On Single Point Forecasts for Fat-Tailed Variables

Nassim Nicholas Taleb*†, Yaneer Bar-Yam‡, and Pasquale Cirillo§
*Universa Investments
†Tandon School of Engineering, New York University
‡New England Complex Systems Institute
§Applied Probability Group, Delft University of Technology

Main Statements

(i)– Forecasting single variables in fat-tailed domains is in violation of both common sense and probability theory.
(ii)– Pandemics are extremely fat-tailed events, with potentially destructive tail risk. Any model ignoring this is necessarily flawed.
(iii)– Science is not about making single points predictions but about understanding properties (which can *sometimes* be tested by single point estimates and predictions).
(iv)– Sound risk management is concerned with extremes, tails and their full properties, and not with averages, the bulk of a distribution or naive estimates.
(v)– Naive fortune-cookie evidentiary methods fail to work under both risk management and fat tails, because the absence of evidence can play a large role in the properties.
(vi)– There are feedback mechanisms between forecast and reaction that cancels the invalidity of some predictions.
(vii)– Exponential dynamics automatically satisfies the mathematical condition for chaos and its unpredictability.

Commentary

Both forecasters and their critics are wrong: At the onset of the COVID-19 pandemic, many research groups and agencies produced single point "forecasts" for the pandemic—most relied upon the compartmental SIR model, sometimes supplemented with cellular automata, or with agent-based models assuming various social rules and behaviors. Apparently, the prevailing idea is that *producing a single numerical estimate* is how science is done, and how science-informed decision-making ought to be done: bean counters producing precise numbers.

Well, no. That is not how "science is done", at least in this domain, and that is not how informed decision-making should develop.

June 25, 2020; nnt1@nyu.edu

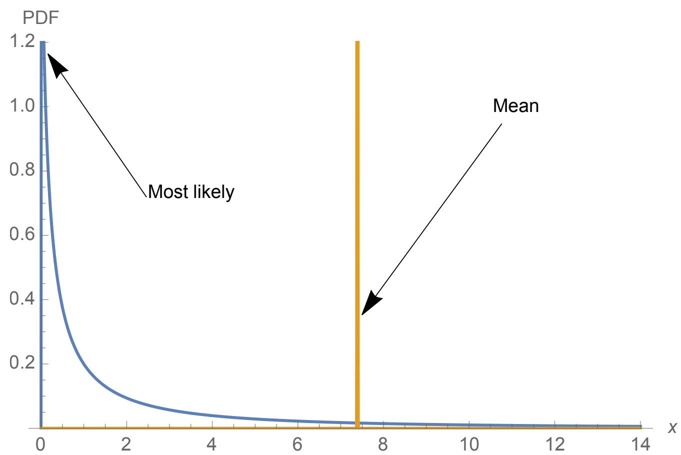

Fig. 1. A high variance Lognormal distributions. 85% of observations fall below the mean. Half the observations fall below 13% of the mean. The lognormal has milder tails than the Pareto which has been shown to represent pandemics.

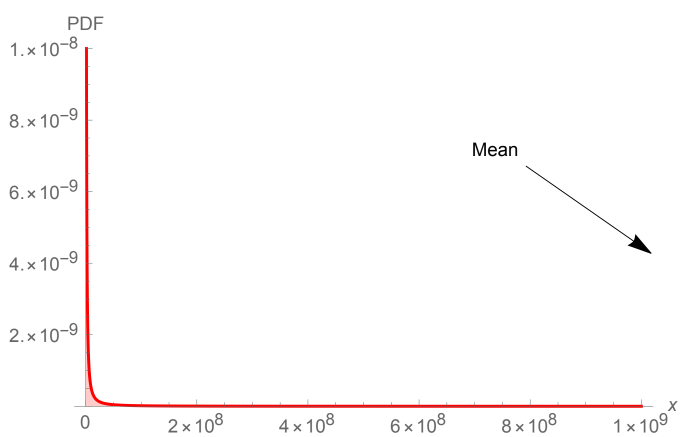

Fig. 2. A Pareto distribution with a tail similar to that of the pandemics. Makes no sense to forecast a single point. The "mean" is so far away you *almost* never observe it. You need to forecast things other than the mean. And most of the density is where there is noise.

Furthermore, subsequently and ironically, many criticized the numerous predictions made by different groups, on grounds that these did not play out (no surprise there). This is also wrong, because both forecasters (who missed) and their critics (complaining) were wrong. Indeed, as we will clarify in what follows, forecasters would have been wrong anyway, even if they had got their predictions right.

