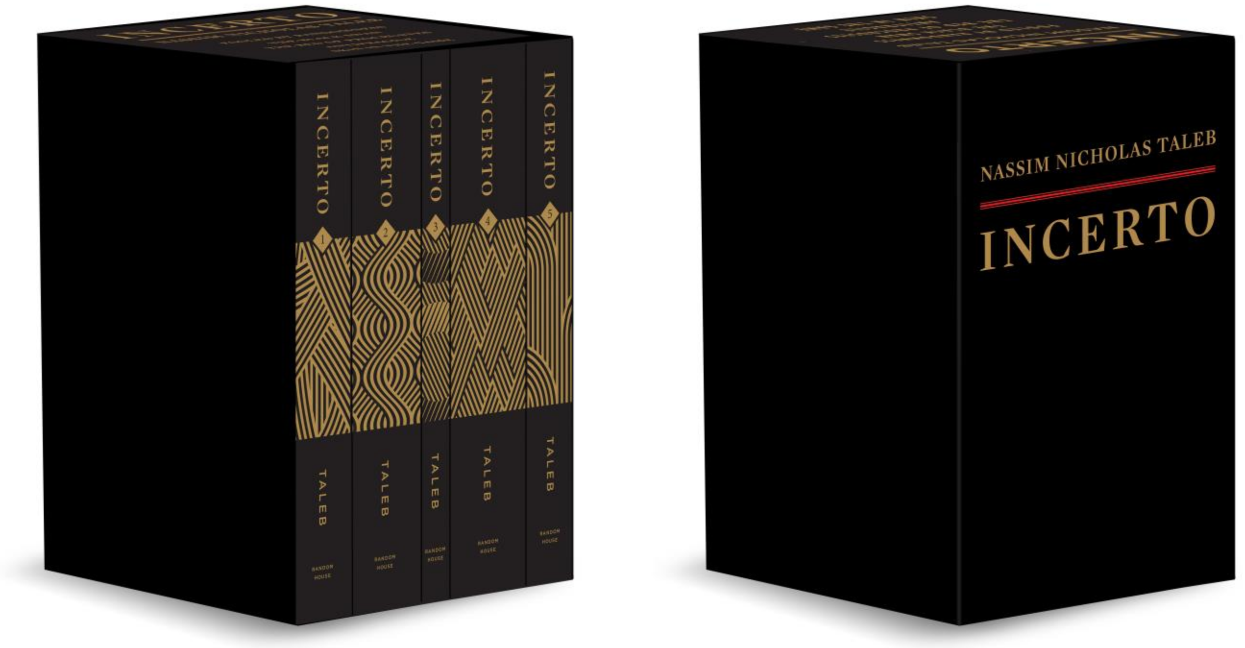
The Technical Incerto

Statistical Consequences of Fat Tails

REAL WORLD PREASYMPTOTICS, EPISTEMOLOGY, AND APPLICATIONS

Papers and Commentary

Nassim Nicholas Taleb


INCERTO
1
2
3
4
5
TALEB
RANDOM HOUSE
NASSIM NICHOLAS TALEB
INCERTO

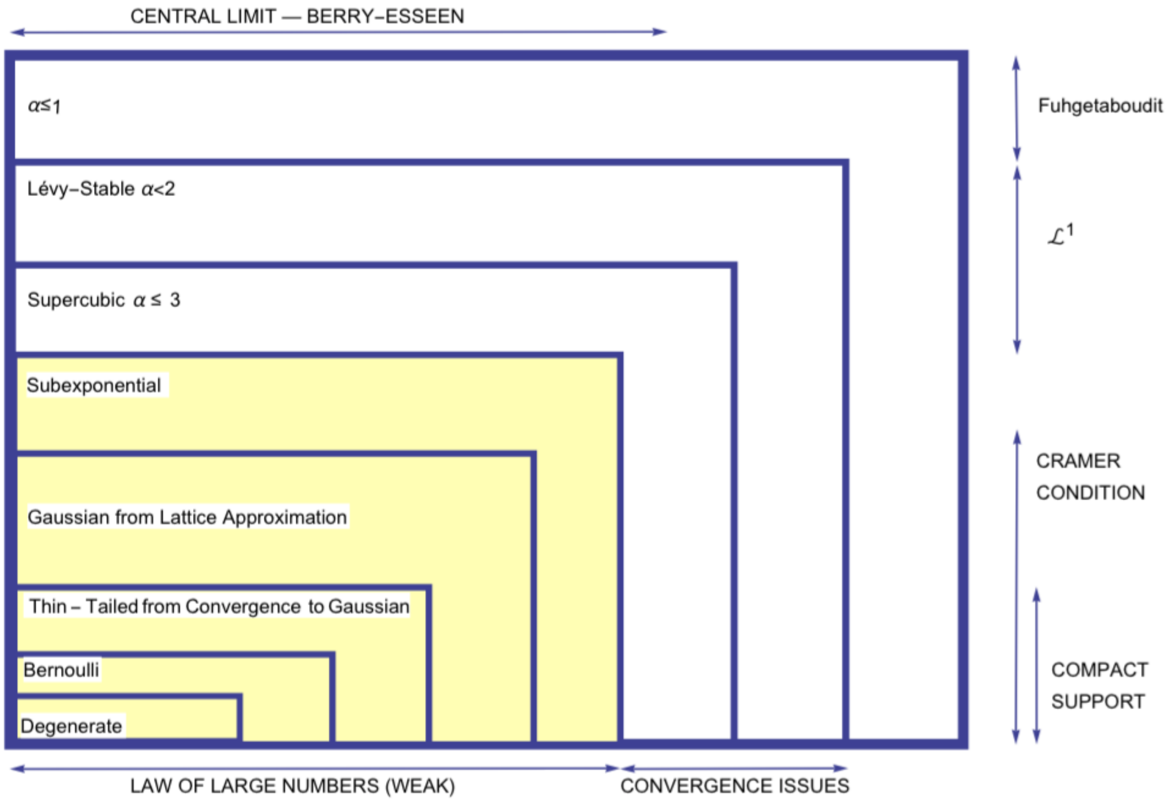


Figure 2.3: *The tableau of Fat tails, along the various classifications for convergence purposes (i.e., convergence to the law of large numbers, etc.) and gravity of inferential problems. Power Laws are in white, the rest in yellow. See Embrechts et al [61].*

Reasoning Errors: It is not changing the color of the dress

Fattest tails [ranking]

Pandemics
Wars
Cultural phenomena
Inflation
Technology companies, Size of cities
Earthquakes (energy)
Financial Returns

A- If the variable is "two-tailed", that is, its domain of support $\mathcal{D}$= (-∞,∞), and where $p^\delta(x) \triangleq \frac{p(x,\sigma+\delta)+p(x,\sigma-\delta)}{2}$,

1. There exist a "high peak" inner tunnel, A_T= (a_2, a_3) for which the δ-perturbed σ of the probability distribution $p^\delta(x) \geq p(x)$ if $x \in (a_2, a_3)$
2. There exists outer tunnels, the "tails", for which $p^\delta(x) \geq p(x)$ if $x \in (-\infty, a_1)$ or $x \in (a_4, \infty)$

