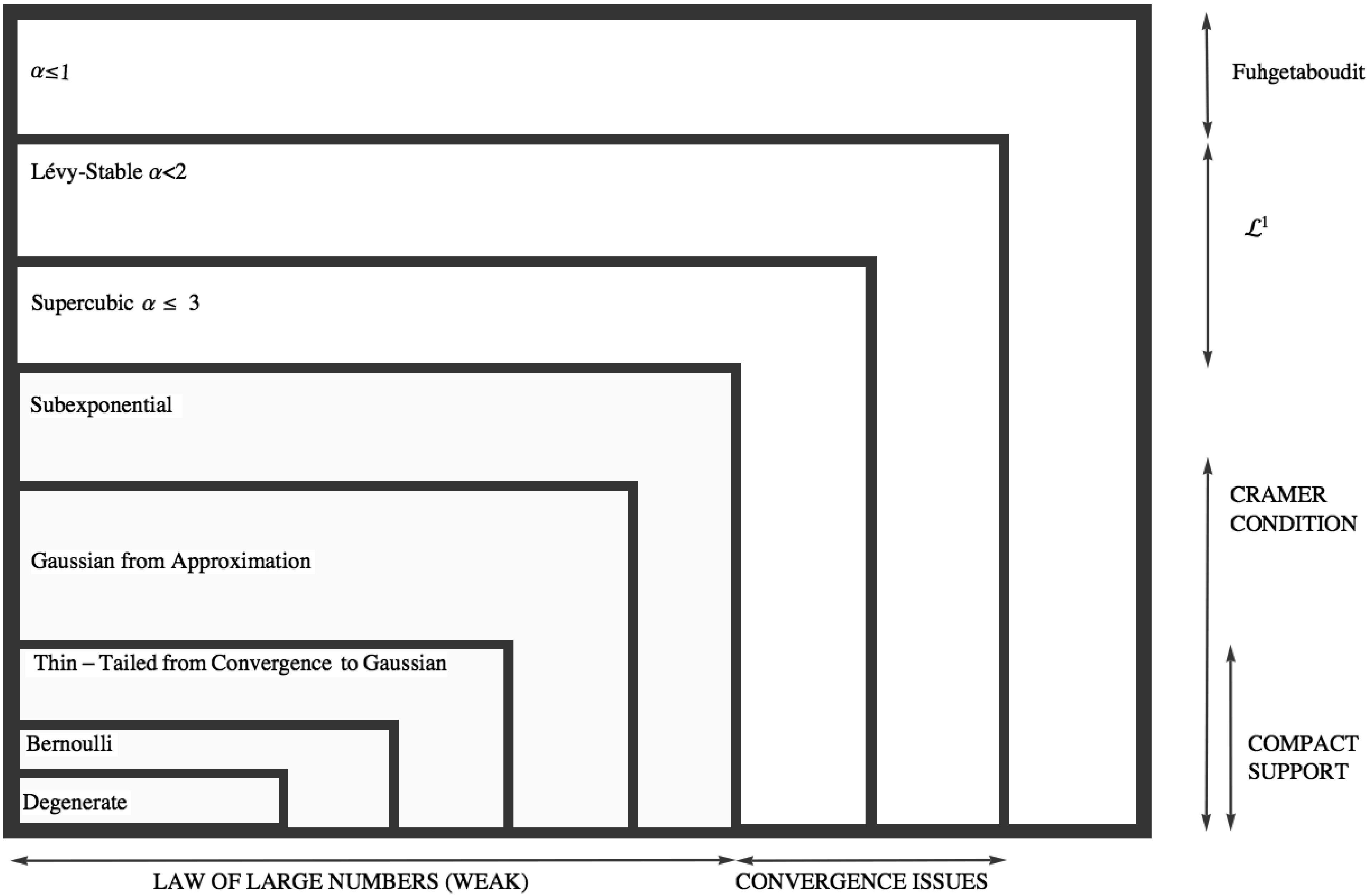
PRE-ASYMPTOTIC MEASURE OF FAT TAILEDNESS

Nassim Nicholas Taleb
Tandon School of Engineering, NYU



$\alpha \leq 1$
Lévy-Stable $\alpha<2$
Supercubic $\alpha \leq 3$
Subexponential
Gaussian from Approximation
Thin – Tailed from Convergence to Gaussian
Bernoulli
Degenerate
Fuhgetaboudit
$\mathcal{L}^1$
CRAMER CONDITION
COMPACT SUPPORT
LAW OF LARGE NUMBERS (WEAK)
CONVERGENCE ISSUES

