The Problem is Beyond Psychology: The Real World is More Random than Regression Analyses

Nassim N. Taleb [a]

Daniel G. Goldstein [b]

[a] Polytechnic Institute of New York University

[b] Yahoo Research
111 W. 40th Street
New York, NY 10028, USA

dgg@yahoo-inc.com

This version: Sept 15, 2011

Corresponding Author Current Address:

Nassim N. Taleb
Polytechnic Institute of New York University
Six MetroTech Center
Brooklyn, NY 11201, USA

Soyer and Hogarth (2011) identify unexpected and severe errors of interpretation of the parameters in linear regression on the part of people on whose expertise we rely upon, namely those involved in econometric and statistical analyses in their professional work. Their results show two major issues: first, a divorce between the analytical definitions and their practical interpretation, second the one-way effect of underestimation of the random character of the process generating the data (or, equivalently the overestimation of the deterministic effect of the parameters). We (Goldstein and Taleb, 2007) have identified both of these problems in the interpretation of the commonly used notion of standard deviation, used in finance as a proxy for "volatility", and found similar errors by persons of similar expertise. First, we observed that despite our participants being able to define standard deviation mathematically, they erred in its practical application, as if there was a severe loss in the translation of an abstract mathematical term into its practical meaning. Second, as in the case of Soyer and Hogarth's study, we observed an underestimation of the role and effect of randomness. Our participants underestimated standard deviation, while those in Soyer and Hogarth's study underestimated the practical effects of it. The most severe problem, however, may lie elsewhere: the tools themselves underestimate randomness.

In the Soyer and Hogarth (2011) case, the matter at hand is standard regression and Gaussian probabilities and participants are asked to make probabilistic interpretations with the Gaussian as the normative framework for computation of frequencies, a general assumption in economics. Econometrics is dominated by standard deviations and more generally measures in the L2 norm[1], based on squares of numbers (SD is the square root of the average of the sum of the squared deviations), all of which are grounded in a class that revolves around the Gaussian family: the Gaussian and related distributions that converge to it under a reasonable amount of summation, such as binomial, Poisson, chi-square, and exponential. The rub is that the Gaussian is of limited applicability outside of textbook examples —the

[1] A norm Lp = $\left(\frac{1}{n}\sum|x|^p\right)^{1/p}$.

type of randomness that prevails in game setups such as coin tosses or, possibly in quantum mechanics. Using it leads to the underestimation of fat tails, to the role of extreme events, and to predictions that underestimate their own errors. For instance, Taleb (2009) showed, using close to 20 million pieces of economic data (most economic variables over a period spanning the past forty years) the following: i) The data have fat tails, meaning that the errors would be dominated by larger deviations than estimated ii) The "fat tailed" nature of the data does not disappear under aggregation, meaning that the sum of the variables remains fat-tailed, and eliminates the hypothesis of convergence to Gaussian thin-tailedness, and iii) The fat-tailedness of the data is impossible to estimate—though we know the process is fat-tailed. Assume we agreed that kurtosis is a measure of the degree of fat-tailedness of the process (a scaled fourth moment of the distribution). For all variables, kurtosis depends on a very small number of observations —for instance, close to 78% of the total kurtosis of the U.S. stock market for 10,000 observations of data depends on one single observation, implying we can't figure out within the L2 norm the fat-tailedness of the process without a huge measurement error. For these reasons, this Gaussian framework fails us severely in economics.

Our point is that while we should be concerned with experts' underappreciation of the role of randomness in a regression framework, we should also be concerned with how regression's assumptions lead to an underestimation of real-world risk.

Now the questions that naturally arise would be: What if we used another, supposedly better fitting distribution? Would that lead to proper estimation of the risks in the real world? The answer is, alas, which distribution, and with which parameters? The problem with the "tails" is that they are not tractable and will be subjected to severe measurement errors. Even if we assumed, generously, that we had the right distribution, small errors in calibration of the parameters lead to disproportionately larger and larger effects in the tails (Taleb, 2011). Since these tail events determine a large share of the properties for almost all socio-economic data, we are left in the dark about the most important information. The conclusions are i) to focus on limiting exposures to these tail events, rather than invent distributions to feel comfortable with them and put people at risk, and ii) limit the use of such probabilistic statements to matters not affected by tail events.

Finally, we deplore the practice in behavioral economics and finance to impart a behavioral anomaly to a mistaken statistical analysis, one that ignore fat tails. Take for instance the "equity premium puzzle" in which equities are held to be vastly outperforming stocks according to some metric. The puzzle goes away once one starts considering that such statements cannot be made about fat tailed processes, as the tools used to derive the existence of the anomaly are themselves erroneous, as explained in Mandelbrot and Taleb (2010). There is a small probability of catastrophic loss that is not taken into account in the analyses. Further, the equity premium puzzle has vanished since the discussions about it as the past decade has witnessed the severe underperformance of stocks. Accordingly, we believe that psychological analyses of many phenomena of the sort can be severely misleading: the psychological should give precedence to the statistical.

REFERENCES

Goldstein, D. G. & Taleb, N. N. (2007). We don't quite know what we are talking about when we talk about volatility. Journal of Portfolio Management, 33(4), 84-86.

Mandelbrot, B. and N. N. Taleb (2010). Mild vs. Wild Randomness: Focusing on Risks that Matter, in Frank Diebold, Neil Doherty, and Richard Herring, eds., The Known, the Unknown and the Unknowable in Financial Institutions. Princeton, NJ: Princeton University Press.

Taleb, N. N. (2009). Errors, robustness, and the fourth quadrant. International Journal of Forecasting 25(4), 744-759.

Taleb, N. N. (2011). The future has thicker tails than the past: Model error as branching counterfactuals. Manuscript in preparation.