Statistical attributes of pandemics: Using tools from extreme value theory (EVT), Cirillo and Taleb [5] have recently shown that pandemic deaths are *patently* fat-tailed[1]–a fact some people such as Benoit Mandelbrot (or one of the authors, in *The Black Swan [17]*) had already guessed, but never formally investigated. Even more, the estimated tail parameter α is smaller than 1, suggesting an apparently infinite risk [5], in line with destructive events like wars [3], [4], and the so-called "dismal" theorem [22]. Pandemics do therefore represent a source of existential risk. The implication is that much of what takes place in the bulk of the distribution is just noise, according to "the tail wags the dog" effect [5], [18]. And one should never forecast, pontificate, or theorise from noise! Under fat tails, all relevant and vital information lies in fact in the tails themselves (hence in the extremes), which can show remarkably stable properties[2].

Remark 1: Distributional Evidence is Strongest Evidence

Random variables with unstable (and uninformative) sample moments may still have extremely stable and informative tail properties, centrally useful for robust inference and risk management.

Furthermore, counter to what one may be led to think from naive "evidence based" claims, these constitute evidence –distributional evidence is the strongest form of evidence.

This is the central problem with the misunderstanding of ***The Black Swan*** *[17]: some events have stable and well-known properties, yet they do not lend themselves to naive conventional point-prediction.*

Fortune-cookie evidentiary methods: In the early stages of the COVID-19 pandemic, scholars like Ioannidis [13] suggested that one should wait for "more evidence" before acting with respect to that pandemic, claiming that "we are making decisions without reliable data".

Firstly, there seems to be some probabilistic confusion, leading towards the so-called delay fallacy [12]: "if we wait we will know more about X, hence no decision about X should be made now." In front of potentially fat-tailed random variables, more evidence is not necessarily needed. Extra (usually imprecise) observations, especially when coming from the bulk of the distribution, will not guarantee extra knowledge. Extremes are rare by definition, and when they manifest themselves it is often too late to intervene. Sufficient –and solid – evidence, in particular for risk management purposes, is already available *in the tail properties themselves*. An existential risk needs to be killed in the egg, when it is still cheap to do so. Events of the last few months have shown that waiting for better data has generated substantial delays, causing thousands of deaths *and* serious economic consequences.

Secondly, unreliable data[3]–or any source of serious uncertainty–should, under some conditions, make us follow the "paranoid" route. More uncertainty in a system makes precautionary decisions more obvious. If you are uncertain about the skills of the pilot, you get off the plane. If there is an asteroid headed for earth, should we wait for it to arrive to see what the impact will be? The logical fallacy runs deeper: "We did not see this particular asteroid yet" misses the very nature of the power of science to generalize (and classify), and the power of actions to possibly change the outcome of events. Similarly, if we had a hurricane headed for Florida, a statement that "We have not seen this hurricane yet, perhaps it will not be like the other hurricanes!" misses the essential role of risk management: to take preventive actions, not to complain ex post. And if people take action boarding up windows, and evacuating, a claim that someone might afterwards make that "look it was not so devastating", such claim should be considered closer to a lunatic conspiracy fringe than scientific discourse.

By definition, evidence follows–and does not precede!–rare impactful events. Waiting for the accident before putting the seat belt on, or evidence of fire before buying insurance would make the perpetrator exit the gene pool. Ancestral wisdom has numerous versions such as *"Cineri nunc medicina datur"* (one does not give remedies to the dead), or the famous saying by Seneca *"Serum est cavendi tempus in mediis malis"* (you don't wait for peril to run its course to start defending yourself).

Remark 2: Fundamental Risk Asymmetry

For matters of survival, particularly when systemic, and in the presence of multiplicative processes (like a pandemic), we require "evidence of no harm" rather than "evidence of harm."

Technical Comments

The Law of Large Numbers (LLN) and Evidence: In order to leave the domain of ancient divination (or modern anecdote) and thus enter proper empirical science, forecasting

[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 $t > 0$ [6], [7], [8]. 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 $\alpha > p$).

[2] In [15], Ioannidis et al. erroneously maintain that choosing tail events as done by [5] is "selection bias". Actually, the standard technique there used is the exact opposite of selection bias: in EVT, one purposely focuses on extremes, to derive properties that nevertheless influence the rest of the distribution as well, especially from a risk management point of view. One could more reasonably argue that the data in [5] do not contain all the extremes, but, by jackknifing and bootstrapping the data, the authors actually show the robustness of their results to variations and holes in historical observations: the tail index α is consistently lower than 1. Finally, when the authors in [15] state that "Tens of millions of outbreaks with a couple deaths must have happened throughout time," to support their selection bias claim against [5], they seem to ignore the fact that the analysis deals with pandemics and not with a single sternutation. The class of events under considerations in [5] is precisely defined as "pandemics with fatalities in excess of $1K$," and their dataset likely contains most (if not all) of them. Worrying about many false alarms in the tail of the distribution of pandemic fatalities is thus misplaced.

