Convexity, Robustness, and Model Error inside the Fourth Quadrant,

Nassim N Taleb

DRAFT VERSION, July 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 the *Fourth Quadrant*). It makes the distinction between fragile an robust to model (or representational) error on the basis of convexity. It also introduces a simple practical method to measure (as an indicator of fragility) the sensitivity of a portfolio (or balance sheet) to model error.

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].

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

[1] This paper is slightly more technical than 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 academics and other nerds that want to provide a critical comment on my work should use this paper and the Fourth Quadrant, not my writing style in *The Black Swan* unless they 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.

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.

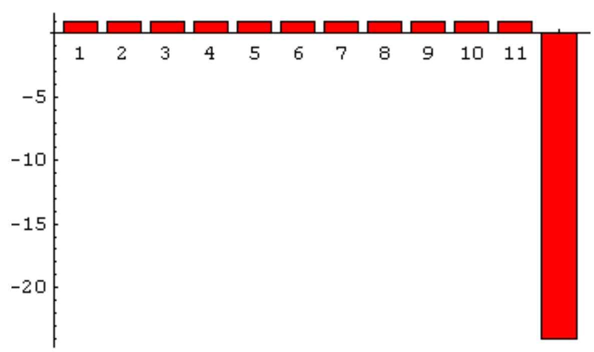

Figure 1 Concave payoff through time, with respect to a source of variation; or concave errors from left-skewed distributions.

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, $\int^{-K} x\, f(x)\, dx$ compared to $\int_K x\, f(x)\, dx$, K being a remote threshold for x the source of variation.

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

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.

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)*.

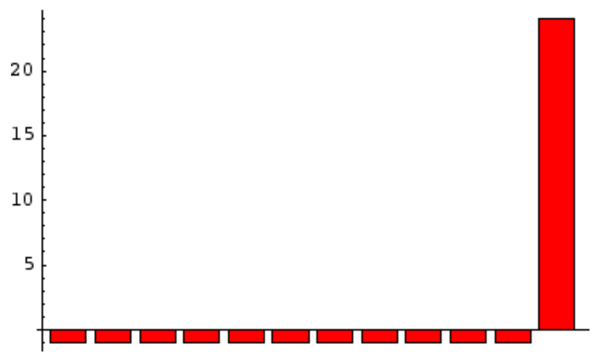

Figure 2 Convex payoff through time, or convex errors.

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".

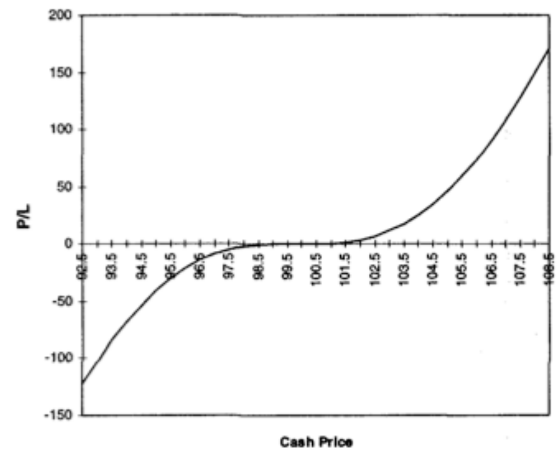

Figure 3 From Dynamic Hedging, Taleb(1997), most payoffs are mixed.

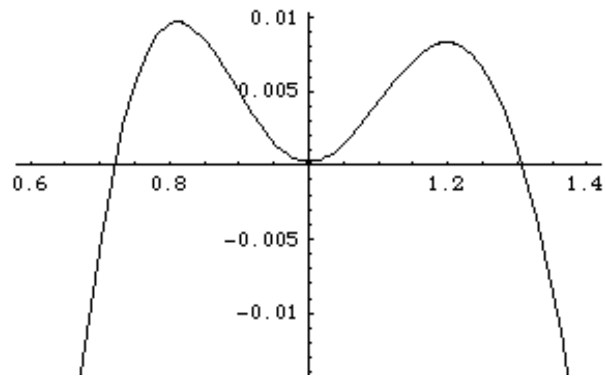

Figure 4 Convexity can be just local for a small or medium size variation, which is why measures need to be broad and fill the tails. Most banks have fallen for this trap: the banking system accumulated concavity where it was invisible.

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, thanks to whom I am now here, 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 $\gamma=1$ is no effect, a $\gamma=2$ is the doubling the shortfall). With δ in [0,1[, and assuming for simplicity $\mu=0$,

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

which yields to a closed form solution

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

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. How to Measure Model Errors with Simple Perturbations

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

[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.

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".

General Mathematical Derivations: Our random variable X is the "unintended exposure". Assume the

[19] Taleb and Tapiero (2010)

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

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 α k /(α-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.

