

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

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”) beyond Jensen’s Inequality, 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”,

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

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

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).

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.

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 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%)

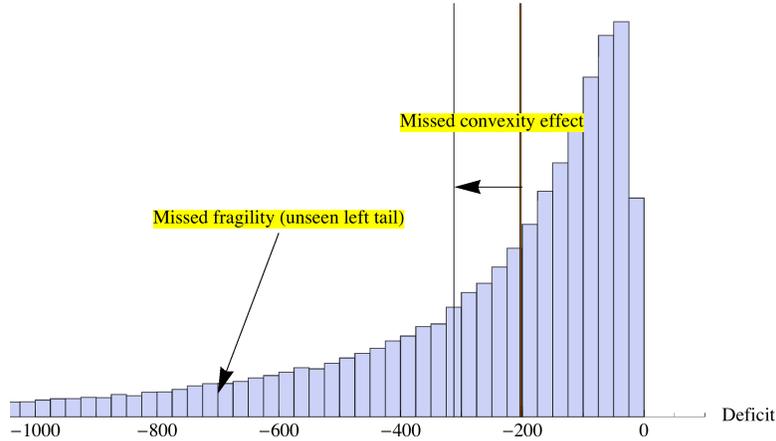


Figure 1 **CONVEXITY EFFECTS ALLOW THE DETECTION OF BOTH MODEL BIAS AND FRAGILITY.** Illustration of the example; histogram from Monte Carlo simulation of government deficit as a left-tailed random variable simply as a result of randomizing unemployment of which it is a convex function. The method of 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_3 since our definition of fragility is grounded in nonlinearities. Further, the “fat tailedness” 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 \Leftarrow 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 unemployment.
- e. Example: hidden tail exposure from dependence on a source of energy, etc. ("squeezability argument").

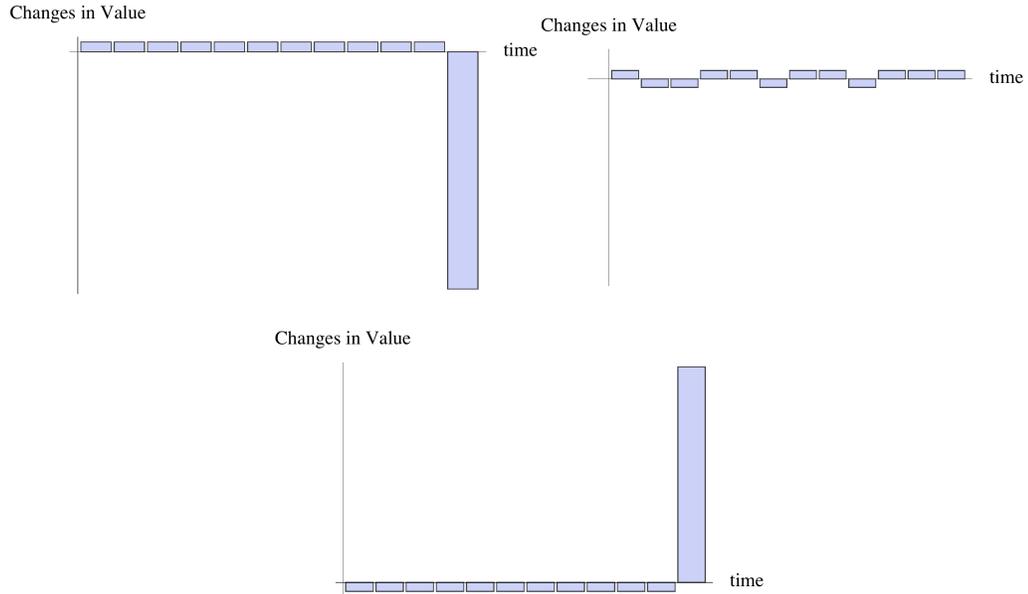


Figure 2 TYPE 1, Fragile variations through time—the horizontal axis shows time. This can apply to anything, a coffee cup, a health indicator, changes in wealth, your happiness, etc. We can see small (or no) benefits most of the time and occasional large adverse outcomes. Uncertainty can hit in a rather 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 Δs 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 Δs instead of the infinitesimal mathematical derivative.