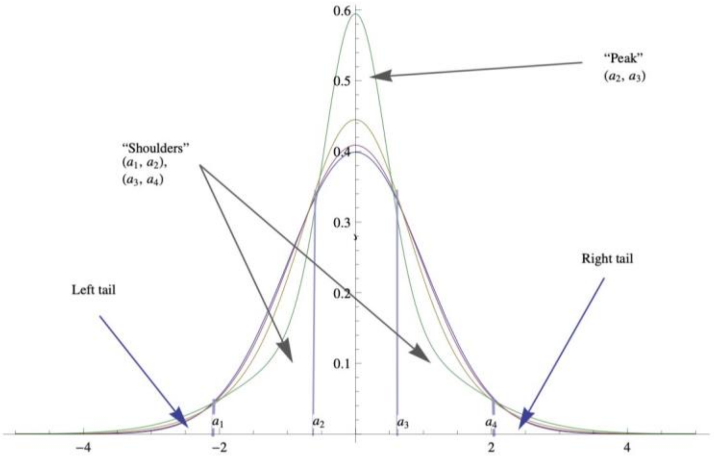


Figure 4.4: *Where do the tails start? Fatter and fatter fails through perturbation of the scale parameter σ for a Gaussian, made more stochastic (instead of being fixed). Some parts of the probability distribution gain in density, others lose. Intermediate events are less likely, tails events and moderate deviations are more likely. We can spot the crossovers a_1 through a_4. The "tails" proper start at a_4 on the right and a_1 on the left.*

$$\{a_1, a_2, a_3, a_4\} = \left\{\mu - \sqrt{\frac{1}{2}\left(5+\sqrt{17}\right)}\sigma, \mu - \sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\sigma, \mu + \sqrt{\frac{1}{2}\left(5-\sqrt{17}\right)}\sigma, \mu + \sqrt{\frac{1}{2}\left(5+\sqrt{17}\right)}\sigma\right\}$$

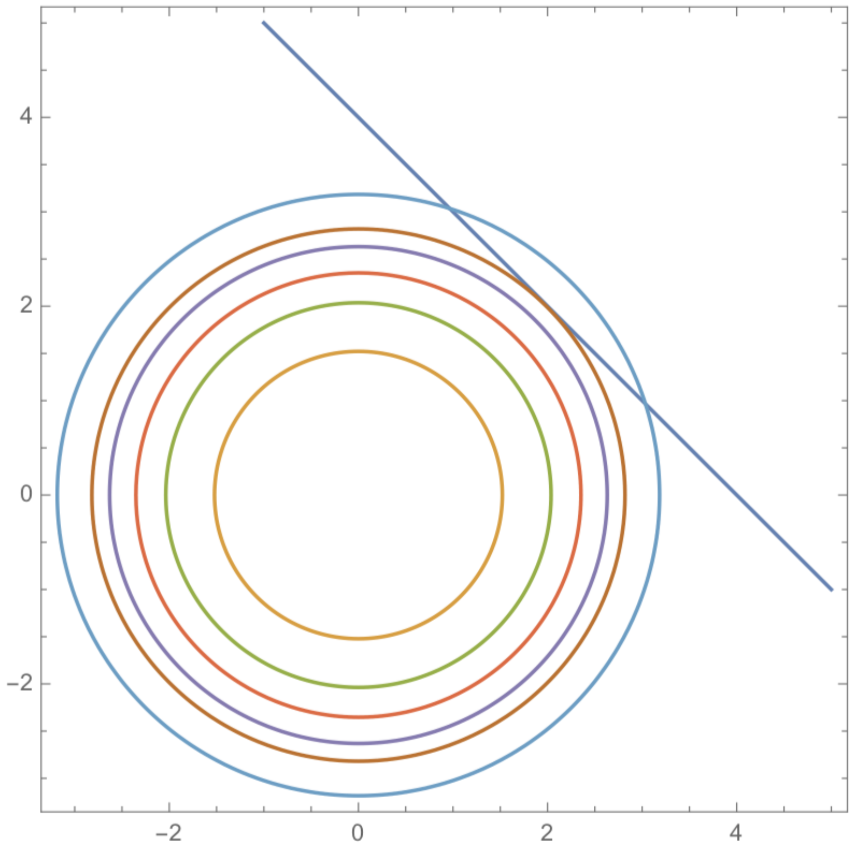


Figure 2.2: *Iso-densities for two independent Gaussian distributions. The line shows* $x + y = 4.1$*. Visibly the maximal probability is for* $x = y = 2.05$*.*

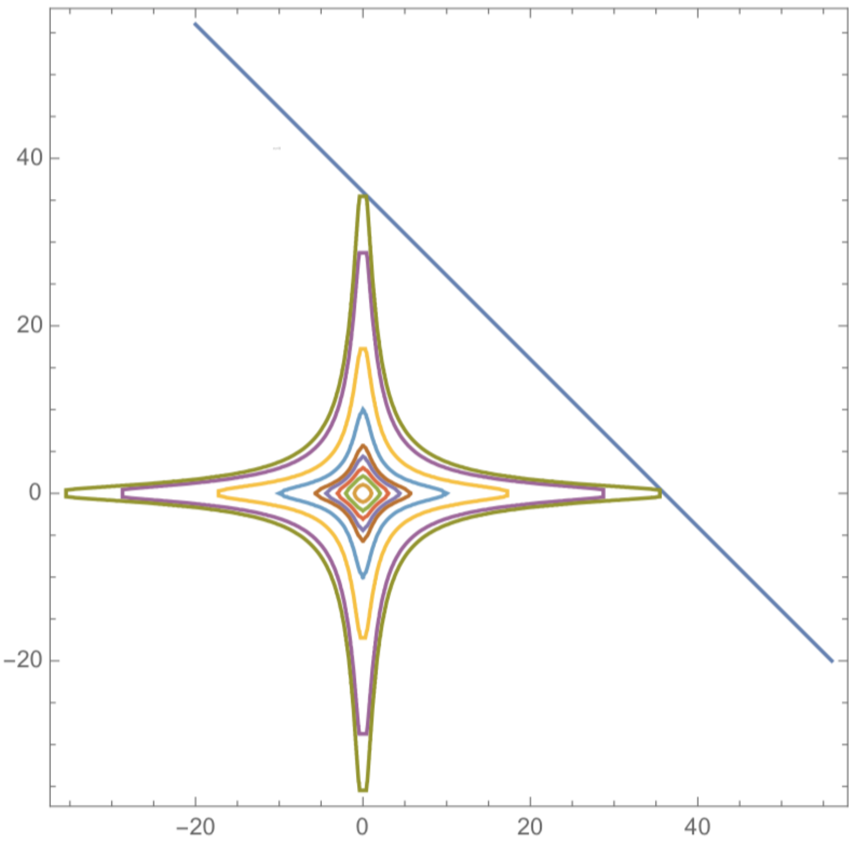
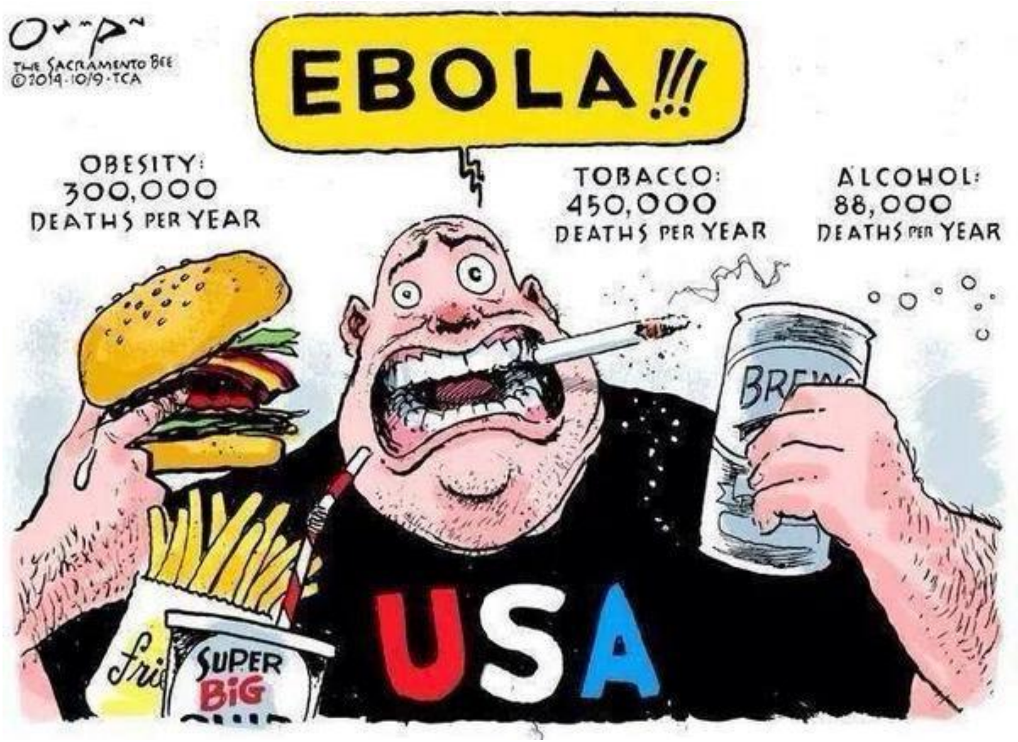
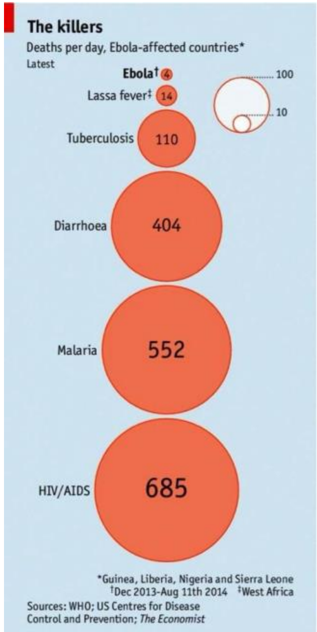


