

Keynes, John Maynard (1937), *The General Theory*. In *Quarterly Journal of Economics*, Vol. LI, 209-233.

Knight, Frank (1965, 1921), *Risk, Uncertainty and Profit*, Harper Torchbook Edition, New York: Harper and Row.

Levi, Isaac (1980), *The Enterprise of Knowledge,* Cambridge, MA: MIT Press.

Luce, R. Duncan and Raiffa, Howard, (1957), *Games and Decisions: Introduction and Critical Survey*. New York: John Wiley and Sons.

Mandelbrot Benoit B (1963) The variation of certain speculative prices. *J. Business* 36 394–419 (Reprinted Mandelbrot, 1997)

Mandelbrot Benoit B (1982) *The Fractal Geometry of Nature*. San Francisco, CA: Freeman)

Mandelbrot Benoit B (1997) *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (Berlin: Springer).

Mandelbrot Benoit B (1999) Multifractals and 1/f Noise: Wild Self-Affinity in Physics. Berlin: Springer.

Markowitz, Harry (1952), Portfolio Selection, *Journal of Finance* 7: 77-91.

Ramsey, Frank P.(1931), *Truth and Probability*. In *The Foundations of Mathematics and Other Logical Essays*. R. B. Braithwaite (ed.) London: Routledge.

Samuelson, P.A. (1983,1947), *Foundations of Economic Analysis,* Cambridge, MA: Harvard University Press.

Samorodnitsky, Gennady, and Murad S. Taqqu (1974), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman & Hull.

Savage, Leonard J.(1954), *The Foundations of Statistics*. New York: John Wiley and Sons.

Shackle, G. L. S. (1955), *Uncertainty in Economics and Other Reflections*. Cambridge: Cambridge University Press.

Taleb, Nassim Nicholas (2005), Finance without variance, preprint.

Von Neumann, John and Morgenstern, Oskar (1944), *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Wald, Abraham (1950), *Statistical Decision Functions*. New York: John Wiley and Sons.

FOREIGN AFFAIRS

MAY/JUNE 2011

The Black Swan of Cairo

How Suppressing Volatility Makes the World Less Predictable and More Dangerous

Nassim Nicholas Taleb and Mark Blyth

Volume 90 • Number 3

The Black Swan of Cairo

How Suppressing Volatility Makes the World Less Predictable and More Dangerous

Nassim Nicholas Taleb and Mark Blyth

Why is surprise the permanent condition of the U.S. political and economic elite? In 2007–8, when the global financial system imploded, the cry that no one could have seen this coming was heard everywhere, despite the existence of numerous analyses showing that a crisis was unavoidable. It is no surprise that one hears precisely the same response today regarding the current turmoil in the Middle East. The critical issue in both cases is the artificial suppression of volatility—the ups and downs of life—in the name of stability. It is both misguided and dangerous to push unobserved risks further into the statistical tails of the probability distribution of outcomes and allow these high-impact, low-probability "tail risks" to disappear from policymakers' fields of observation. What the world is witnessing in Tunisia, Egypt, and Libya is simply what happens when highly constrained systems explode.

Complex systems that have artificially suppressed volatility tend to become extremely fragile, while at the same time exhibiting no visible risks. In fact, they tend to be too calm and exhibit minimal variability as silent risks accumulate beneath the surface. Although the stated intention of political leaders and economic policymakers is to stabilize the system by inhibiting fluctuations, the result tends to be the opposite. These artificially constrained systems become prone to "Black Swans"—that is, they become extremely vulnerable to large-scale events that lie far from the statistical norm and were largely unpredictable to a given set of observers.

Such environments eventually experience massive blowups, catching everyone off-guard and undoing years of stability or, in some cases, ending up far worse than they were in their initial volatile state. Indeed, the longer it takes for the blowup to occur, the worse the resulting harm in both economic and political systems.

Seeking to restrict variability seems to be good policy (who does not prefer stability

---

Nassim Nicholas Taleb is Distinguished Professor of Risk Engineering at New York University's Polytechnic Institute and the author of *The Black Swan: The Impact of the Highly Improbable*. Mark Blyth is Professor of International Political Economy at Brown University.

to chaos?), so it is with very good intentions that policymakers unwittingly increase the risk of major blowups. And it is the same misperception of the properties of natural systems that led to both the economic crisis of 2007–8 and the current turmoil in the Arab world. The policy implications are identical: to make systems robust, all risks must be visible and out in the open—*fluctuat nec mergitur* (it fluctuates but does not sink) goes the Latin saying.

