THE BELL–CURVE IS WRONG: SO WHAT.

PAUL EMBRECHTS
DEPARTMENT OF MATHEMATICS, ETH ZÜRICH

In a recent conversation with a banker, upon introducing myself as one of those mathematicians interested in finance and insurance, he asked me whether I had read the *Financial Times* article ([8]) pointing out that the bell–curve is wrong. After I informed him that I even co–authored a book ([2]) on this topic, he replied by saying that this must be akin to a physicist discovering a new elementary particle. I informed him that, though the comparison was flattering for the scientists working "beyond the bell–curve", I would personally scale down a bit this comparison. The point made in [8] was that in various studies in different scientific fields, it was shown that rare events are more common than the bell–shaped curve predicts. How can we understand the above discussion, and what, if any, is its relevance for finance.

Well, first of all, the bell–curve referred to above is of course the famous normal (or Gaussian) distribution. For a random variable X (think for instance of X as the P&L position of a trading book, or the return of a financial instrument), X has a normal distribution with mean m and variance σ^2, assumed finite, if the probability that X takes values between a and b is given by

$$P(a < X \leq b) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx \,. \tag{1}$$

We denote this by $X \sim N(m, \sigma^2)$; the important standardised case, $m = 0$, $\sigma = 1$, is referred to as the standard–normal case. By (1), if we know m and σ (through statistical estimation, say), we can estimate probabilities by surfaces under the bell–shaped curve; see Figure 1.

FIGURE 1. $P(a < X \leq b)$ for $X \sim N(m, \sigma^2)$

From (1), such results as "68% of the data lie between $m - \sigma$, $m + \sigma$", "95% between $m \pm 2\sigma$" and "99% between $m \pm 3\sigma$" can be numerically derived. But also "a 5σ move or more has probability 10^{-6}, a 10σ move 10^{-23}, a 20σ move 10^{-88}". On these kind of figures we have built our "normal" intuition. At this point, it may be interesting to know that modern reliability theory is pushing the quality control boundaries from 3σ to 6σ, the latter corresponding to 10^{-9} events; pretty rare.

If now, like in the FT article [8], someone has observed that certain rare events occur more frequently than predicted by the bell–curve, something must be wrong with the modelling. One such example is the 1987 crash which was a 10^{-68} event in a bell–curve world with 20% yearly volatility. A further example is the LTCM crash. As noted in [5], by August 31 1998, the LTCM portfolio had lost \$1.71 trillion in a month; this translates (in LTCM's risk management system) into a 8.3 monthly standard deviation event. And, as the author in [5] rightly concludes: "Assuming a normal distribution, such an event would occur once every 800 trillion years, or 40 000 times the age of the universe. Surely the model was wrong." This "surely the model was wrong" is a key point we need to address in finance.

Before we do this, a comment on "When is the bell–curve right?". The (mathematical) solution lies in the so–called Central Limit Theorem giving an approximation of the distribution function of averages. Again in finance language, suppose $X_1, \dots, X_n$ are the daily returns on a financal instrument with mean return m and volatility σ (and I won't discuss where we got these numbers from), and denote the average daily return by

$$\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}\,. \tag{2}$$

Then, under precise assumptions on the X_i's (they are independent, each with the same distribution, denoted by iid), we have that, for n large,

$$\overline{X}_n \sim N\left(m, \frac{\sigma^2}{n}\right)\,, \tag{3}$$

and hence the bell–curve (i.e. N) comes out of the blue. But note that N in (3) plays a measuring role for **averages** $\overline{X}_n$. If now we have reason to believe that financial returns (the X_i's themselves) come about as some sort of averaging process; for instance, very many "small" traders constantly hitting (i.e. buying, selling) the price process, then, in such a micro model, we may come up with the bell–curve as a good first approximation. Incidently, the same result underlying (3) also yields the so–called Lévy (α–)stable distributions; the σ (volatility) of the X_i's making up for $\overline{X}_n$ must in that case be infinite! With the α–stables (the normal case corresponds to $\alpha = 2$) we do believe in averaging effects.

Now at least two things can go wrong in the above construction, and both are relevant for finance. First of all, we still believe in the micro trading argument underlying the normal (Black–Scholes) world, but the underlying assumptions (iid, σ constant) are violated. Then, at first, there is no reason even for averages to always come up with a result like (3). Econometricians have looked at these problems and have established the following, so–called stylised facts for return data (only a partial list):

- returns are uncorrelated, but the absolute and squared returns are strongly correlated (this hints at an intricate dependence structure in return date),
- returns are heavy–tailed, extremes occur in clusters, leptokurtosis,
- volatility (σ) is stochastic (fluctuations, clustering).