[3] Ironically, many of those complaining about the quality of data and asking for more evidence before taking action, even in extremely risky situations, rarely treat the inputs of their predictive models as imprecise [3], [21], stressing them, and performing serious robustness checks of their claims.

must abide by both evidentiary and probabilistic rigor. Any forecasting activity about the mean (or a given parameter) of a phenomenon requires the working of the law of large numbers (LLN), guaranteeing the convergence of the sample mean at a known rate, when the number n of observations increases. This is surely well-known and established, except that some are not aware that, even if the theory remains the same, the actual story changes under fat tails.

Even in front of the most well-behaved and non-erratic random phenomenon, if one claimed fitness or non-fitness of a forecasting ability on the basis of a single observation ($n = 1$), she would be rightly accused of unscientific claim. Unfortunately, with fat-tailed variables that "$n = 1$" error can be made with $n = 10^6$. In the case of events like pandemics, even larger $n \to \infty$ can still be anecdotal.

Remark 3: What is not Forecastable

Fat-tailed random variables with tail exponent $\alpha \leq 1$ are simply not forecastable in the traditional sense. They do not obey the LLN, as their theoretical mean is not defined, so there is nothing the sample mean can converge to. But we can still understand several useful tail properties.
And even for random variables with $1 < \alpha \leq 2$, the LLN can be extremely slow, requiring an often unavailable number of observations to produce somehow reliable forecasts.

As a matter of fact, owing to preasymptotic properties, a conservative heuristic is to consider variables with $\alpha \leq \frac{5}{2}$ as not forecastable in practice. Their sample mean will be too unstable and will require way too much data for forecasts to be reliable in a reasonable amount of time. Notice that 10^{14} observations are needed for the sample mean of a Pareto "80/20", with $\alpha \approx 1.13$, to emulate the gains in reliability of the sample average of a 30-data-points sample from a Normal distribution [18].

Assuming significance and reliability with a low n is an insult to everything we have studied since Bernoulli, or perhaps even Cardano.

Science is about understanding properties, not forecasting single outcomes: Figures 1 and 2 show the extent of the problem of forecasting the average (and so other quantities) under fat tails. Most of the information is away from the center of the distribution. The most likely observations are far from the true mean of the phenomenon and very large samples are needed for being able of performing a reliable forecast. In the lognormal case of Figure 1, 85% of all observations fall below the mean; half the observations even fall below 13% of the mean. In the Paretian situation of Figure 2, mimicking the distribution of pandemic deaths [5], the situation gets even worse: the mean is so far away that we will almost never observe it. It is therefore preferable to look at other quantities, like for example the tail exponent.

In some situations of fast-acting LLN, as (sometimes) in physics, properties can be revealed by single predictive experiments. But it is a fallacy to assume that a single predictive experiment can actually validate any theory–rather remember that a single tail event can falsify a theory [17].

Sometimes, as recently shown in the IJF by one of the authors [19]), a forecaster may find a single quantity that is actually forecastable, say the survival function. For n observations a tail survival function has an error of $o(\frac{1}{n})$, even when tail moments are not tractable, which is why many predict binary "outcomes"–as with the "superforecasters" masquerade. In [18], it is actually shown how the more intractable the higher moments of the variable, the more tractable the survival function. Metrics such as the Brier score are well adapted to binary survival functions, though not to the corresponding random variables. That is why survival functions are essentially useless for risk management purposes [19]. One never uses survival functions for hedging, but rather expected shortfalls–binary functions are reserved to (illegal) gambling[4].

We do not observe properties of empirical distributions: On his blog, Andrew Gelman [14] wrote: "The sad truth, I'm afraid, is that Taleb is right: point forecasts are close to useless, and distributional forecasts are really hard."

The problem is actually worse. Even if one moves away from point forecasts and looks at "distributional forecasts," then he may find out that these are often hard to obtain, and possibly uninformative. Playing with empirical survival functions under fat tails in many cases does not reveal tail properties, since observations will likely be censored and missing the real tail–the object that under strong fat tails ($\alpha \leq 2$) harbors not most but literally all of the story [18]. However, as also shown in [5], the tail parameters are themselves thin-tailed distributed, so they reveal their properties rather rapidly. The correct analysis of tail parameters via EVT thus allows for a more reliable study of tails, and a more solid approach to risk, while empirical survival functions simply do not.