Figure 2.3: *Iso-densities for two independent Fat tailed distributions (in the power law class). The line shows* $x + y = 36$*. Visibly the maximal probability is for either* $x = 36 - \epsilon$ *or* $y = 36 - \epsilon$*, with* ϵ *going to* 0 *as the sum* $x + y$ *becomes larger, with the iso-densities looking more and more like a cross.*

Fattest tails [ranking]

Pandemics
Wars
Cultural phenomena
Inflation
Technology companies, Size of cities
Earthquakes (energy)
Financial Returns


THE SACRAMENTO BEE
©2014 · 10/9 · TCA
EBOLA!!!
OBESITY:
300,000
DEATHS PER YEAR
TOBACCO:
450,000
DEATHS PER YEAR
ALCOHOL:
88,000
DEATHS PER YEAR
USA
SUPER
BIG


The killers
Deaths per day, Ebola-affected countries*
Latest
Ebola† 4
Lassa fever‡ 14
Tuberculosis 110
Diarrhoea 404
Malaria 552
HIV/AIDS 685
100
10
*Guinea, Liberia, Nigeria and Sierra Leone
†Dec 2013-Aug 11th 2014 ‡West Africa
Sources: WHO; US Centres for Disease Control and Prevention; The Economist

Dr. Phil appears on Laura Ingraham and says we don't shut the country down for automobile deaths, cigarette related deaths, and swimming pool deaths

Contents lists available at ScienceDirect

International Journal of Forecasting

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

On single point forecasts for fat-tailed variables

Nassim Nicholas Taleb [a,b,*], Yaneer Bar-Yam [c], Pasquale Cirillo [d,e]

[a] Universversa Investments, Miami, United States of America
[b] Tandon School of Engineering, New York University, New York City, NY, United States of America
[c] New England Complex Systems Institute, Cambridge, United States of America
[d] M Open Forecasting Center and Institute For the Future, University of Nicosia, Nicosia, Cyprus
[e] S-T-A-T-S GmbH, Bern, Switzerland

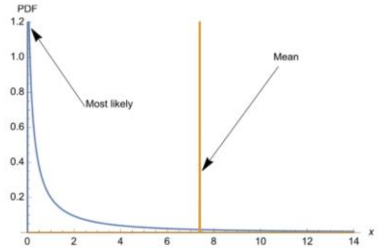


Fig. 1. A high variance lognormal distribution: 85% of observations fall below the mean; half the observations fall below 13% of the mean. The lognormal has milder tails than the Pareto which has been shown to represent pandemics.

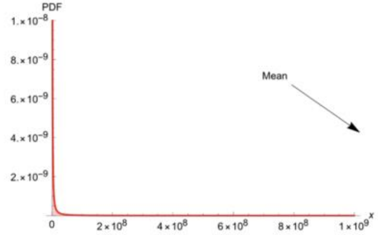


Fig. 2. A Pareto distribution with a tail similar to that of the pandemics. It makes no sense to forecast a single point. The "mean" is so far away you *almost* never observe it. You need to forecast things other than the mean. And most of the density is where there is noise.

Black Swan (2007)

prove how globalization takes us into Extremistan: the notion of species density. Simply, larger environments are more scalable than smaller ones—allowing the biggest to get even bigger, at the expense of the smallest, through the mechanism of preferential attachment we saw in Chapter 14. We have evidence that small islands have many more species per square meter than larger ones, and, of course, than continents. As we travel more on this planet, epidemics will be more acute—we will have a germ population dominated by a few numbers, and the successful killer will spread vastly more effectively. Cultural life will be dominated by fewer persons: we have fewer books per reader in English than in Italian (this includes bad books). Companies will be more uneven in size. And fads will be more acute. So will runs on the banks, of course.

Once again, I am not saying that we need to stop globalization and prevent travel. We just need to be aware of the side effects, the trade-offs—and few people are. I see the risks of a very strange acute virus spreading throughout the planet.

Preasymptotics for Summands

- **There is *no such thing* as infinite summands in the real world**
- ***n* “large” but not asymptotic is not necessarily in the perceived distributional class**

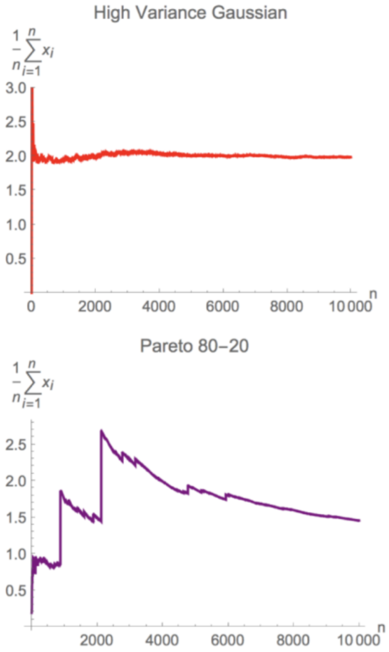


Figure 2.2: *The law of large numbers, that is how long it takes for the sample mean to stabilize, works much more slowly in Extremistan (here a Pareto distribution with 1.13 tail exponent , corresponding to the "Pareto 80-20". Both have the same mean absolute deviation. Note that the same applies to other forms of sampling, such as portfolio theory.*

1. **The law of large numbers, when it works, works too slowly in the real world.**

[1]A non-negative continuous random variable X has a fat-tailed distribution, if its survival function $S(x) = P(X \geq x)$ is regularly varying, formally $S(x) = L(x)x^{-\alpha}$, where $L(x)$ is a slowly varying function, for which $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$ for $t > 0$ [9], [10], [11]. The parameter α is known as the tail parameter, and it governs the fatness of the tail (the smaller α the fatter the tail) and the existence of moments ($E[X^p] < \infty$ if and only if $\alpha > p$).