Further "facts" like long–range dependence, chaotic behaviour, self–similarity, fractality, regime switching and so many more have been looked at. For an excellent critical view on these stylised facts, and how they can be replicated in specific models, see [7]. Criticising the bell–curve world and its friends for not fitting empirical data well is one thing, coming up with a better (the "correct") model is quite a different matter. Some institutions have gone for a parametric mixing of the normal, i.e. assume that the σ in $N(m, \sigma^2)$ is random with a gamma distribution, say; the dynamics built around this model are still of the so–called Lévy–process type, i.e. a process with stationary, independent increments. Others went for the, again parametric, class of GARCH–type processes in which the volatility dynamics are modelled directly. More general stochastic volatility models may use semi–parametric, or even non–parametric models. In this model hunting, we stick for a while by our favorite one and whenever a new stylised fact comes around the corner, we start hunting for the next (rarer) animal. Several of the newer models come from physics. One of the main differences between economics and physics however (from a modelling point of view) is that the former has no agreed set of fundamental laws, also most economic processes are irreversible, and empirical model testing is nearly impossible. I personally am a bit hesitant in pushing this model hunting too far. As so often in the social sciences, there is no universally correct model. As long as we understand the assumptions underlying a particular model choice we can also stress test the model in order to go beyond its often narrow boundaries. Model risk is therefore an issue that will stay with us forever.

The stylised facts above tell us that the micro–model underlying financial markets is much more intricate than the assumptions underlying the validity of (3). Looking for better models here is definitely useful. However, in many cases applications of bell–curve thinking are applied to situations where there is absolutely no averaging going on! For instance, if the X_i's stand for losses in a credit portfolio, from a risk management point of view, I am not only interested in the average loss, but much more importantly in the so–called stress loss. The latter corresponds to the $p100\%$ largest losses with p very small. What is the loss level for my portfolio which will only be surpassed with probability p small (VaR thinking)? Questions like these are not governed by the average loss $\overline{X}_n$ but by the extreme loss $M_n = \max(X_1, \ldots, X_n)$, or an average of the 1% largest losses, say. In order to handle the behaviour of these extremes, a theory similar to (3), but now for M_n has to be worked out. In Figure 2 we give the limiting laws which occur (there are three types) instead of the normal law: they are the Gumbel, Fréchet and Weibull. Do note the asymmetry (skewness) and indeed heavy–tailedness. For instance, when the X_i's are themselves normally distributed, then their maximum M_n is governed by the two–sided skew Gumbel distribution (bold solid curve) in Figure 2. The resulting Extreme Value Theory (see [2]) is just one tool which can be used for traveling beyond the bell–curve. For a very readable introduction to EVT, see [6]. One place where EVT can play an important role is quantile (VaR) estimation given a historical or simulated P&L or credit loss distribution. Many banks go to great length getting a proper P&L histogram say, but then finish the job by reporting a VaR empirically often as the smallest (or close to the smallest) P&L observation. EVT allows for a parametric fitting in