Sources of Fragility: y is a function, that is, a derivative of some primitive 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 function 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 $V(y, f, K, \Delta s)$ and the alternative distribution $V(y, f^*, K, \Delta s)$, 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 distributions. 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, $V(y,f,K,\Delta p)$ 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.

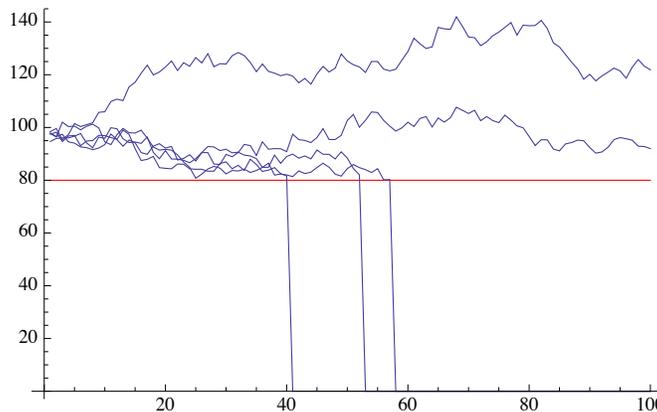
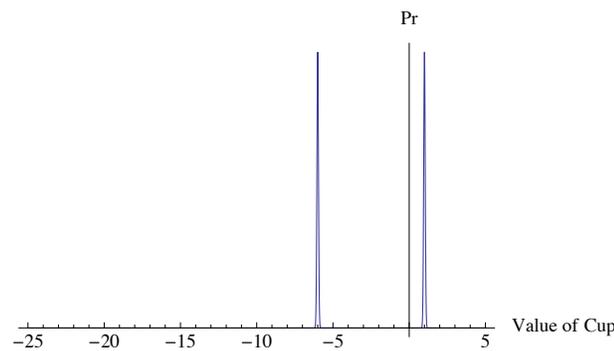
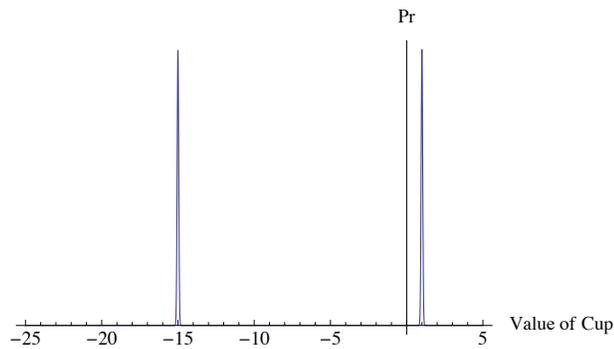


Figure 5 : $L=80, S_0 = 100$. Absorbing barrier (down and out) causes a special brand of left tails, when the barrier causes a collapse to a certain value (here, 0, with no residual)

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 $\zeta(y, f, K, s)$ 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

$$y \text{ is robust below } L$$

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 further how antifragility benefits from Jensen’s inequality.

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

Our definition is based on philostochasticity, love of variations (fragile is stochastophobe). Positive sensitivity to variations 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) \implies Left tailedness (in probability space) \implies 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 asymmetric 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 \rightarrow 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 $g(x) = 1 - e^{-a \cdot x}$ (typically used with utility models) and x Gaussian (μ, σ) distributed. With $a=1$, the

distribution becomes
$$v(x) = -\frac{e^{-\frac{(\mu+\log(1-x))^2}{2\sigma^2}}}{\sqrt{2\pi} \sigma (X-1)}$$

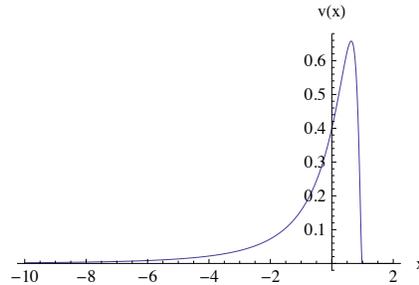


Figure 9- Left tails, pure skewness from concave transformation $1 - e^{-x}$ of a Gaussian variable x .