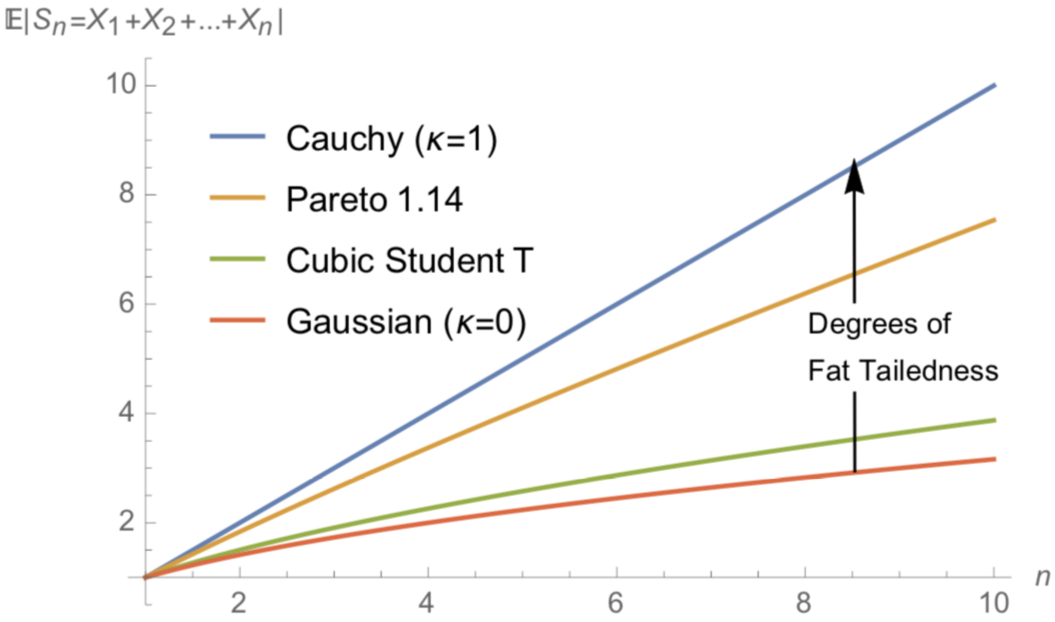
Consequence 1

The law of large numbers, when it works, works too slowly in the real world.

This is more shocking than you think as it cancels most statistical estimators. See Figure 2.4 in this chapter for an illustration. The subject is treated in Chapter 7 and distributions are classified accordingly.[9]

Table 2.1: *Corresponding n_α, or how many observations to get a drop in the error around the mean for an equivalent α-stable distribution (the measure is discussed in more details in Chapter 7). The Gaussian case is the $\alpha = 2$. For the case with equivalent tails to the 80/20 one needs at least 10^{11} more data than the Gaussian.*

α	n_α Symmetric	$n_\alpha^{\beta=\pm\frac{1}{2}}$ Skewed	$n_\alpha^{\beta=\pm 1}$ One-tailed
1	*Fughedaboudit*	-	-
$\frac{9}{8}$	6.09×10^{12}	2.8×10^{13}	1.86×10^{14}
$\frac{5}{4}$	574,634	895,952	1.88×10^{6}
$\frac{11}{8}$	5,027	6,002	8,632
$\frac{3}{2}$	567	613	737
$\frac{13}{8}$	165	171	186
$\frac{7}{4}$	75	77	79
$\frac{15}{8}$	44	44	44
2	30.	30	30



$\mathbb{E}|S_n=X_1+X_2+...+X_n|$
10
8
6
4
2
2
4
6
8
10
n
Cauchy (κ=1)
Pareto 1.14
Cubic Student T
Gaussian (κ=0)
Degrees of
Fat Tailedness

Behavior of sums **before** the limit

Definition 1 (the κ metric). *Let $X_1, \ldots, X_n$ be i.i.d. random variables with finite mean, that is $\mathbb{E}(X) < +\infty$. Let $S_n = X_1 + X_2 + \ldots + X_n$ be a partial sum. Let $\mathbb{M}(n) = \mathbb{E}(|S_n - \mathbb{E}(S_n)|)$ be the expected mean absolute deviation from the mean for n summands. Define the "rate" of convergence for n additional summands starting with n_0:*

$$\kappa_{n_0,n} = \min\left\{\kappa_{n_0,n} : \frac{\mathbb{M}(n)}{\mathbb{M}(n_0)} = \left(\frac{n}{n_0}\right)^{\frac{1}{2-\kappa_{n_0,n}}}, n_0 = 1, 2, \ldots\right\},$$

κ_n is indicative of both the rate of convergence under the law of large number, and for $\kappa_n \to 0$, for rate of convergence of summands to the Gaussian under the central limit, as illustrated in Figure 7.2.