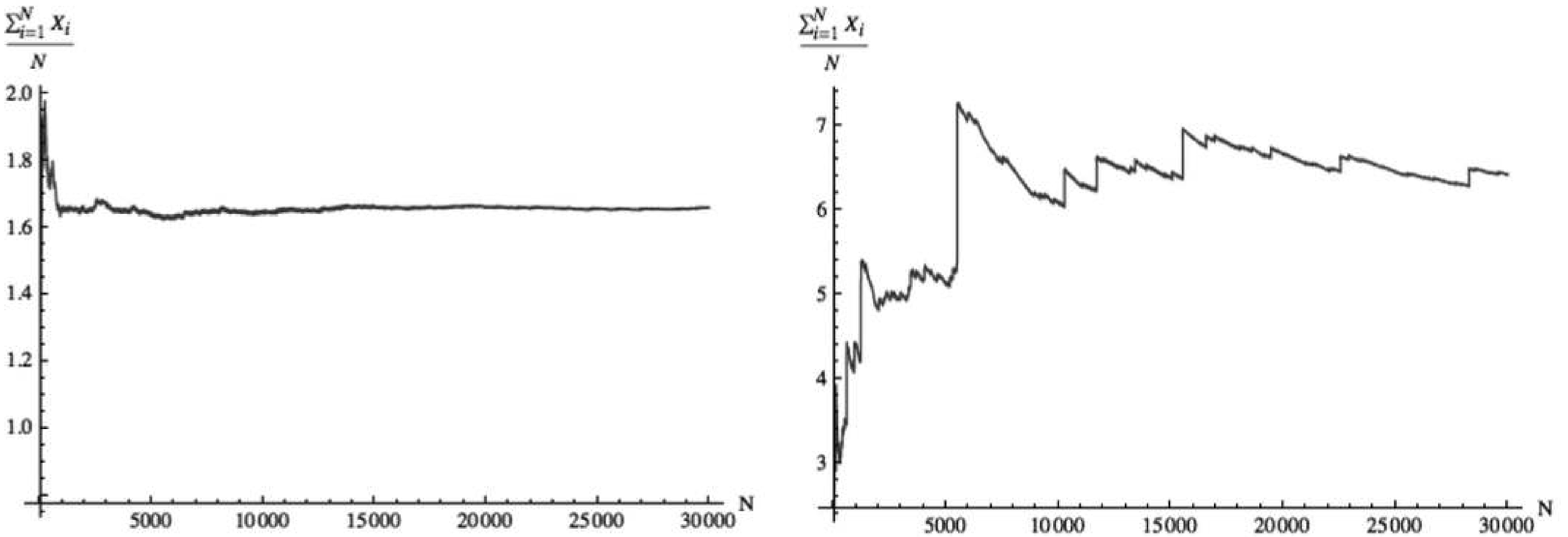


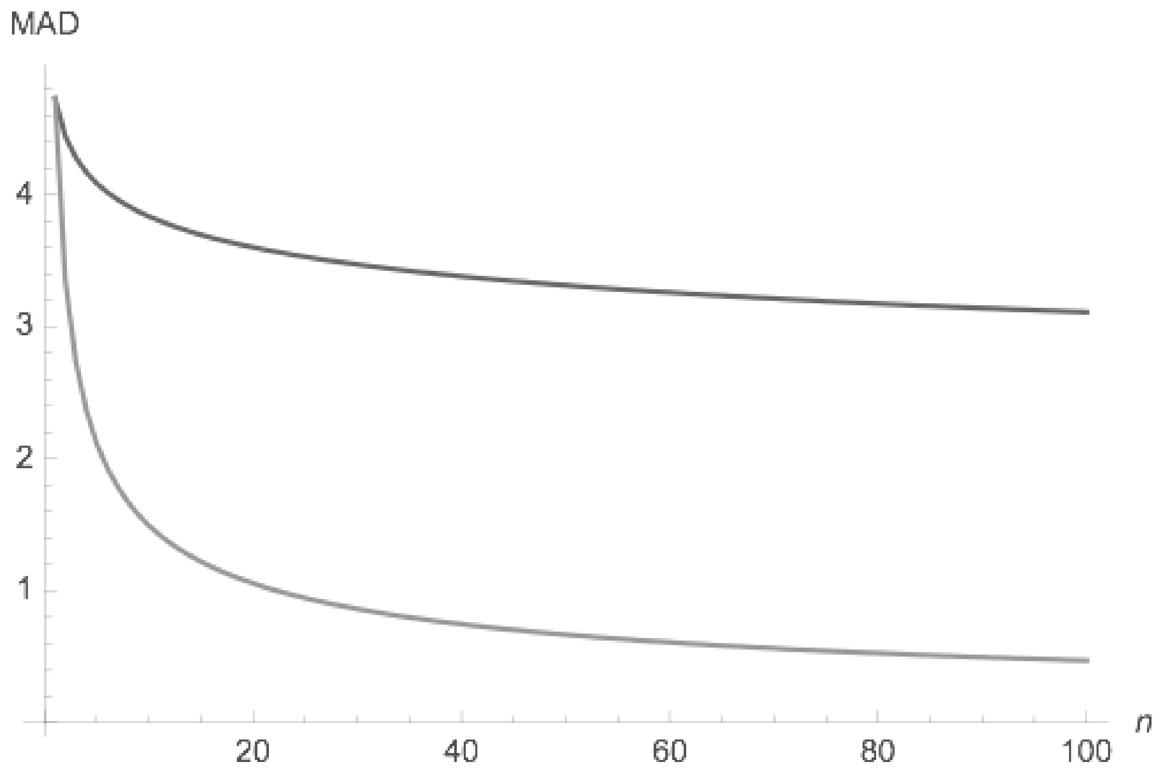
Figure 7.1: How thin tails (Gaussian) and fat tails ($1 < \alpha \leq 2$) converge to the mean.

Table 7.1: Corresponding n_α, or how many for equivalent α-stable distribution. The Gaussian case is the $\alpha = 2$. For the case with equivalent tails to the 80/20 one needs 10^{11} more data than the Gaussian.

α	n_α	$n_\alpha^{\beta=\pm\frac{1}{2}}$	$n_\alpha^{\beta=\pm 1}$
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

kappa and Portfolio “Risk”

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



Measures of Fat Tailedness

MAIN

- NONPARETAN CLASS (Finite Moments): Kurtosis

- PARETAN CLASS: Tail exponent

Other:

- GINI/concentration measures
- Quantile contribution
- Other

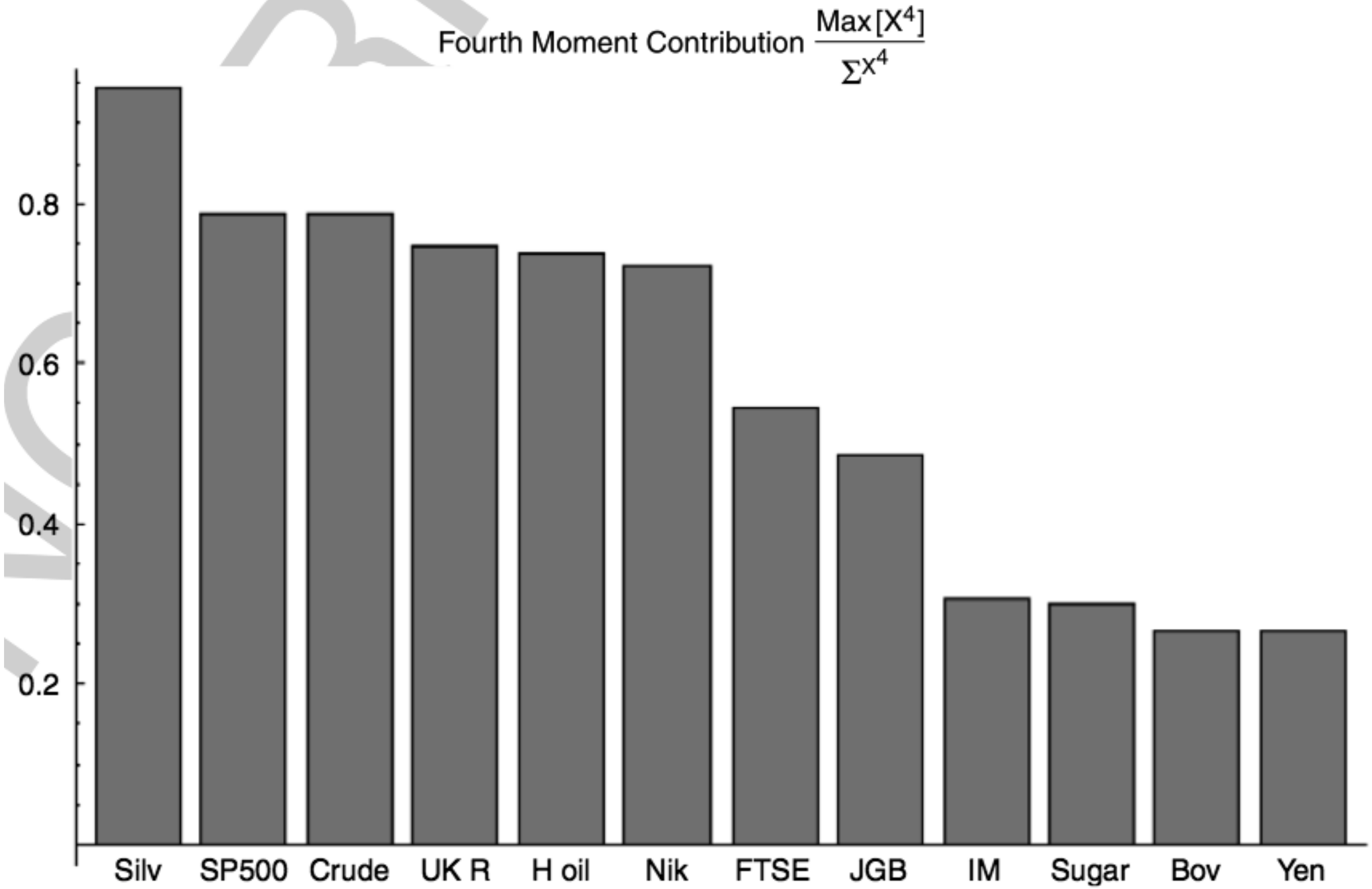


Fig. 3. A selection of 12 most acute cases among the 43 economic variables.

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**

Behavior of sums **before** the limit

Definition 1. *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 $MD(n) = \mathbb{E}(|S_n - \mathbb{E}(S_n)|)$ be the expected mean absolute deviation from the mean for n summands. Define the "speed" of convergence for n summands:*

$$\kappa_{n_0,n} = \left\{ \kappa_{n_0,n} : \frac{ME(n_0+n)}{ME(n_0)} = n^{\frac{1}{\kappa_{n_0,n}}}, n = 1, 2, ... \right\} \tag{1}$$

Further, for the default value $n_0 = 1$ we write the unsummed initial value $\kappa_{1,n}$ as κ_n.

Stable Dist Equiv

Remark 1 (Equality for stable distributions). *We note that* $\kappa_{(\cdot,\cdot)} = \alpha$ *for all* n_0 *and* n *in the Stable* $\mathfrak{S}$ *class.*

obtain a factorized σ, and since the scale for n summands is $n^{\frac{1}{\alpha}}\sigma$. □

Generalized Central Limit

$$\tilde{\alpha} \triangleq \begin{cases} \alpha\, \mathbb{1}_{\alpha<2} + 2\, \mathbb{1}_{\alpha \geq 2} & \text{if X is in } \mathfrak{P} \\ 2 & \text{otherwise.} \end{cases}$$

The problem of the preasymptotics for n summands reduces to:

- What is the property of the distribution for $n = 1$?
- What is the property of the distribution for n summands?
- How does $\kappa_n \to \tilde{\alpha}$ and at what speed?

Why MAD not STD

- We are interested in class of finite mean not necessarily finite variance.
- MAD more “natural”
- MAD more “efficient” except for non Gaussian
- MAD only “efficient” asymptotically

  hence not for ***finite*** “small” *n*

Some History of STD

Why The [CENSORED] did statistical science pick Mean Deviation over STD? Here is the story, with analytical derivations not seemingly available in the literature. In Huber [28]:

> There had been a dispute between Eddington and Fisher, around 1920, about the relative merits of dn (mean deviation) and Sn (standard deviation). Fisher then pointed out that for **exactly normal** observations, Sn is 12% more efficient than dn, and this seemed to settle the matter. (My emphasis)

Let us rederive and see what Fisher meant.