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

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

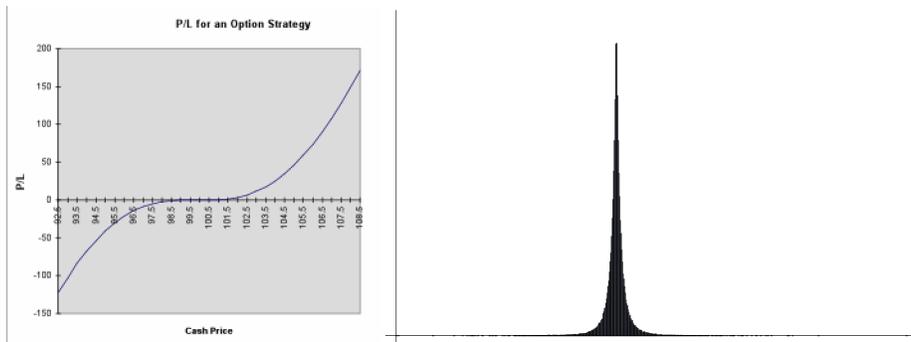


Figure 10 A "Fence" Payoff. From Dynamic Hedging, Taleb(1997), (short series of OTM puts long series OTM call, delta neutral). The shape can be expressed with $f = e^x$ if $x > 0$, $2 - e^{-x}$ otherwise.

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 function, 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)

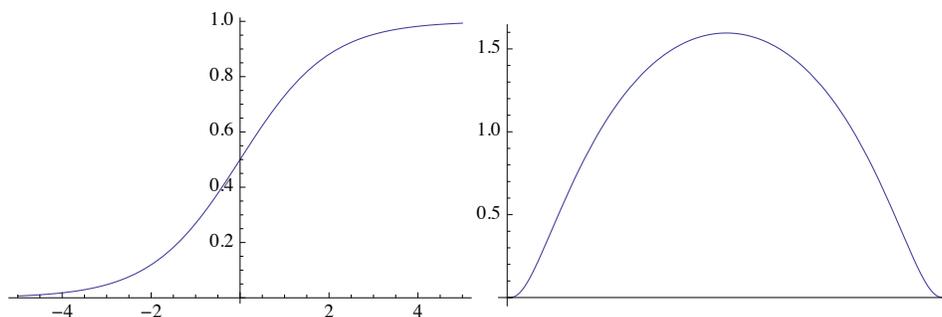


Figure 7- The "digital" payoff, opposite style to the "fence", closer to a 1st or 3rd quadrant payoff (Taleb, 1998), that is, bounded on both sides. Its properties can be captured with the sigmoid function $\frac{1}{1+e^{-ax}}$

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

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

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.