- Another way to “measure” both CLT & speed of LLN

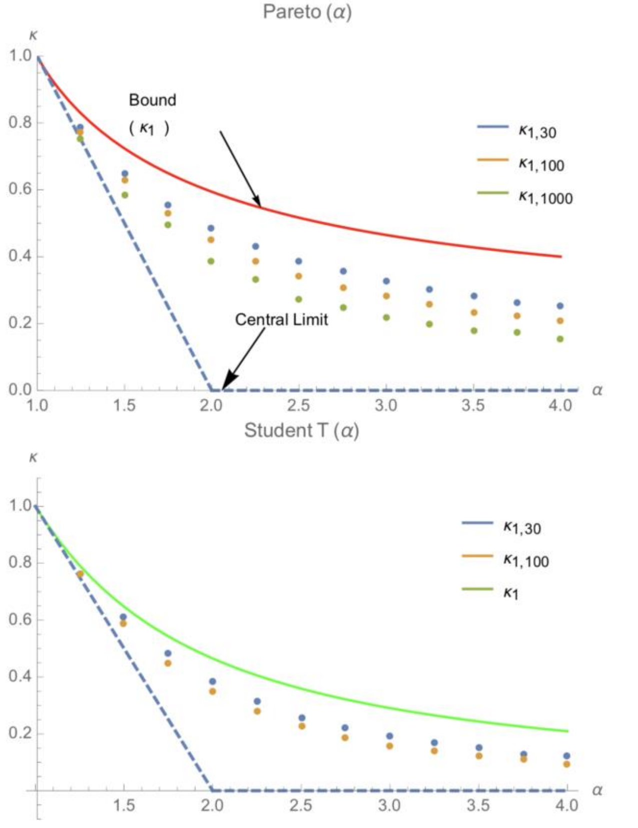
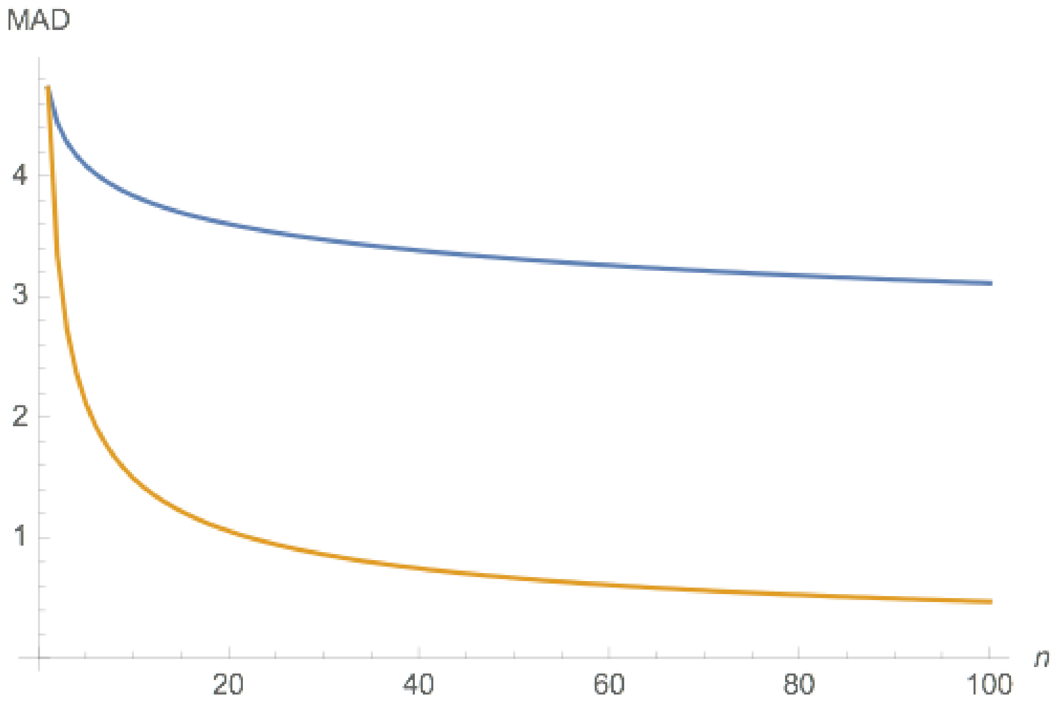
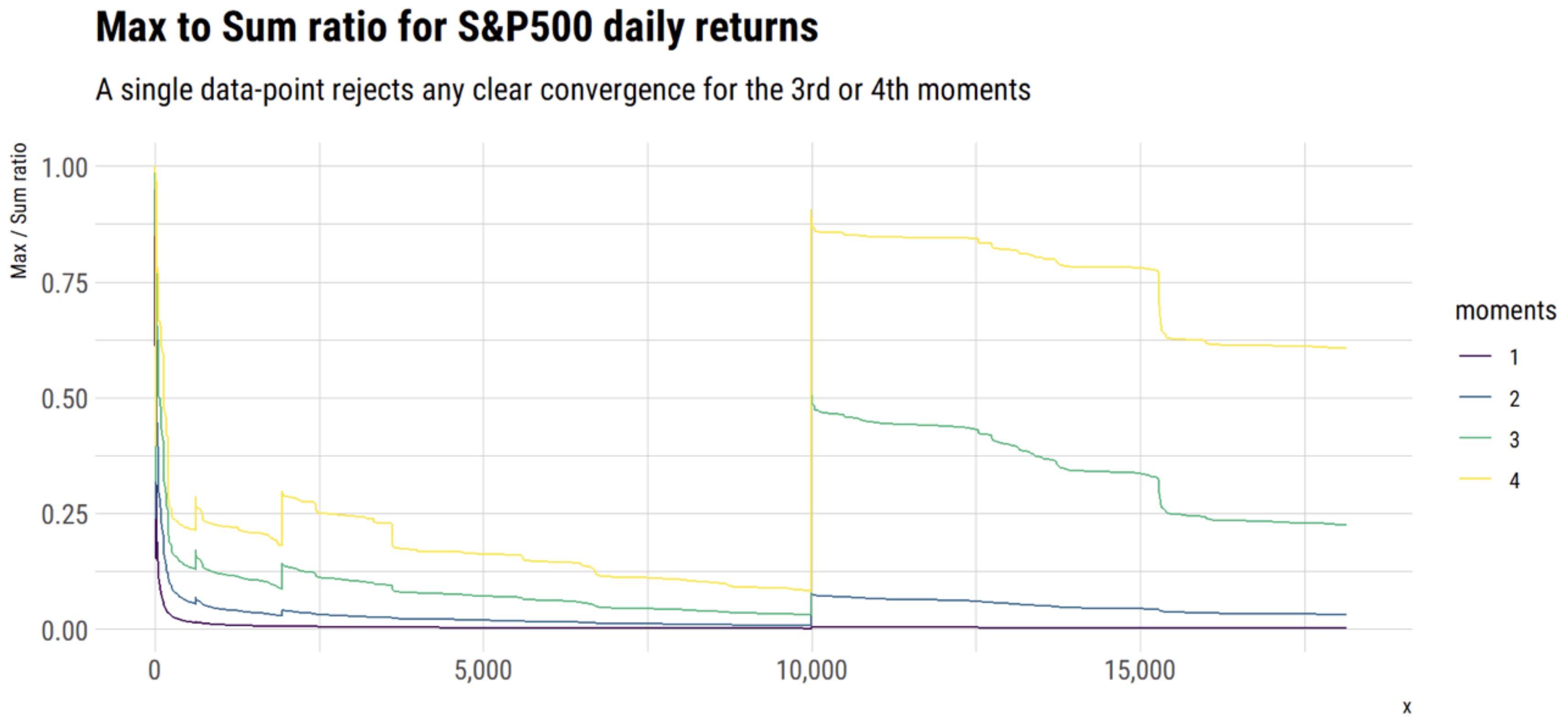


Fig. 2. Watching the effect of the Generalized Central Limit Theorem: Pareto and Student T Distribution, in the $\mathfrak{P}$ class, with α exponent, κ converge to

kappa and Portfolio "Risk"

Speed of statistical inference (number of summands) and diversification effects are same.




Max to Sum ratio for S&P500 daily returns
A single data-point rejects any clear convergence for the 3rd or 4th moments
Max / Sum ratio
1.00
0.75
0.50
0.25
0.00
0
5,000
10,000
15,000
x
moments
1
2
3
4

Stable Dist Equiv

A. Equivalence for stable distributions

For all n_0 and $n \geq 1$ in the Stable $\mathfrak{S}$ class with $\tilde{\alpha} \geq 1$:

$$\kappa_{(n_0,n)} = 2 - \tilde{\alpha},$$

simply from the property that

$$\mathbb{M}(n) = n^{\frac{1}{\alpha}} \mathbb{M}(1)$$

WITTGENSTEIN'S RULER: WAS IT REALLY A "10 SIGMA EVENT"?

IN THE SUMMER OF 1998, the hedge fund called "Long Term Capital Management" (LTCM) proved to have a very short life; it went bust from some deviations in the markets –those "of an unexpected nature". The loss was a *yuuuge deal* because two of the partners received the Swedish Riksbank Prize, marketed as the "Nobel" in economics. More significantly, the fund harbored a large number of finance professors; LTCM had imitators among professors (at least sixty finance PhDs blew up during that period from trades similar to LTCM's, and owing to risk management methods that were identical). At least two of the partners made the statement that it was a "10 sigma" event (10 standard deviations), hence they should be absolved of all accusations of incompetence (I was first hand witness of two such statements).

Let us apply what the author calls "Wittgenstein's ruler": are you using the ruler to measure the table or using the table to measure the ruler?

Assume to simplify there are only two alternatives: a Gaussian distribution and a Power Law one. For the Gaussian, the "event" we define as the survival function of 10 standard deviations is 1 in 1.31×10^{-23}. For the Power law of the same scale, a student T distribution with tail exponent 2, the survival function is 1 in 203.

What is the probability of the data being Gaussian conditional on a 10 sigma event, compared to that alternative?