Let N be the number of summands:

$$\text{Asymptotic Relative Efficiency} = \lim_{N\to\infty}\left(\frac{\mathbb{V}(Std)}{\mathbb{E}(Std)^2}\Big/\frac{\mathbb{V}(Mad)}{\mathbb{E}(Mad)^2}\right)$$

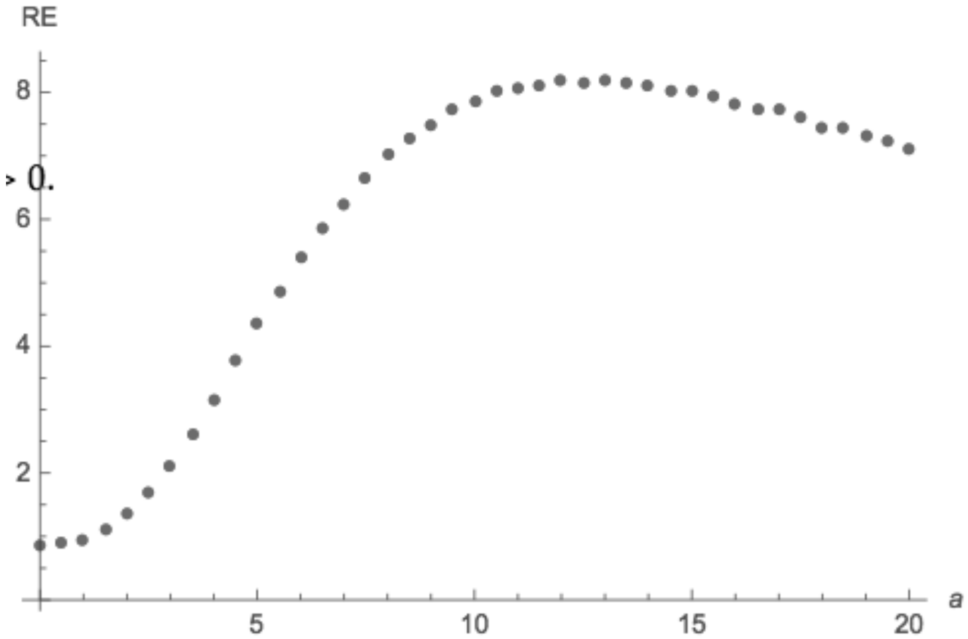
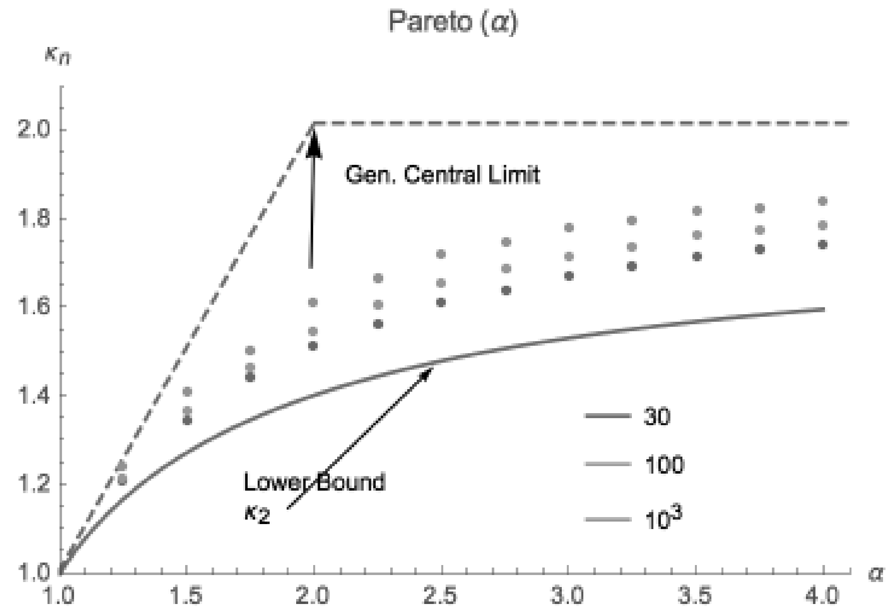


Figure 1.8: A simulation of the Relative Efficiency ratio of Standard deviation over Mean deviation when injecting a jump size $(1+a)\sigma$, as a multiple of σ the standard deviation.

Finalemente, the Asymptotic Relative Efficiency For a Gaussian

$$\text{ARE} = \lim_{N\to\infty}\frac{N\left(\frac{N\Gamma\left(\frac{N}{2}\right)^2}{\Gamma\left(\frac{N+1}{2}\right)^2}-2\right)}{\pi-2} = \frac{1}{\pi-2} \approx .875$$

- Another way to "measure" both CLT & speed of LLN



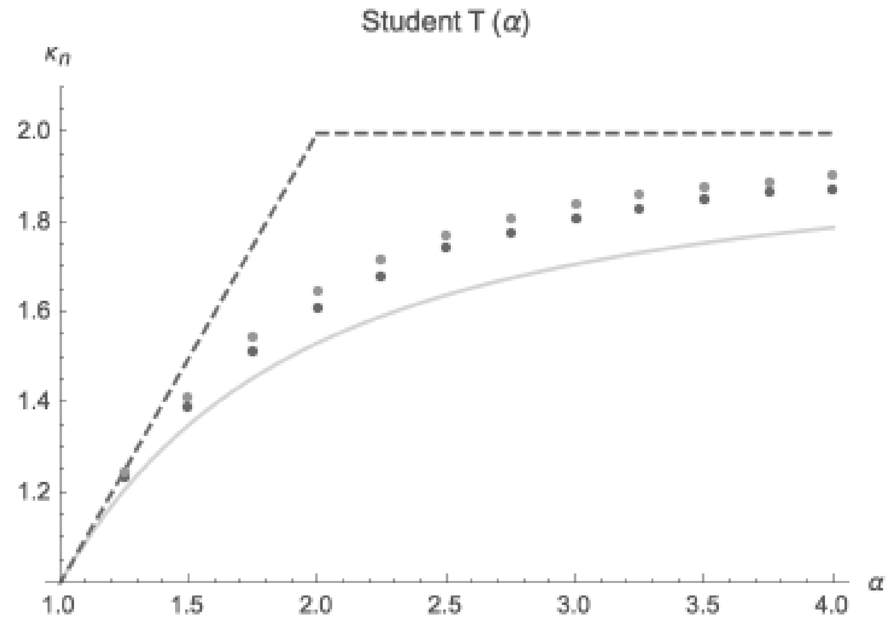


TABLE III
COMPARING PARETO TO STUDENT T

α	Pareto κ_2	Pareto κ_{30}	Pareto κ_{100}	Student κ_2	Student κ_{30}	Student κ_{100}
1.25	1.171	1.113	1.229	1.208	1.235	1.244
1.5	1.276	1.350	1.369	1.353	1.391	1.413
1.75	1.35	1.444	1.47	1.457	1.517	1.549
2.	1.406	1.516	1.551	1.535	1.613	1.648
2.25	1.449	1.569	1.612	1.594	1.684	1.718
2.5	1.483	1.614	1.659	1.641	1.744	1.773
2.75	1.512	1.644	1.693	1.679	1.776	1.811
3.	1.535	1.6754	1.719	1.71	1.809	1.841
3.25	1.555	1.695	1.742	1.735	1.833	1.862
3.5	1.572	1.716	1.765	1.757	1.851	1.879
3.75	1.587	1.737	1.778	1.775	1.870	1.891
4.	1.6	1.7468	1.789	1.791	1.874	1.907