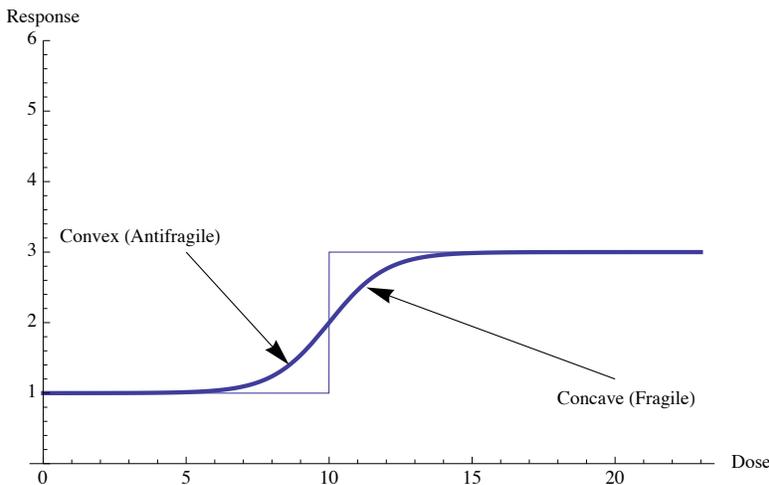


Figure 13- Dose response in biological systems (similar to digital payoffs). The mixture regularizes say, caloric consumption.

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.

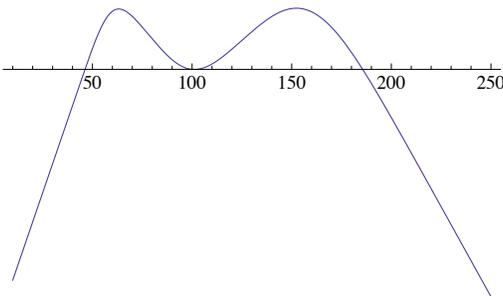


Figure14 : Many risks are hidden under local convexity and tails concavity. From *Dynamic Hedging* (1997). There are simply remedied with slightly larger size of perturbations.

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_3 missing 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 $2 \Delta x$ satisfies the following inequality:

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

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 $\{X_i\}$,

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

For example, take a conventional die (six sides) and consider a payoff equal to the number it lands on. The expected (average) payoff is $\frac{1}{6} \sum_{i=1}^6 i = 3.5$. Now consider that we get the squared payoff, $\frac{1}{6} \sum_{i=1}^6 i^2 = \frac{91}{6} \approx 15.17$, while $(\frac{1}{6} \sum_{i=1}^6 i)^2 = 12.25$, 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 defiiect/employment example used in the introduction, say we are using a simple function:

$$f(x | \bar{\alpha}) \tag{7}$$

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:

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

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:

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

where Φ is the Lognormal distribution, S_0 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:

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

(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 6σ away from the money, with $a= 1/5$, the relative bias approaches 5000% in option value. Note that $a= 1/5$ is “mild compared to the variations we see, with a often = $4/5$.

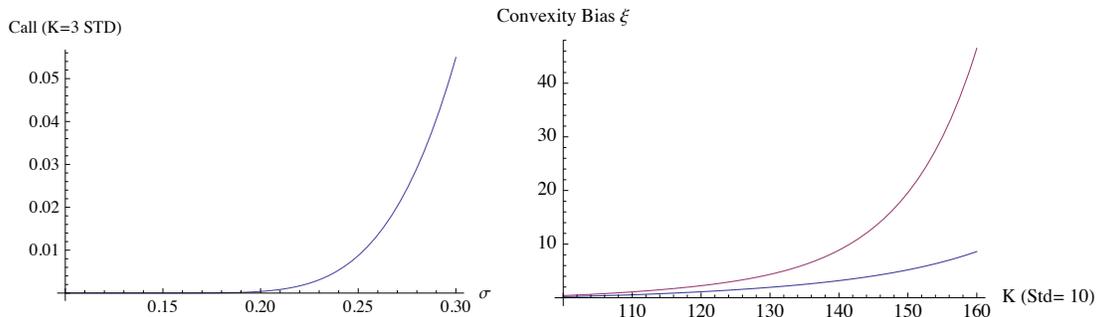


Figure 15 The value of a call as a function of σ ; out of the money options are extremely convex to σ .