Just as a robust economic system is one that encourages early failures (the concepts of "fail small" and "fail fast"), the U.S. government should stop supporting dictatorial regimes for the sake of pseudostability and instead allow political noise to rise to the surface. Making an economy robust in the face of business swings requires allowing risk to be visible; the same is true in politics.

SEDUCED BY STABILITY

Both the recent financial crisis and the current political crisis in the Middle East are grounded in the rise of complexity, interdependence, and unpredictability. Policymakers in the United Kingdom and the United States have long promoted policies aimed at eliminating fluctuation—no more booms and busts in the economy, no more "Iranian surprises" in foreign policy. These policies have almost always produced undesirable outcomes. For example, the U.S. banking system became very fragile following a succession of progressively larger bailouts and government interventions, particularly after the 1983 rescue of major banks (ironically, by the same Reagan administration that trumpeted free markets). In the United States, promoting these bad policies has been a bipartisan effort throughout. Republicans have been good at fragilizing large corporations through bailouts, and Democrats have been good at fragilizing the government. At the same time, the financial system as a whole exhibited little volatility; it kept getting weaker while providing policymakers with the illusion of stability, illustrated most notably when Ben Bernanke, who was then a member of the Board of Governors of the U.S. Federal Reserve, declared the era of "the great moderation" in 2004.

Putatively independent central bankers fell into the same trap. During the 1990s, U.S. Federal Reserve Chair Alan Greenspan wanted to iron out the economic cycle's booms and busts, and he sought to control economic swings with interest-rate reductions at the slightest sign of a downward tick in the economic data. Furthermore, he adapted his economic policy to guarantee bank rescues, with implicit promises of a backstop—the now infamous "Greenspan put." These policies proved to have grave delayed side effects. Washington stabilized the market with bailouts and by allowing certain companies to grow "too big to fail." Because policymakers believed it was better to do something than to do nothing, they felt obligated to heal the economy rather than wait and see if it healed on its own.

The foreign policy equivalent is to support the incumbent no matter what. And just as banks took wild risks thanks to Greenspan's implicit insurance policy, client governments such as Hosni Mubarak's in Egypt for years engaged in overt plunder thanks to similarly reliable U.S. support.

Those who seek to prevent volatility on the grounds that any and all bumps in the road must be avoided paradoxically increase the probability that a tail risk will cause a major explosion. Consider as a thought

experiment a man placed in an artificially sterilized environment for a decade and then invited to take a ride on a crowded subway; he would be expected to die quickly. Likewise, preventing small forest fires can cause larger forest fires to become devastating. This property is shared by all complex systems.

In the realm of economics, price controls are designed to constrain volatility on the grounds that stable prices are a good thing. But although these controls might work in some rare situations, the long-term effect of any such system is an eventual and extremely costly blowup whose cleanup costs can far exceed the benefits accrued. The risks of a dictatorship, no matter how seemingly stable, are no different, in the long run, from those of an artificially controlled price.

Such attempts to institutionally engineer the world come in two types: those that conform to the world as it is and those that attempt to reform the world. The nature of humans, quite reasonably, is to intervene in an effort to alter their world and the outcomes it produces. But government interventions are laden with unintended—and unforeseen—consequences, particularly in complex systems, so humans must work with nature by tolerating systems that absorb human imperfections rather than seek to change them.

Take, for example, the recent celebrated documentary on the financial crisis, *Inside Job*, which blames the crisis on the malfeasance and dishonesty of bankers and the incompetence of regulators. Although it is morally satisfying, the film naively overlooks the fact that humans have always been dishonest and regulators have always been behind the curve. The only difference this time around was the unprecedented magnitude of the hidden risks and a misunderstanding of the statistical properties of the system.

What is needed is a system that can prevent the harm done to citizens by the dishonesty of business elites; the limited competence of forecasters, economists, and statisticians; and the imperfections of regulation, not one that aims to eliminate these flaws. Humans must try to resist the illusion of control: just as foreign policy should be intelligence-proof (it should minimize its reliance on the competence of information-gathering organizations and the predictions of "experts" in what are inherently unpredictable domains), the economy should be regulator-proof, given that some regulations simply make the system itself more fragile. Due to the complexity of markets, intricate regulations simply serve to generate fees for lawyers and profits for sophisticated derivatives traders who can build complicated financial products that skirt those regulations.