Results

Distribution	κ_n
Exponential/Gamma	Explicit
Lognormal (μ, σ)	No explicit κ_n but explicit lower bound (low or high σ or n). Approximated with Pearson IV for σ in between.
Pareto (α) (Constant)	Explicit for κ_2 (lower bound for all α).
Student $T(\alpha)$ (slowly varying function)	Explicit for κ_2 , $\alpha = 3$.

TABLE I
INDEX OF "FAT-TAILEDNESS" FOR NO-SUMMANDS, κ_1.

Distribution	κ_2
Student T (α)	$\frac{\log(4)}{2\log\left(\frac{2^{2-\alpha}\Gamma\left(\alpha-\frac{1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)^2}\right)+\log(\pi)}$
Exponential/Gamma	$\frac{\log(2)}{\log(4)-1} \approx 1.79$
Pareto (α)	$\frac{\log(2)}{\log\left((\alpha-1)^{2-\alpha}\alpha^{\alpha-1}\int_0^{\frac{2}{\alpha-1}} -2\alpha^2(y+2)^{-2\alpha-1}\left(\frac{2}{\alpha-1}-y\right)\left(B_{\frac{1}{y+2}}(-\alpha,1-\alpha)-B_{\frac{y+1}{y+2}}(-\alpha,1-\alpha)\right)dy\right)}$
Normal (μ,σ) with switching variance $\sigma^2 a$ w.p p.	$\frac{\log(2)}{\log\left(\frac{\sqrt{2}\left(\sqrt{\frac{ap}{p-1}+\sigma^2}+p\left(-2\sqrt{\frac{ap}{p-1}+\sigma^2}+p\left(\sqrt{\frac{ap}{p-1}+\sigma^2}-\sqrt{2a\left(\frac{1}{p-1}+2\right)+4\sigma^2}+\sqrt{a+\sigma^2}\right)+\sqrt{2a\left(\frac{1}{p-1}+2\right)+4\sigma^2}\right)\right)}{p\sqrt{a+\sigma^2}-(p-1)\sqrt{\frac{ap}{p-1}+\sigma^2}}\right)}$
Lognormal (μ,σ)	$\approx \frac{\log(2)}{\log\left(\frac{2\,\mathrm{Erf}\left(\frac{\sqrt{\log\left(\frac{1}{2}\left(e^{\sigma^2}+1\right)\right)}}{2\sqrt{2}}\right)}{\mathrm{Erf}\left(\frac{\sigma}{2\sqrt{2}}\right)}\right)}.$

E. A Pareto is not a Stable Distribution

In Uchaikin and Zolotarev [3]:

> Mandelbrot called attention to the fact that the use of the extremal stable distributions (corresponding to $\beta = 1$) to describe empirical principles was preferable to the use of the Zipf-Pareto distributions for a number of reasons. It can be seen from many publications, both theoretical and applied, that Mandelbrot's ideas receive more and more wide recognition of experts. In this way, the hope arises to confirm empirically established principles in the framework of mathematical models and, at the same time, to clear up the mechanism of the formation of these principles.

These are not the same animals, even for large number of summands.

Problems in literature

- Tendency to treat all tail exponents >2 as Gaussian

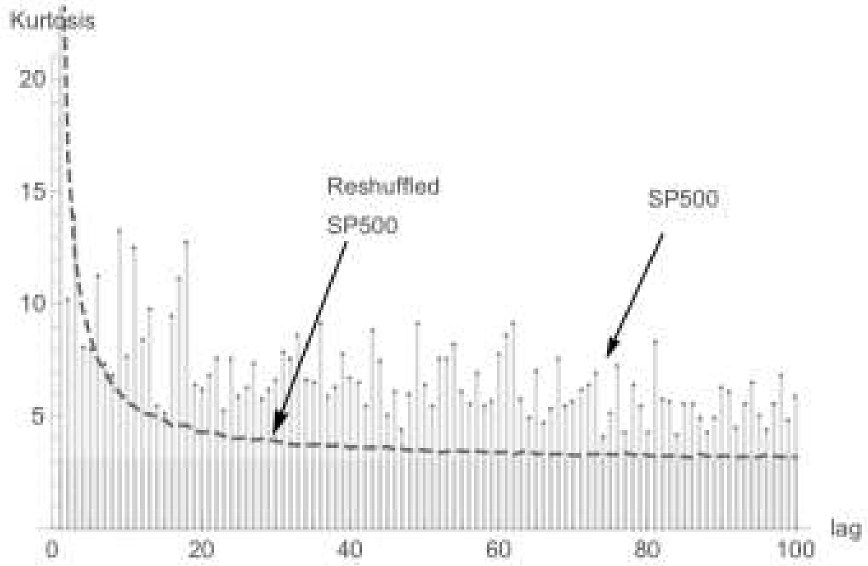


Fig. 8. Visual convergence diagnostics for the Kurtosis of the SP500 over the past 17000 observations. We compute the kurtosis at different lags for the raw SP500 and reshuffled data. While the 4^{th} norm is not convergent for raw data, it is clearly so for the reshuffled series. We can thus assume that the "fat tailedness" is attributable to the temporal structure of the data, particularly the clustering of its volatility. See Table I for the expected drop at speed $1/n^2$ for thin-tailed distributions.