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 stochasticizing 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)

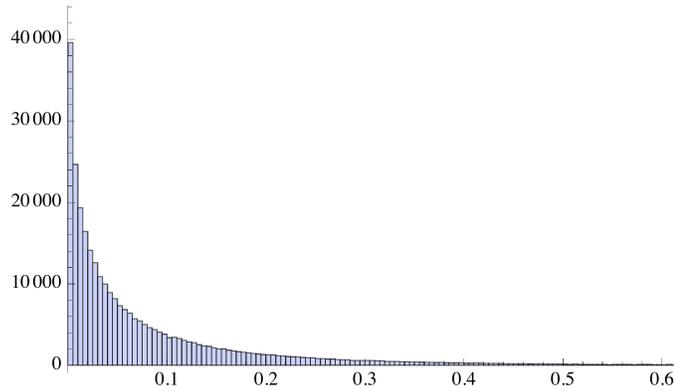


Figure 17 Distribution of the value of the option in previous example across stochastic σ [n=300K simulations]. Much more skewed.

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

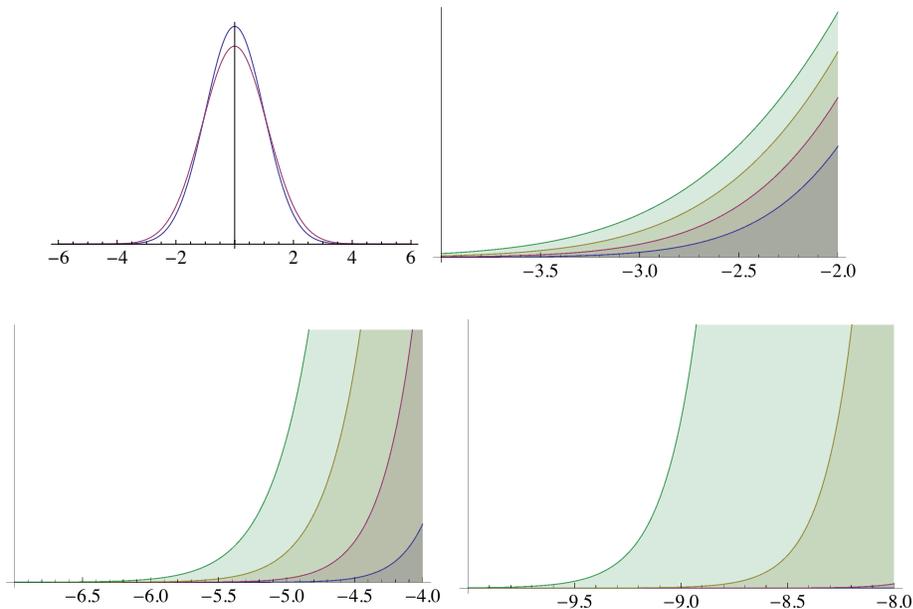
$$\xi \equiv \frac{P > K | p(1 + a) + P > K | p(1 - a)}{P > K | p} \tag{14}$$

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 $\frac{1}{n} a(1)(1 \pm a(2))(1 \pm a(3)) \dots (1 \pm a(n))$ for additional convexity effects.

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

K	$\xi (a=1/10)$	$\xi (a=1/5)$
0	1.000	1.000
1	0.9924	0.9706
2	1.047	1.187
3	1.342	2.333
4	2.252	6.778
5	4.805	26.96
10	3222.	2.579×10^6
15	1.658×10^8	5.084×10^{14}
20	6.529×10^{14}	2.079×10^{26}
25	1.970×10^{23}	1.766×10^{41}

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 perturbrates 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 unemployment forecast of 8%, as a point estimate. We ask the government 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 mainly 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 miserably 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), $H = \frac{V(x, f, K, \Delta s + 1/2 \Delta s) + V(x, f, K, \Delta s - 1/2 \Delta s)}{2 V(x, f, K, \Delta s)}$ 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 unemployment
- 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 perturbing a certain parameter p :

- a. Size and pseudoeconomies 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, Emanuel Derman, Jean-Philippe Bouchaud.

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, Columbia University September 24 2011.

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