We start with Bayes' rule. $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$. Replace $P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A})$ and apply to our case.

$$P(\text{Gaussian}|\text{Event}) = \frac{P(\text{Gaussian})P(\text{Event}|\text{Gaussian})}{(1 - P(\text{Gaussian}))P(\text{Event}|\text{NonGaussian}) + P(\text{Gaussian})P(\text{Event}|\text{Gaussian})}$$

P(Gaussian)	P(Gaussian|Event)
0.5	2×10^{-21}
0.999	2×10^{-18}
0.9999	2×10^{-17}
0.99999	2×10^{-16}
0.999999	2×10^{-15}
1	1

Moral: If there is a tiny probability, $< 10^{-10}$ that the data might not be Gaussian, one can firmly reject Gaussianity in favor of the fat tailed distribution. The heuristic is to reject Gaussianity in the presence of any event > 4 or > 5 STDs –we will see throughout the book why patches such as conditional variance are inadequate and can be downright fraudulent.[a]

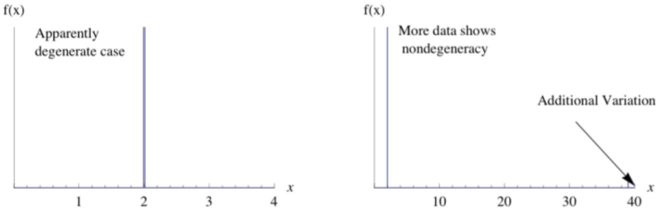


Figure 2.13: ***The Masquerade Problem (or Central Asymmetry in Inference)***. *To the left, a degenerate random variable taking seemingly constant values, with a histogram producing a Dirac stick. One cannot rule out nondegeneracy. But the right plot exhibits more than one realization. Here one can rule out degeneracy. This central asymmetry can be generalized and put some rigor into statements like "failure to reject" as the notion of what is rejected needs to be refined. We produce rules in Chapter* **??**.

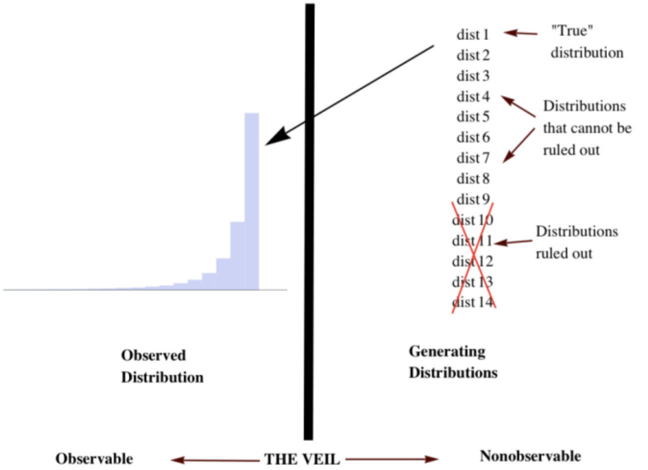


Figure 2.14: ***"The probabilistic veil"***. *Taleb and Pilpel [228] cover the point from an epistemological standpoint with the "veil" thought experiment by which an observer is supplied with data (generated by someone with "perfect statistical information", that is, producing it from a generator of time series). The observer, not knowing the generating process, and basing his information on data and data only, would have to come up with an estimate of the statistical properties (probabilities, mean, variance, value-at-risk, etc.). Clearly, the observer having incomplete information about the generator, and no reliable theory about what the data corresponds to, will always make mistakes, but these mistakes have a certain pattern. This is the central problem of risk management.*

Consequence 8

The gap between disconfirmatory and confirmatory empiricism is wider than in situations covered by common statistics i.e., the difference between absence of evidence and evidence of absence becomes larger.

Consequence 2

The mean of the distribution will rarely correspond to the sample mean; it will have a persistent small sample effect (downward or upward) particularly when the distribution is skewed (or one-tailed).

This is another problem of sample insufficiency. In fact, there is no very fat tailed- one tailed distribution in which the population mean can be properly estimated directly from the sample mean –rare events determine the mean, and these, *being rare*, take a lot of data to show up[10]. Consider that some power laws (like the one described as the "80/20" in common parlance have 92 percent of the observations falling below the true mean). For the sample average to be informative, we need orders of magnitude more data than we do (people in economics still do not understand this, though traders have an intuitive grasp of the point). The problem is discussed briefly further down in 2.7, and more formally in the "shadow mean" chapters, Chapters 13 and 14. Further, we will introduce the notion of hidden properties are in 2.7. Clearly by the same toke, variance will be likely to be underestimatwd.

Consider K_n the maximum of a sample of n independent identically distributed variables in the power law class; $K_n = \max(X_1, X_2, \ldots, X_n)$. Let $\phi(.)$ be the density of the underlying distribution. We can decompose the moments in two parts, with the "hidden" moment above K_0, as shown in Fig 9.6:

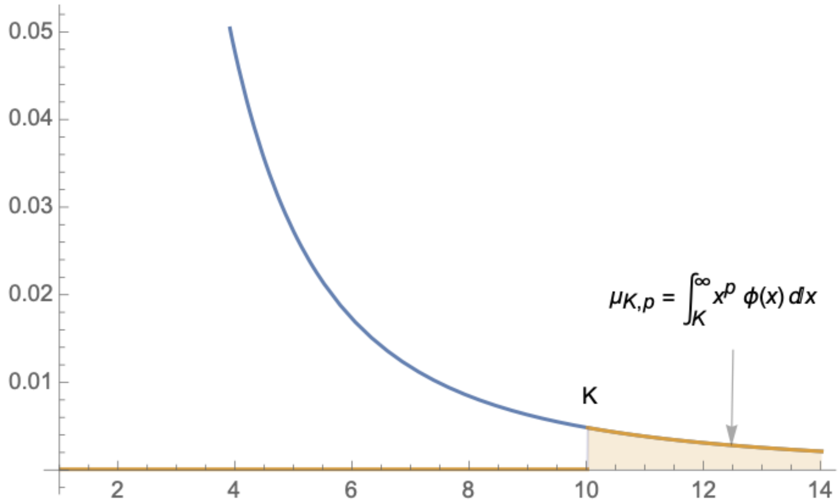


Figure 9.6: *The p^{th} moment above K*

$$\mathbb{E}(X^p) = \underbrace{\int_L^{K_n} x^p \phi(x)\, dx}_{\mu_{0,p}} + \underbrace{\int_{K_n}^{\infty} x^p \phi(x)\, dx}_{\mu_{K,p}}$$

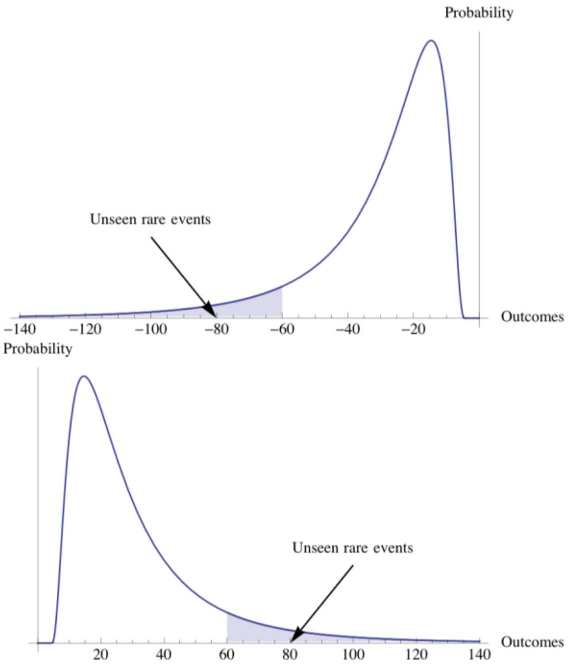


Figure 2.26: *Shadow Mean at work: Below: Inverse Turkey Problem –Ť The unseen rare event is positive. When you look at a positively skewed (antifragile) time series and make (nonparametric) inferences about the unseen, you miss the good stuff an underestimate the benefits. Above: The opposite problem. The filled area corresponds to what we do not tend to see in small samples, from insufficiency of data points. Interestingly the shaded area increases with model error.*