Lognormal Bounds

- A lognormal with high σ stays lognormal under summation
- A lognormal with low σ behaves like normal

$$\kappa_{1,n}^{*} \leq \kappa_{1,n} \leq 2$$

$$\kappa_{1,n} = \frac{n \log}{\frac{\log\left(\operatorname{erf}\sqrt{\log\left(n+e^{\sigma^2}-1\right)-n\log}\right)}{2\sqrt{2}} - \frac{\log(\operatorname{erf}\sigma)}{2\sqrt{2}} + n\log}$$

$$\lim_{\sigma \to \infty} \kappa_{1,n} = \kappa_{1,n}^{*}$$

$$\lim_{n \to \infty} \kappa_{1,n} = 2$$

Infinity Shminfinity

- For a Lognormal, “X large but not infinity” is even more ambiguous.
- For a lognormal, “σ low is ambiguous”

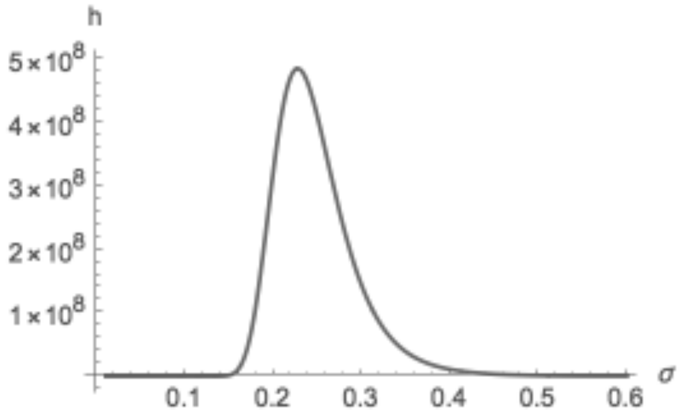
SubExponentiality and the Confusing Lognormal

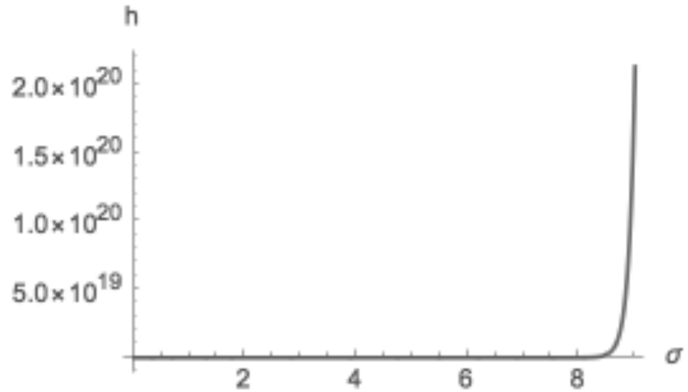
Recall the story of “two randomly selected people are 4.4 meters tall in total. What is the most likely distribution?” Clearly, 2.2 meters each. But “two randomly selected people have $30 million net worth. What is the most likely breakdown?” Here .1 and 29.9. This is another way to express the “basin” differences.

The Lognormal is again ambiguous. Take the ratio $h := \frac{P(X > n\, s\, x)}{P(X > s\, x)^n}$, with n=2 and s is the standard deviation of the Lognormal, here $s = \sqrt{e^{\sigma^2}\left(e^{\sigma^2} - 1\right)}$

$$h = \frac{2\, \text{Erfc}\left[\frac{\text{Log}\left[2\, e^{\frac{\sigma^2}{2}} \sqrt{-1+e^{\sigma^2}}\; x\right]}{\sqrt{2}\, \sigma}\right]}{\text{Erfc}\left[\frac{\text{Log}\left[e^{\frac{\sigma^2}{2}} \sqrt{-1+e^{\sigma^2}}\; x\right]}{\sqrt{2}\, \sigma}\right]^2};$$

The weirdest think is a kink at low σ, the "crossover"





A nice equality

The Pearson family is defined for an appropriately scaled density f satisfying the following differential equation.

$$f'(x) = -\frac{(a_0 + a_1 x)}{b_0 + b_1 x + b_2 x^2} f(x) \tag{3}$$

Let m be the mean. Using the identity $\mathbb{E}(|X - m|) = 2\int_m^\infty x f(x)\mathrm{d}x$ and integrating by parts,

$$\mathbb{E}(|X - m|) = -\frac{2\left(b_0 + b_1 m + b_2 m^2\right)}{a_1 - 2b_2} f(m) \tag{4}$$

Cumulants

- Simply, since

  Cumulants order p (n summands) = n cumulants order p (1).
- We match cumulants of Pearson to those of n-summed Lognormal
- Parametrization determines the Pearson "class" (here Pearson IV)

We use cumulants of the n-summed lognormal to match the parameters.Setting $a_1 = 1$, and $m = \frac{b_1 - a_0}{1 - 2b_2}$, we get

$$a_0 = \frac{-12\kappa_1\kappa_2{}^3 + 6\kappa_3\kappa_2{}^2 - 10\kappa_1\kappa_4\kappa_2 + 12\kappa_1\kappa_3{}^2 + \kappa_3\kappa_4}{2\left(6\kappa_2{}^3 + 5\kappa_4\kappa_2 - 6\kappa_3{}^2\right)}$$

$$b_2 = \frac{3\kappa_3{}^2 - 2\kappa_2\kappa_4}{-12\kappa_2{}^3 - 10\kappa_4\kappa_2 + 12\kappa_3{}^2},$$

$$b_1 = \frac{6\kappa_3\kappa_2{}^2 - 4\kappa_1\kappa_4\kappa_2 + 6\kappa_1\kappa_3{}^2 + \kappa_3\kappa_4}{2\left(6\kappa_2{}^3 + 5\kappa_4\kappa_2 - 6\kappa_3{}^2\right)},$$

$$b_0 = \frac{1}{-12\kappa_2{}^3 - 10\kappa_4\kappa_2 + 12\kappa_3^2}\left(-12\kappa_2{}^4 + 6\kappa_1\kappa_3\kappa_2^2 - 4\kappa_4\kappa_2^2 + 3\kappa_3^2\kappa_2 - 2\kappa_1^2\kappa_4\kappa_2 + 3\kappa_1^2\kappa_3^2 + \kappa_1\kappa_3\kappa_4\right)$$