DON'T BE A TURKEY

The life of a turkey before Thanksgiving is illustrative: the turkey is fed for 1,000 days and every day seems to confirm that the farmer cares for it—until the last day, when confidence is maximal. The "turkey problem" occurs when a naive analysis of stability is derived from the absence of past variations. Likewise, confidence in stability was maximal at the onset of the financial crisis in 2007.

The turkey problem for humans is the result of mistaking one environment for another. Humans simultaneously inhabit two systems: the linear and the complex. The linear domain is characterized by its predictability and the low degree of interaction among its components, which

allows the use of mathematical methods that make forecasts reliable. In complex systems, there is an absence of visible causal links between the elements, masking a high degree of interdependence and extremely low predictability. Nonlinear elements are also present, such as those commonly known, and generally misunderstood, as "tipping points." Imagine someone who keeps adding sand to a sand pile without any visible consequence, until suddenly the entire pile crumbles. It would be foolish to blame the collapse on the last grain of sand rather than the structure of the pile, but that is what people do consistently, and that is the policy error.

U.S. President Barack Obama may blame an intelligence failure for the government's not foreseeing the revolution in Egypt (just as former U.S. President Jimmy Carter blamed an intelligence failure for his administration's not foreseeing the 1979 Islamic Revolution in Iran), but it is the suppressed risk in the statistical tails that matters—not the failure to see the last grain of sand. As a result of complicated interdependence and contagion effects, in all man-made complex systems, a small number of possible events dominate, namely, Black Swans.

Engineering, architecture, astronomy, most of physics, and much of common science are linear domains. The complex domain is the realm of the social world, epidemics, and economics. Crucially, the linear domain delivers mild variations without large shocks, whereas the complex domain delivers massive jumps and gaps. Complex systems are misunderstood, mostly because humans' sophistication, obtained over the history of human knowledge in the linear domain, does not transfer properly to the complex domain. Humans can predict a solar eclipse and the trajectory of a space vessel, but not the stock market or Egyptian political events. All man-made complex systems have commonalities and even universalities. Sadly, deceptive calm (followed by Black Swan surprises) seems to be one of those properties.

THE ERROR OF PREDICTION

As with a crumbling sand pile, it would be foolish to attribute the collapse of a fragile bridge to the last truck that crossed it, and even more foolish to try to predict in advance which truck might bring it down. The system is responsible, not the components. But after the financial crisis of 2007–8, many people thought that predicting the subprime meltdown would have helped. It would not have, since it was a symptom of the crisis, not its underlying cause. Likewise, Obama's blaming "bad intelligence" for his administration's failure to predict the crisis in Egypt is symptomatic of both the misunderstanding of complex systems and the bad policies involved.

Obama's mistake illustrates the illusion of local causal chains—that is, confusing catalysts for causes and assuming that one can know which catalyst will produce which effect. The final episode of the upheaval in Egypt was unpredictable for all observers, especially those involved. As such, blaming the CIA is as foolish as funding it to forecast such events. Governments are wasting billions of dollars on attempting to predict events that are produced by interdependent systems and are therefore not statistically understandable at the individual level.

As Mark Abdollahian of Sentia Group, one of the contractors who sell predictive analytics to the U.S. government, noted regarding Egypt, policymakers should

"think of this like Las Vegas. In blackjack, if you can do four percent better than the average, you're making real money." But the analogy is spurious. There is no "four percent better" on Egypt. This is not just money wasted but the construction of a false confidence based on an erroneous focus. It is telling that the intelligence analysts made the same mistake as the risk-management systems that failed to predict the economic crisis—and offered the exact same excuses when they failed. Political and economic "tail events" are unpredictable, and their probabilities are

not scientifically measurable. No matter how many dollars are spent on research, predicting revolutions is not the same as counting cards; humans will never be able to turn politics into the tractable randomness of blackjack.

Most explanations being offered for the current turmoil in the Middle East follow the "catalysts as causes" confusion. The riots in Tunisia and Egypt were initially attributed to rising commodity prices, not to stifling and unpopular dictatorships. But Bahrain and Libya are countries with high GDPs that can afford to import grain and other commodities. Again, the focus is wrong even if the logic is comforting. It is the system and its fragility, not events, that must be studied—what physicists call "percolation theory," in which the properties of the terrain are studied rather than those of a single element of the terrain.

When dealing with a system that is inherently unpredictable, what should be done? Differentiating between two types of countries is useful. In the first, changes in government do not lead to meaningful differences in political outcomes (since political tensions are out in the open). In the second type, changes in government lead to both drastic and deeply unpredictable changes.