Table 2.4: *Shadow mean, sample mean and their ratio for different minimum thresholds. In bold the values for the 145k threshold. Rescaled data. From Cirillo and Taleb [46]*

Thresh.$\times 10^3$	Shadow$\times 10^7$	Sample$\times 10^7$	Ratio
50	1.9511	1.2753	1.5299
100	2.3709	1.5171	1.5628
145	**3.0735**	**1.7710**	**1.7354**
300	3.6766	2.2639	1.6240
500	4.7659	2.8776	1.6561
600	5.5573	3.2034	1.7348

Consequence 5

Robust statistics is not robust and the empirical distribution is not empirical.

Crash Beliefs From Investor Surveys

William N. Goetzmann — *Yale School of Management, Yale University*

Dasol Kim — *Weatherhead School of Management, Case Western Reserve University*

Robert J. Shiller — *Yale University*

Draft: March 19, 2016

Please do not quote without permission

Abstract: Historical data suggest that the base rate for a severe, single-day stock market crash is relatively low. Surveys of individual and institutional investors, conducted regularly over a 26 year period in the United States, show that they assess the probability to be much higher. We examine the factors that influence investor responses and test the role of media influence. We find evidence consistent with an availability bias. Recent market declines and adverse market events made salient by the financial press are associated with higher subjective crash probabilities. Non-market-related, rare disasters are also associated with higher subjective crash probabilities.

Keywords: Crash Beliefs, Availability Bias, Investor Surveys

JEL: G00, G11, G23, E03, G02

Figure 8.2: *The base rate fallacy, revisited —or, rather in the other direction. The "base rate" is an empirical evaluation that bases itself on the worst past observations, an error identified in [212] as the fallacy identified by the Roman poet Lucrecius in De rerum natura of thinking the tallest future mountain equals the tallest on has previously seen. Quoted without permission after warning the author.*

Consequence 3

Standard deviations and variance are not useable.

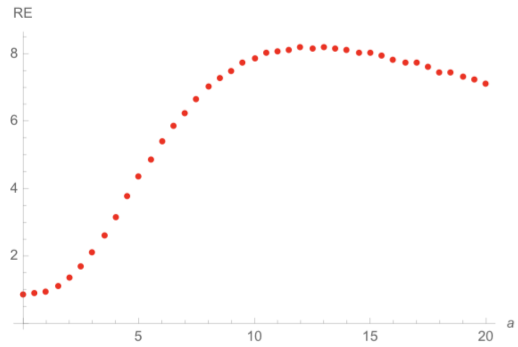


Figure 3.8: *A simulation of the Relative Efficiency ratio of Standard deviation over Mean deviation when injecting a jump size $\sqrt{(1+a)} \times \sigma$, as a multiple of σ the standard deviation.*

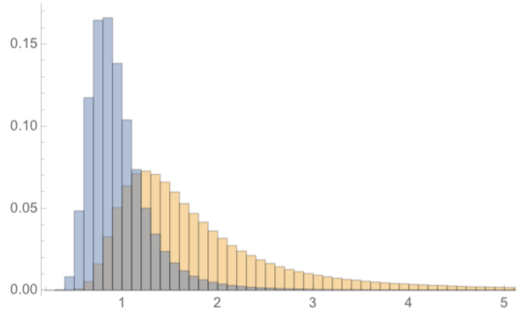


Figure 3.9: *Mean deviation vs standard deviation for a finite variance power law. The result is expected (MD is the thinner distribution), complicated by the fact that standard deviation has an infinite variance since the square of a Paretan random variable with exponent α is Paretan with an exponent of $\frac{1}{2}\alpha$. In this example the mean deviation of standard deviation is 5 times higher.*

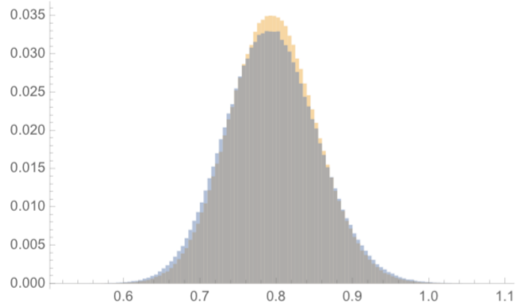


Figure 3.10: *For a Gaussian, there is small difference between MD and STD.*

Consequence 4

Beta, Sharpe Ratio and other common hackneyed financial metrics are uninformative.

Table D.1: *Maximum contribution to the fourth moment from a single daily observation*

Security	Max Q	Years.
Silver	0.94	46.
SP500	0.79	56.
CrudeOil	0.79	26.
Short Sterling	0.75	17.
Heating Oil	0.74	31.
Nikkei	0.72	23.
FTSE	0.54	25.
JGB	0.48	24.
Eurodollar Depo 1M	0.31	19.
Sugar #11	0.3	48.
Yen	0.27	38.
Bovespa	0.27	16.
Eurodollar Depo 3M	0.25	28.
CT	0.25	48.
DAX	0.2	18.

Consequence 6

Linear least-square regression doesn't work (failure of the Gauss-Markov theorem).

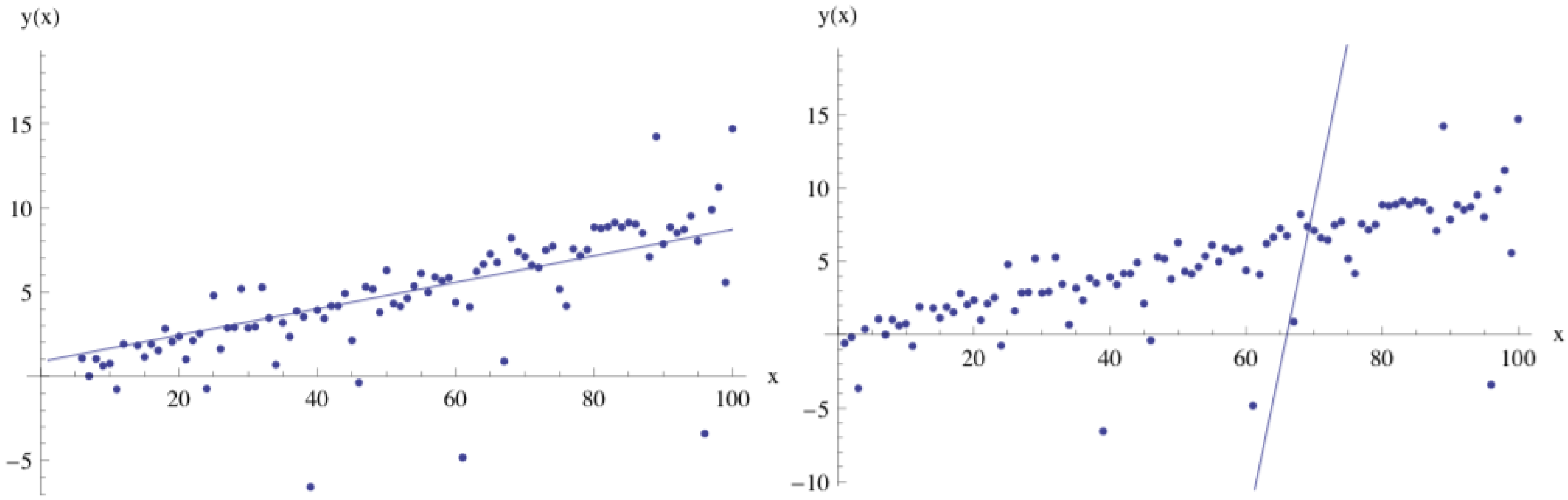


Figure 2.4: *In the presence of fat tails, we can fit markedly different regression lines to the same story (the Gauss-Markov theorem doesn't apply anymore). Left: a regular (naive) regression. Right: a regression line that tries to accommodate the large deviation —a "hedge ratio" so to speak, one that protects the agent from a large deviation, but mistracks small ones. Missing the largest deviation can be fatal. Note that the sample doesn't include the critical observation, but it has been guessed using "shadow mean" methods.*