Finalmente

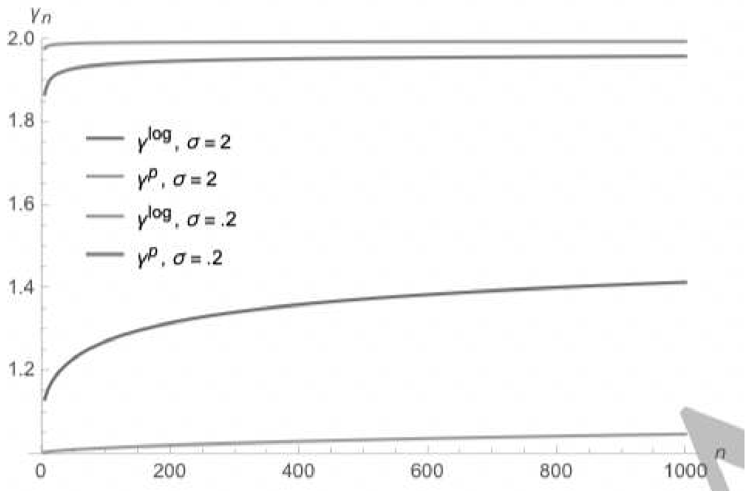


Fig. 2. The lognormal case, assuming $\mu = 0$. We show κ_n^{log} for the bounds (where the sum stays lognormal) and κ_n^{p} for the Pearson class.

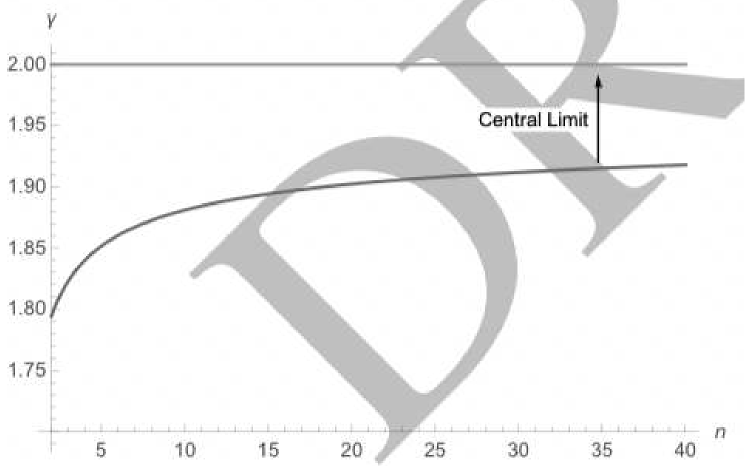


Fig. 3. Exponential Distribution

TABLE III

COMPARING PARETO TO STUDENT T

α	Pareto κ_2	Pareto κ_{30}	Pareto κ_{100}	Student κ_2	Student κ_{30}	Student κ_{100}
1.25	1.171	1.113	1.229	1.208	1.235	1.244
1.5	1.276	1.350	1.369	1.353	1.391	1.413
1.75	1.35	1.444	1.47	1.457	1.517	1.549
2.	1.406	1.516	1.551	1.535	1.613	1.648
2.25	1.449	1.569	1.612	1.594	1.684	1.718
2.5	1.483	1.614	1.659	1.641	1.744	1.773
2.75	1.512	1.644	1.693	1.679	1.776	1.811
3.	1.535	1.6754	1.719	1.71	1.809	1.841
3.25	1.555	1.695	1.742	1.735	1.833	1.862
3.5	1.572	1.716	1.765	1.757	1.851	1.879
3.75	1.587	1.737	1.778	1.775	1.870	1.891
4.	1.6	1.7468	1.789	1.791	1.874	1.907

Cubic alpha

Let X be a random variable distributed with density $p(x)$:

$$p(x) = \frac{6\sqrt{3}}{\pi\left(x^2+3\right)^2},\ x \in (-\infty, \infty) \tag{9}$$

Theorem 1. *Let Y be a sum of $X_1, \ldots, X_n$, n identical copies of X. Let $MD(n)$ be the mean absolute deviation from the mean for n summands. The "speed" of convergence $\gamma_n = \left\{\gamma : \frac{MD(n)}{MD(1)} = \left(\frac{1}{n}\right)^{1-\frac{1}{\gamma}}\right\}$ is:*

