Additional comments on Cirillo and Taleb(2014)

Nassim Nicholas Taleb

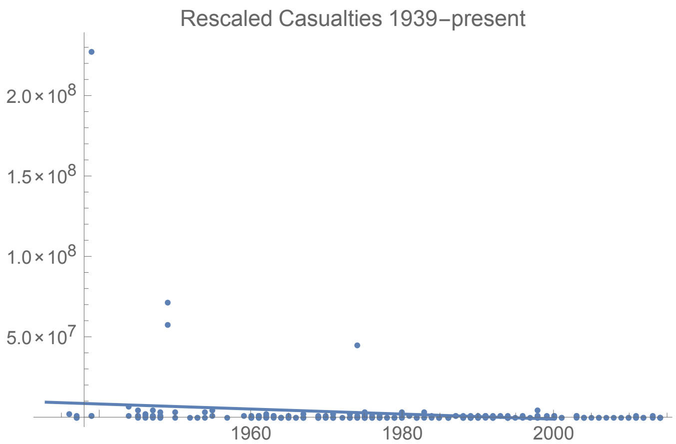

Fig. 1. Apparent drop in violence after 1945 which we test to show that it is natural optical illusion after a large deviation (particularly a cherry-picked one) and even if it weren't it would not be **statistically significant**.

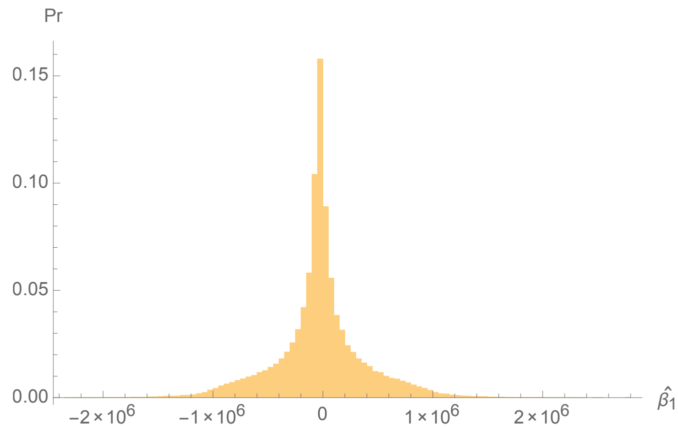

Fig. 2. Distribution of β_1 as a tail coefficient from independent random sampling from the process, 100K simulations.

***Abstract*—These are follow-up comments on our paper (Cirillo and Taleb) in response to Pol Science people and nonspecialists of extreme deviations and risk management of fat tailed processes.**

The points are too obvious to put the main paper aiming at a real probability journal, but can be useful as case study for students or those who want to debunk BS such as the "Pinker problem" about making anecdotes pass for scientific claims. It can be also useful to show students how to use Monte Carlo to get test statistics and track sampling error.

I. Did Violence Drop since 1945? The answer is "not from the data"

S. Pinker has been working the press showing a "drop" in violence since 1945 and cited some bloggers in social science talking about possible "trends"... from fattailed data without awareness of the notion of **sampling error**.

Figure 1 shows there is a "drop" since 1945. This is natural after every spike. But let us see how. So let us us a pinned process at a value y_τ, here the 1939-1944 war: You set a specific realization y_τ at period τ and look for activity in periods $t > \tau$. In this case we have $y_\tau = 2.27 \times 10^8$ (rescaled to today's populations). With the data generating process we estimated as most likely in our paper (under change of variable, Pareto with tail $\alpha \approx \frac{1}{2}$) and the most common time process $y_t = \beta_0 + \beta_1 t + \eta$ (simplified with no loss of generality), where t, to solve technical problems with Poisson arrival times, is not a calendar time but an "event" time (sort of using Clark's clock). With some abuse of notation, $\frac{\partial Y}{\partial t}$ is our regression coefficient, so what claim can one make about the negativity of β_1? Translating Pinker into a coefficient $\hat{\beta}_1 \approx -156K$ per event since the 1945 spike. What would be a statistically significant $\beta_1 < k$, to establish such claim?

Using $\|$ for order-preserving concatenation of sequences, we have a new sequence of the pinned variable followed by independent realizations drawn from the data generating process, Monte Carlo runs indexed by i:

$$\mathbf{Y}^{(i)} = (y_\tau) \| (y^{(i)}_{\tau+1}, y^{(i)}_{\tau+2}, \dots, y^{(i)}_{\tau+n})$$

so we end with pairs $(\mathbf{Y}^{(i)}, \hat{\beta}^{(i)}_1)$, where $\hat{\boldsymbol{\beta}}_1$ is obtained by least-squares from:

$$\hat{\boldsymbol{\beta}}_1 = \left\{ \beta^{(i)}_1 : \min_{\{\beta_0,\beta_1\}} \sum_t \left(\beta_0 + \beta_1 t - y^{(i)}_t \right)^2 \right\}$$

where

$$y^{(i)} \triangleq \varphi^{-1}(x^{(i)}; L, H) = (L - H) e^{\frac{L - x^{(i)}}{H}} + H$$

and where $X \sim ParetoDistribution[84.36K., 0.536, 10^3]$, and $L = 10^4, H = 7.210^9$. It is helpful that the retransformed variable has finite variance, hence the Gauss-Markov theorem holds. But the finite variance is too high to allow such claims as Pinker's, at the sample of ≈ 206 events.

Results:

m simulations	10^5
Mean	-28396.1
Standard Deviation	$404120.$
Mean Deviation	$266549.$
Min	-2.344×10^6
Max	2.78078×10^6

Conclusion: changes in β_1 are within noise, $< .31$ STD. Even without a pinned process (Mean ≈ 0), the results are not significant (.37 STD). Scientific claims require > 1.6 or even 2 STD.