Uncertainty goes one way: errors in growth rates induce biases and massive fat tails.: Consider the simple model

$$X_t = X_0 e^{r(t-t_0)},$$

where X_t represents the number of fatalities between periods t_0 and t. Set then

$$r = \frac{1}{(t-t_0)} \int_{t_0}^{t} r_s ds,$$

with r_s being the instantaneous rate.

A small change in r can have a tremendous impact on X, because of the connection between the two quantities. As shown in Figure 3 an exponentially distributed r, which thus follows a well-behaved thin-tailed distribution, leads to an extremely fat-tailed X, which may turn out to be intractable. Furthermore the more volatile r, the more downward-biased your observation of the mean of X.

The implications is clear: one cannot translate between the rate of growth r and the quantity X, because changes in r can be small (but likely nonzero), but they may have an explosive impact on X, because of exponentiation. The tail exponent α

[4]For the fat-tailed random variable X and a high threshold value K, one has $\lim_{K\to\infty} \frac{1}{K}\mathbb{E}_{\mathbb{P}}(X|X > K) > 1$, where the expectation is under the real-world measure $\mathbb{P}$. We refer to [19] for all details and implications.

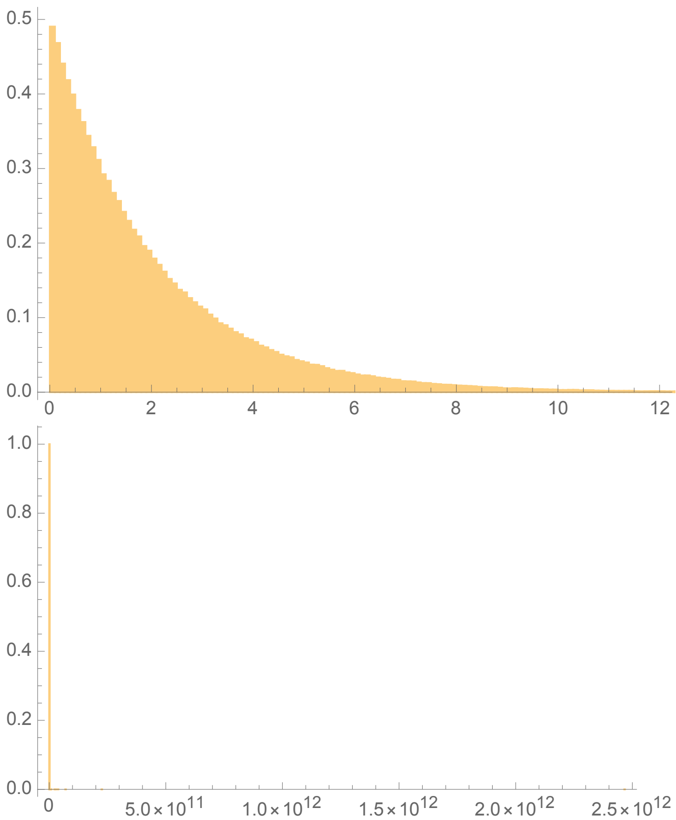

Fig. 3. Above, a histogram of 10^6 realisations of r, from an exponential distribution with intensity $\lambda = \frac{1}{2}$. Below, the histogram of $X = e^r$. We can see the difference between the two distributions. The sample kurtosis are 9 and 10^6 respectively (in fact it is theoretically infinite for the second); all values for the latter are dominated by a single large deviations.

of X is a direct function of the variance of r: the more volatile r, the fatter the tail of X.

Remark 4: Exponential Growth

Errors in the growth rates of a disease increase the fatness of the tail in the distribution of fatalities.
Errors in growth rates translate, on balance, into higher expected casualties and a magnification of unpredictability.

We note that in the context of dynamical systems an exponential dynamics is defined as chaotic [1]. While the study of chaos often considers systems with fixed parameters and variable initial conditions, the same sensitivities arise due to variations in parameters. In the case of pandemics, the value of the contagion rate (R) and the social behaviors affecting it. This means that by changing human behavior, the dynamics can be strongly affected (something that can be both good and bad news, depending on the decisions taken) [11].

Never cross a river that is 4 feet deep on average : Risk management (or policy making) should focus on tail properties and not on the body of probability distributions. For instance, The Netherlands have a policy of building and calibrating their dams and dykes not on the average height of the sea level, but on the extremes, and not only on the historical ones, but also on those one can expect by modelling the tail using EVT, mainly via semi-parametric approaches [6], [7].