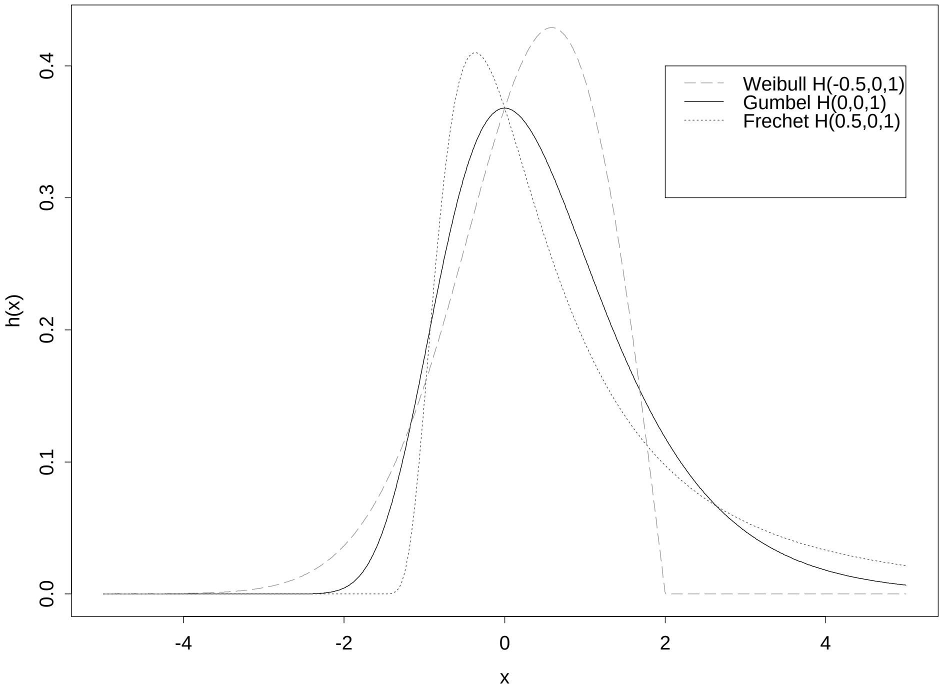

FIGURE 2. Extreme value distributions

the tail area of the P&L; there are of course assumptions underlying such fitting, but practice has shown (using backtesting) that such procedures work well and are definitely better than some gormless guessing. For the estimation of extreme quantiles (like VaR, stress losses in credit, scenario testing) non–parametric methods in general and bootstrapping in particular have to be treated with great care. Quantile estimation for heavy–tailed distributions is the key example where the standard bootstrap methodology does not work.

Most of the above points are certainly (hopefully) not new for most finance experts: the normal model is a good first approximation, various changes/alternatives are being studied/used. Reality in finance is to an important degree (especially within Risk Management) non–normal. And yet in practice we go on using implicitly normal assumptions. Examples are:

- the use of VaR as a risk measure,

- the use of linear correlation for measuring dependence,
- mean–variance portfolio optimisation,
- market risk capital measures based on averaging (think of the Basle 60 day averaging rule),
- using the Sharpe–ratio for ranking different long–tailed investment opportunities,
- constructing symmetric confidence intervals around risk measures.

In all of these, and many more examples, one uses normal thinking. This can lead to wrong or misleading conclusions, especially in risk management. Going back to the LTCM case discussed earlier, an easy model improvement from normality to a heavier–tailed t–model shows that the loss incurred was an eight years event (instead of an 800 trillion years one in the wrong model); see [5]. A final example to show how the non–normal world differs. Many authors have rightly criticised VaR for not summarising the tail–loss properly; an alternative, so–called coherent risk measure in use (see [1]) is the expected loss given that a loss above VaR occurs, also referred to as shortfall. In a normal world, for VaRs calculated for high confidence limits, this conditional expected loss is roughly VaR itself. In a non–normal, heavy–tailed world, the difference between the two risk measures can easily be a factor 2 or 3.

Conclusions. Of course the bell–curve is right, when applied to the right sort of problem. It is however very "wrong" in many (if not all) of the financial and insurance Integrated Risk Management models. This definitely in credit and operational risk management where the loss distributions are much more akin to the very skew insurance–loss type distributions. This is not just an academic statement. By now, many cases are documented where bell–curve based calculations are just wrong, and in some cases, terribly wrong. The methods needed to further improve existing IRM systems have now been around for a while; it is up to the individual risk manager to find out how they can be put to work in practice. For some further reading on this issue, see [3], [4] and the references therein.

References

[1] Artzner, P., Delbaen, F., Eber, J.–M. and Heath, D. (1999) Coherent measures of risk. *Mathematical Finance* **9**, 203–228.

[2] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

[3] Embrechts, P., McNeil, A. and Straumann, D. (1999) Correlation: Pitfalls and Alternatives. *RISK* (5), 69–71.

[4] Embrechts, P., McNeil, A. and Straumann, D. (1999) Correlation and dependency in risk management: properties and pitfalls. Preprint ETH, Zürich, www.math.ethz.ch/∼embrechts.

[5] Jorion, P. (1999) How Long–Term lost its capital. *RISK* (9), 31–36.

[6] McNeil, A. (1999) Extreme value theory for risk managers, *RISK*, to appear, www.math.ethz.ch/∼mcneil.

[7] Mikosch, T. and Stărică, C. (1999) Change of structure in financial time series, long range dependence and the GARCH model. Preprint. University of Groningen and Chalmers University of Technology, www.cs.rug.nl/∼eke/iwi/preprints.

[8] Peel, M. Redrawn curve reveals new pattern of events. *Financial Times*, Thursday September 2, 1999, p. 8