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n} (y_i - (a x_i + b + \epsilon_i))^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

We can show that, for large n

$$R^2 = \frac{a^2}{a^2 + \mathbb{E}(\epsilon_i^2)} + O\left(\frac{1}{n^2}\right). \quad (5.11)$$

And of course, for infinite variance:

$$\lim_{E(\epsilon^2) \to +\infty} \mathbb{E}(R^2) = 0.$$

When ϵ is T-distributed with α degrees of freedom, clearly ϵ^2 will follow an FRatio distribution $(1, \alpha)$ –a power law with exponent $\frac{\alpha}{2}$.

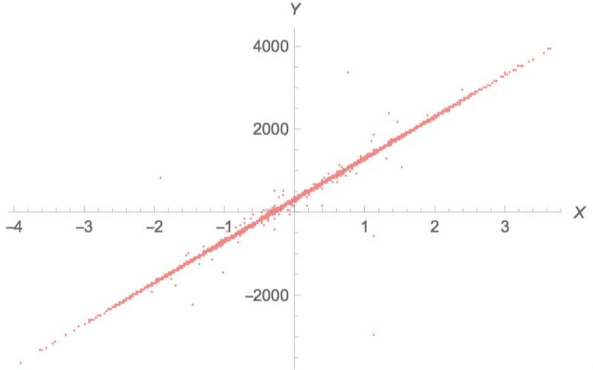


Figure 5.12: *A Cauchy regression with an expected* $R^2 = 0$*, faking it but showing higher values in small samples (here .985).*

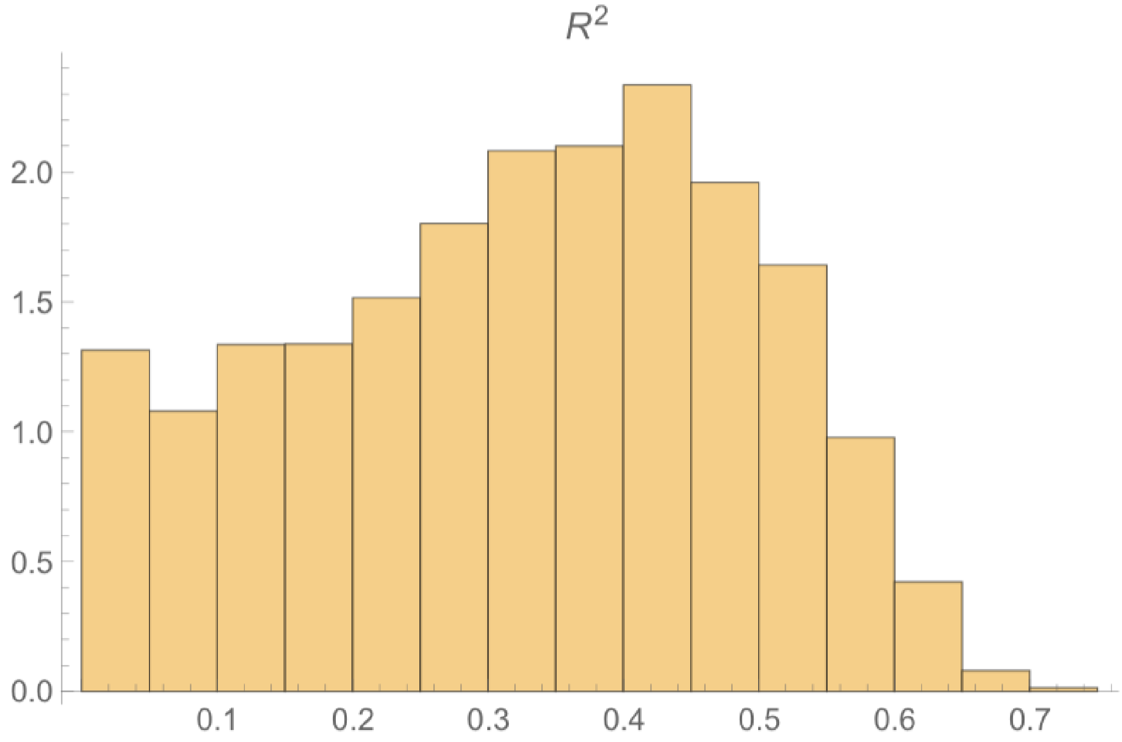


Figure 5.11: *An infinite variance case that shows a high R^2 in sample; but it ultimately has a value of 0. Remember that R^2 is stochastic. The problem greatly resembles that of P values in Chapter 17 owing to the complication of a metadistribution in* $[0,1]$

Consequence 9

Principal component analysis (PCA) and factor analysis are likely to produce spurious factors and loads.

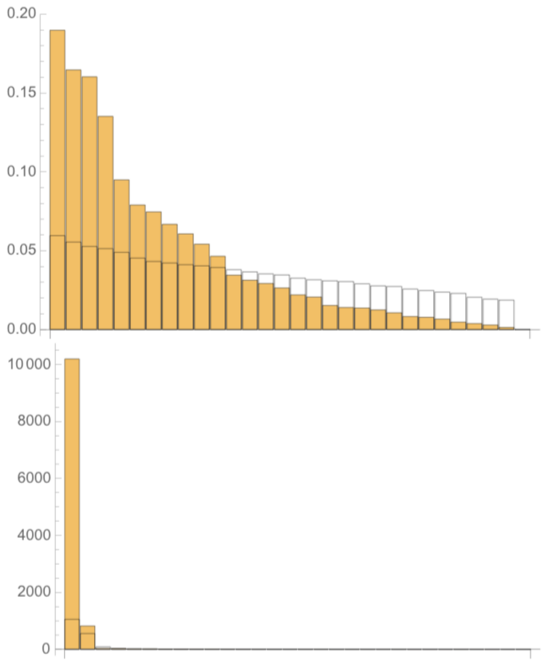


Figure 2.24: ***Spurious PCAs Under Fat Tails:*** *A Monte Carlo experiment that shows how spurious correlations and covariances are more acute under fat tails. Principal Components ranked by variance for 30 Gaussian uncorrelated variables, n=100 (above) and 1000 data points, and principal components ranked by variance for 30 Stable Distributed (with tail $\alpha = \frac{3}{2}$, symmetry $\beta = 1$, centrality $\mu = 0$, scale $\sigma = 1$) (below). Both are "uncorrelated" identically distributed variables. We can see the "flatter" PCA structure with the Gaussian as n increases (the difference between PCAs shrinks). Such flattening does not occur in reasonable time under fatter tails.*

Consequence 7

Maximum likelihood methods can work well for some parameters of the distribution (good news).


STOP