Science is not about safety: Science is a procedure to update knowledge; and it can be wrong provided it produces interesting discussions that lead to more discoveries. But real life is not an experiment. If we used a p-value of .01 or other methods of statistical comfort for airplane safety, few pilots and flight attendants would still be alive. For matters that have systemic effects and/or entail survival, the asymmetry is even more pronounced.

Forecasts can result in adjustments that make forecasts less accurate: It is obvious that if forecasts lead to adjustments, and responses that affect the studied phenomenon, then one can no longer judge these forecasts on their subsequent accuracy. Yet the point does not seem to be part of the standard discourse on COVID-19.

By various mechanisms, including what is known as Goodhart's law [16], a forecast can become a target that is gamed by participants–see also the Lucas' critique applying the point more generally to dynamical systems. In that sense a forecast can be a warning of the style "if you do not act, these are the costs".

More generally any game theoretical framework has an interplay of information and expectation that causes forecasts to be self-canceling. The entire apparatus of efficient markets–and most of modern economics–is based on such a self-canceling aspect of prediction under both rational expectations and an arbitrage-free world.

References

[1] E.N. Lorenz (1963). Deterministic non-periodic flow. Journal of the Atmospheric Sciences. 20(2): 130-141.
[2] R. Albert and A.-L. Barabasi (2002). Statistical mechanics of complex networks. Reviews of Modern Physics 74: 47.
[3] P. Cirillo, N.N. Taleb (2016). On the statistical properties and tail risk of violent conflicts. Physica A: Statistical Mechanics and its Applications 452, 29-45.
[4] P. Cirillo, N.N. Taleb (2016). Expected shortfall estimation for apparently infinite-mean models of operational risk. Quantitative Finance 16, 1485-1494.
[5] P. Cirillo, N.N. Taleb (2020). Tail risk of contagious diseases. Nature Physics 16, 606-613.
[6] L. de Haan, A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer.
[7] P. Embrechts, C. Klüppelberg, T. Mikosch (2003). *Modelling Extremal Events*. Springer.
[8] M. Falk, J. Hüsler J, R. D. Reiss R-D (2004). *Laws of small numbers: extremes and rare events*, Birkhäuser.
[9] R. Lucas (1976). Econometric Policy Evaluation: A Critique. In K. Brunner, A. Meltzer (eds.). The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy 1, 19-46. American Elsevier.
[10] J. Norman, Y. Bar-Yam, N.N. Taleb (2020). Systemic Risk of Pandemic via Novel Pathogens - Coronavirus: A Note. New England Complex Systems Institute.
[11] A.F. Siegenfeld, N. N. Taleb, Y. Bar-Yam (2020), What models can and cannot tell us about COVID-19, Proceedings of the National Academy of Sciences
[12] S.O. Hansson (2004). Fallacies of Risk. Journal of Risk Research 7, 353-360.
[13] J.P.A. Ioannidis (2020). A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data. Stat March 17, https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/

[14] Andrew Gelman (2020), Some forecasting for COVID-19 has failed:a discussion of taleb and ioannidis et al, available at https://statmodeling.stat.columbia.edu/2020/06/17/some-forecasting-for-covid-19-has-failed-a-discussion-of-taleb-and-ioannidis-et-al.

[15] J.P.A. Ioannidis, S. Cripps, M.A. Tanner (2020). Forecasting for COVID-19 has failed. International Institute of Forecasters, June 11 2020, available at https://forecasters.org/blog/2020/06/14/forecasting-for-covid-19-has-failed

[16] M. Strathern (1997). Improving ratings: Audit in the British University system. European Review 5, 305-321.

[17] N.N. Taleb (2001-2018). *Incerto: Fooled by Randomness, The Black Swan, The Bed of Procrustes, Antifragile, and Skin in the Game* . Penguin

[18] N.N. Taleb (2020a). *Statistical Consequences of Fat Tails*. STEM Academic Press.

[19] N.N. Taleb (2020b). On the statistical differences between binary forecasts and real-world payoffs, International Journal of Forecasting, doi: https://doi.org/10.1016/j.ijforecast.2019.12.004.

[20] N.N. Taleb, P. Cirillo (2019). The Decline of Violent Conflict: What do the data really say? In A. Toje, N.V.S. Bård, eds. The Causes of Peace: What We Know Now. Nobel Symposium Proceedings. Norwegian Nobel Institute, 57-85.

[21] R. Viertl (1995). *Statistical Methods for Non-Precise Data*. CRC Press.

[22] M.L. Weitzman (2009). On modeling and intepreting the economics of catastrophic climate change. Review of Economics and Statistics 1, 1-19.