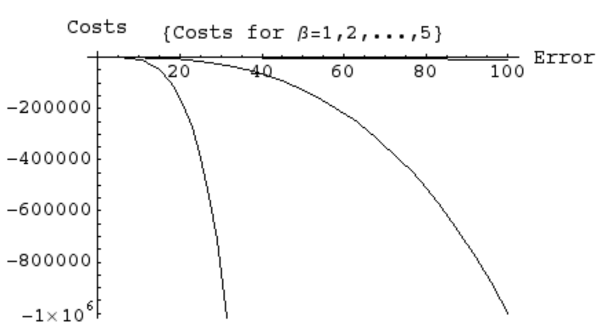

Figure 5- Various loss functions of increasing convexity: -b x^{β} for b=1, a=2,...,5

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 α/β

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

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

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

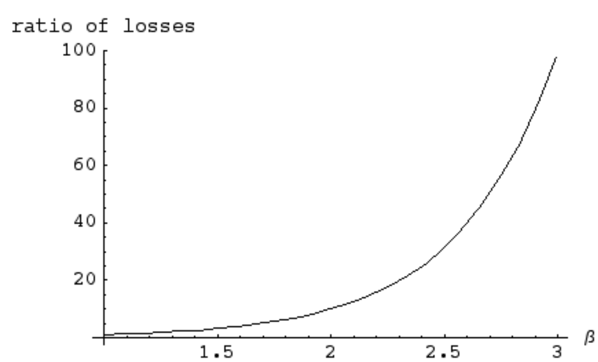

Figure 6 ratio of losses for N=1 entity/ Losses for N=10 entities as β increases. As β reaches α, the expectation of the losses becomes infinite.

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

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

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"

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.

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

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:

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 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.

BELIEF IN THE LAW OF SMALL NUMBERS AND OVERCONFIDENCE

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) writes "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 inductive 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 .

It is easy to see that the "small numbers" gets exacerbated with skewness since most of the time the observed mean will be different from the true mean and most of the time the observed variance will 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.

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

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. Since Alpert and Raiffa (1982) 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 happen 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 or making), 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 of 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.

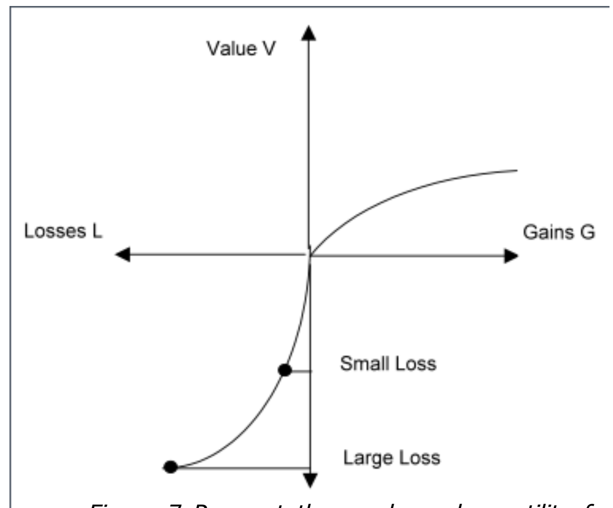

Figure 7 Prospect theory shows how utility for losses is convex, utility for profits is concave (Taleb 2004a) which acts as an incentive for negatively skewed trades.

It is key that prospect theory was empirically derived from one-shot experiments with agents subjected to questions in which the odds were supplied. Nor has it been subjected to streams of payoffs, the concerns of this discussion, a point to which we will return. It is the value function of the prospect theory that we examine next, not the 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.

The empirically derived version of utility theory presents asymmetric higher order properties. The Kahneman-Tversky value-function v for changes in wealth is convex to in the loss domain v(L) and concave in the profit domain v(G). Since the second derivative of v(L) is positive, we have by convexity the value of a large loss higher than the sum of the value of losses: $v(L) > n\, v(L/n)$. 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).

BY comparison, the conventional Von Neuman-Morgenstern utility of wealth (instead of payoffs), while making no such distinction, results in asymmetry in skewness preferences: U(W) is concave for all levels of wealth W which makes the investor favor negative skewness on the right and positive skewness on the left for incremental changes (How? $U(W+\ \Delta W) < U(W)+U(\Delta W)$ if $\Delta W>0$, since $U''(W) <0$ for all W, and < otherwise). In the domain of gains or increase in wealth there is an convergence between the two methods of viewing utility: over a single period the right-side utilities are both concave.

The result here is sufficiently firm to require no additional testing: Prospect theory has been subjected to all manner of experiments and the concavity in the domain of losses has shown to be robust.

Take an example from the parametrization of Tversky and Kahneman (1992). V+(x) for x positive and V-(x) for x strictly negative.

V+(x)= xα

V-(x)= (-λ)(-x)α

Take α=2.25, λ= .65

This point is not fraught with a modicum of ambiguity – the . Ignorance about convergence of distributions leads to underestimate the probabilities of large deviations owing to their mathematical properties, but there is a documented countervailing tendency to overestimate the small odds. The probability weighting function in the Kahneman-Tversky prospect theory (see Tversky and Kahneman, 1992) implies that the agent, in his decision making, overestimates small probabilities and underestimates the higher ones (the ones closer to 100). It seems to contradict the earlier effect but this implies that the agent knows these probabilities, which, in a framework of purely inductive inference, he doesn't.

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) .

HEDONIC ADAPTATION

The central idea behind recent research on well being is the existence of a set-point of happiness, to which the agent tends to revert after some departure –the Brickman and Campbell (1971) hedonic treadmill. Such mechanism seems to be the backbone for the research on happiness and economics. The idea provides an explanation to prospect theory, as the sensitivity decreases on both sides, and the agent is sensitive to differences rather than to absolute conditions, as he resets his utility curve at the origin.

The problems, however seem to be that adaptation is selective and domain specific. In some cases, repetition or duration of a constant stimulus results in an increasing hedonic response –a process the literature calls sensitization. The literature (Frederick and Loewenstein (1998)) shows evidence that there are some things to which we adapt rapidly: (imprisonment, increases in wealth, and disabilities like paralysis), condition to which we adapt slowly (the death of a loved one), and things to which we do not seem to adapt (noise, debilitating diseases, foods, or an annoying roommate). Now the question: do people adapt to bleed? In other words do people increase in sensitivity to the pain of the "Chinese torture" treatment of slow losses?