Consider that Italy, with its much-maligned "cabinet instability," is economically and politically stable despite having had more than 60 governments since World War II (indeed, one may say Italy's stability is because of these switches of government). Similarly, in spite of consistently bad press, Lebanon is a relatively safe bet in terms of how far governments can jump from equilibrium; in spite of all the noise, shifting alliances, and street protests, changes in government there tend to be comparatively mild. For example, a shift in the ruling coalition from Christian parties to Hezbollah is not such a consequential jump in terms of the country's economic and political stability. Switching equilibrium, with control of the government changing from one party to another, in such systems acts as a shock absorber. Since a single party cannot have total and more than temporary control, the possibility of a large jump in the regime type is constrained.

In contrast, consider Iran and Iraq. Mohammad Reza Shah Pahlavi and Saddam Hussein both constrained volatility by any means necessary. In Iran, when the shah was toppled, the shift of power to Ayatollah Ruhollah Khomeini was a huge, unforeseeable jump. After the fact, analysts could construct convincing accounts about how killing Iranian Communists, driving the left into exile, demobilizing the democratic opposition, and driving all dissent into the mosque had made Khomeini's rise inevitable. In Iraq, the United States removed the lid and was actually surprised to find that the regime did not jump from hyperconstraint to something like France. But this was impossible to predict ahead of time due to the nature of the system itself. What can be said, however, is that the more constrained the volatility, the bigger the regime jump is likely to be. From the French Revolution to the triumph of the Bolsheviks, history is replete with such examples, and yet somehow humans remain unable to process what they mean.

THE FEAR OF RANDOMNESS

Humans fear randomness—a healthy ancestral trait inherited from a different environment. Whereas in the past, which was a more linear world, this trait enhanced

fitness and increased chances of survival, it can have the reverse effect in today's complex world, making volatility take the shape of nasty Black Swans hiding behind deceptive periods of "great moderation." This is not to say that any and all volatility should be embraced. Insurance should not be banned, for example.

But alongside the "catalysts as causes" confusion sit two mental biases: the illusion of control and the action bias (the illusion that doing something is always better than doing nothing). This leads to the desire to impose man-made solutions. Greenspan's actions were harmful, but it would have been hard to justify inaction in a democracy where the incentive is to always promise a better outcome than the other guy, regardless of the actual, delayed cost.

Variation is information. When there is no variation, there is no information. This explains the CIA's failure to predict the Egyptian revolution and, a generation before, the Iranian Revolution—in both cases, the revolutionaries themselves did not have a clear idea of their relative strength with respect to the regime they were hoping to topple. So rather than subsidize and praise as a "force for stability" every tin-pot potentate on the planet, the U.S. government should encourage countries to let information flow upward through the transparency that comes with political agitation. It should not fear fluctuations per se, since allowing them to be in the open, as Italy and Lebanon both show in different ways, creates the stability of small jumps.

As Seneca wrote in *De clementia*, "Repeated punishment, while it crushes the hatred of a few, stirs the hatred of all . . . just as trees that have been trimmed throw out again countless branches." The imposition of peace through repeated punishment lies at the heart of many seemingly intractable conflicts, including the Israeli-Palestinian stalemate. Furthermore, dealing with seemingly reliable high-level officials rather than the people themselves prevents any peace treaty signed from being robust. The Romans were wise enough to know that only a free man under Roman law could be trusted to engage in a contract; by extension, only a free people can be trusted to abide by a treaty. Treaties that are negotiated with the consent of a broad swath of the populations on both sides of a conflict tend to survive. Just as no central bank is powerful enough to dictate stability, no superpower can be powerful enough to guarantee solid peace alone.

U.S. policy toward the Middle East has historically, and especially since 9/11, been unduly focused on the repression of any and all political fluctuations in the name of preventing "Islamic fundamentalism"—a trope that Mubarak repeated until his last moments in power and that Libyan leader Muammar al-Qaddafi continues to emphasize today, blaming Osama bin Laden for what has befallen him. This is wrong. The West and its autocratic Arab allies have strengthened Islamic fundamentalists by forcing them underground, and even more so by killing them.

As Jean-Jacques Rousseau put it, "A little bit of agitation gives motivation to the soul, and what really makes the species prosper is not peace so much as freedom." With freedom comes some unpredictable fluctuation. This is one of life's packages: there is no freedom without noise—and no stability without volatility.