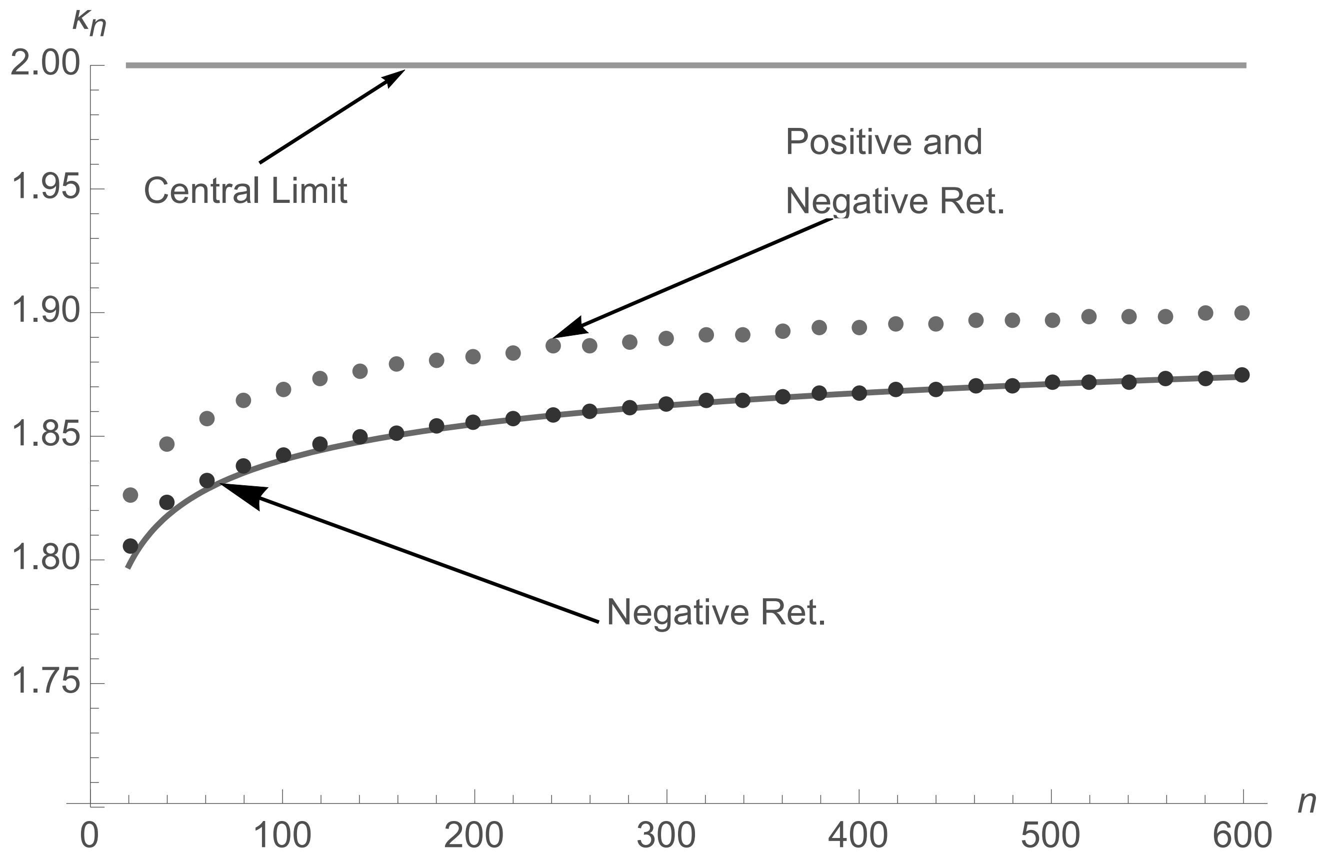
$$\gamma_n = \frac{\log(n)}{\log\left(e^n n E_{-n}(n) - 1\right)} \tag{10}$$

where $E_{(.)}(.)$ is the exponential integral $E_n z = \int_1^\infty \frac{e^{t(-z)}}{t^n} dt$.

Further, the PDF of Y can be written as

$$p(y) = \frac{e^{n-\frac{iy}{\sqrt{3}}}\left(e^{\frac{2iy}{\sqrt{3}}} E_{-n}\left(n+\frac{iy}{\sqrt{3}}\right) + E_{-n}\left(n-\frac{iy}{\sqrt{3}}\right)\right)}{2\sqrt{3}\pi} \tag{11}$$

Empirical Kappa SP500



Doing Statistics Under Fat Tails: The Program

Research Project started in 2015 by Nassim Nicholas Taleb and colleagues

(so far Pasquale Cirillo, Raphael Douady and other members of the Real World Risk Institute)

Background: *The technical papers below are part of a systematic approach to uncover mismeasurement of statistical metrics under fattailedness and propose corrections and alternative tools. Conventional statistics fail to cover fat tails; physicists who use power laws do not usually produce statistical estimators, leading to a large —and consequential — gap.* ***It is not just changing the color of the dress*** *(see discussion below).*
The initial aim was to establish a network of Bourbaki-style collaborators in a synchronized way working on the gap and injecting rigor in policy-making and decision-making under fat tails.

Taleb, N.N., "**The law of large numbers under fat tails"**(in progress**)**. This the central idea; it shows where statistical inference is BS and explores more rigorous estimation of the mean of the sum of fat-tailed random variables. A YouTube presentation here at MIT Big Data Luncheon.

Taleb, N.N., "**Stochastic Tail Exponent for Asymmetric Power Laws**"

Taleb, N.N., "**Preasymptotic behavior of subexponential and non-stable powerlaw sums"**

Taleb, N.N., "**The mathematical foundations of the precautionary principle"**(in progress**)**. Actually shows how the entire structure of probability in the social sciences is messed-up.

The inequality papers (apply to all measures of concentration, not just inequality)**:**
The next two papers apply the idea showing the flaw in using "averages" and "sums" as estimators of inequality

Silent Risk, a book in progress (freely available PDF, ~450 pages).

N. N. Taleb's Home Page

Precautionary Principle Page

Real World Risk Institute

The fragility heuristic paper (with IMF, non technical)

A Mathematical

Cirillo, P. and Taleb, N.N., 2016, "What are the odds of a thirld world war?", (*Significance*).

Taleb, N.N., Cirillo and P., Taleb, N.N., 2016,"Expected shortfall estimation for apparently infinite-mean models of operational risk", forthcmoing, *Quantitative Finance.*

P-Value Problem

Taleb, N.N., 2016, The meta-distribution of p-values, P-values (although with compact support) are fat-tailed, with effects on p-hacking.

Option Theory

Taleb, N.N., 2015, Unique Option Pricing Measure with neither Dynamic Hedging nor Complete Markets, *European Financial Management* . It proves using measure theory how a distribu- tion with finite first moment can produce a risk-neutral option price, and why we can dispense with both the dynamic hedging and pricing kernel arguments – hence price options with fat-tails.

Formalization of the barbell strategy using information theory
We are clueless about downside probability, particularly under fat tails. We look at constructions with severe tail constraints and compatible with gambler's ruin (a generalization of Kelly's criterion).

Geman, D., Geman, H. and Taleb, N.N., 2015. "Tail risk constraints and maximum entropy". *Entropy*, 17, pp.1-14.

Dimensionality and Model Error

Taleb, N.N., "Model error and dimensionality". In progress

Undecidability: *amply covered in Silent Risk (it is its theme), here is the formalization.*

Douady, R. and Taleb, N.N. Statistical Undecidability Under what conditions on the metadistribution of the probability measure is a statistical formally decidable.

Power laws and stochastic tail Exponents mixtures of power laws**.**

TBA