FORESIGHT

The International Journal of Applied Forecasting

ARTICLES *by* FORECASTERS *for* FORECASTERS: Q3:2023

Opinion Editorial: The Technological Limits to Forecasting

The Technological Limits to Forecasting

MALTE TICHY

The efficiency of power plants (how much heat is converted to electricity), the resolution of microscopes (how small an object can one discern), and the fuel consumption of cars (how far they go on a gallon of fuel) are limited by the laws of nature. These fundamental bounds to everyday technologies are insuperable. They are well established, and even to some extent intuitively understood by nonscientists.

Forecasts are not that different from power plants, microscopes, and cars: they are also produced by a technology, the difference being that forecasts are not constrained by the laws of physics, but by those of information theory and statistics. Nevertheless, natural laws impose natural limits. But what are these limits? How good can a forecast possibly become? Spectacular advances have nurtured the impression that accuracies can be pushed to ever higher values, that all imperfection constitutes merely an engineering challenge.

The variety of today's forecasting methodologies (statistical, machine-learning, time-series, etc.) obfuscates the fact that all forecasts are limited in the accuracy they can yield. The behavior of the process that is being modeled, and the available information that is exploited ultimately define what accuracy is possible.

Under a given set of information, we can characterize the best forecast as the *sharpest* one that is still *calibrated*. These concepts arise from the 2007 article "Probabilistic Forecasts, Calibration and Sharpness" by Gneiting, Balabdaoui, and Raftery, appearing in the *Journal of the Royal Statistical Society*. *Sharpness* refers to the strength of the statement that a forecast makes: "With 90% probability, we'll sell between 12 and 15 boxes of strawberries" is a sharper, stronger, more valuable statement than "With 90% probability, we'll sell between 2 and 25 boxes of strawberries." The sharper forecast is preferable – but only under the condition that the forecast remains *calibrated* (meaning that we can rely on finding 90% of the realizations within the predicted 90% intervals). The best forecast is thus the sharpest, least uncertain one that is reliably calibrated.

Assuming forecasts to be well calibrated, how sharp can they become? Since humans generally don't like uncertainty, they strive for the sharpest thinkable forecast, a deterministic one, which suggests that an event will occur with certainty: "With 100% probability, we'll sell 14 boxes of strawberries." For any nontrivial situation, such zero-uncertainty prediction is never calibrated, and its occasional realization amounts to pure luck. Point forecasts mimic certainty, giving a strong sense of security, but a false and dangerous one. When you stock those 14 boxes of strawberries, you should expect either leftovers (waste) or unfulfilled customer demand. You can hide the uncertainty, but that doesn't make it go away.

Accepting and quantifying the uncertain nature of every forecast facilitates more suitable decisions than what a point forecast allows. Depending on the cost of strawberries, the number of missed sales that are acceptable, and other business considerations, the optimal level of stock will vary.

There is the kind of uncertainty that we can control and reduce by more accurate information and better algorithms, and there is irreducible uncertainty that we will have to live with. The latter sets an upper bound to the achievable level of forecast sharpness. This distinction is not always easy, but I argue that it is possible in many domains and should be established whenever one wants to produce a forecast.

Understanding the upper limits of sharpness – the "speed of light" of forecasting – would yield immense benefits. Benchmarking a given forecast against the ideal helps identify the low-hanging fruit and prioritize improvement efforts. This would discourage futile attempts to improve already excellent forecasts and encourage attention toward clearly suboptimal forecasts. Just like I can compute how much fuel I'll need to get from New York to Boston before the ride, I should articulate what sharpness I can ideally expect from a forecast even before constructing it.

As an example, we can estimate the upper, often not even achievable, bounds to forecast sharpness in retail. Consider demand for a single item in a supermarket that serves many customers every day. The customers are not individually known, so the best that the forecaster can achieve is to predict the total number of customers on a given day (say, 1,000) and the probability of a customer buying the item on that day (say, 2%). This gives the prediction 1,000 * 0.02 = 20. Charitably assuming that the total number of visitors and the average buying probability are known perfectly, the resulting item sales will follow a binomial distribution centered around 20. If there were 2,000 customers buying the item with probability 1% instead, the resulting distribution would not change much. In this regime (many customers, small average probability), what eventually counts is the expected number of sales: 20. The limit of the binomials, the Poisson distribution with mean 20, is to nonpersonalized retail forecasting the sharpest possible forecast.

Could we improve a Poisson-limited forecast? That would require us to model a different process and to predict each individual customer's buying probability: "Alice never buys orange juice, but Bob does in 20% of his visits." Simple numerical experiments show that the level of certainty required to achieve a sharper-than-Poisson forecast is considerable. The forecaster would need to know the buying probabilities extremely well to yield a substantially better-than-Poisson forecast.

The retail example takes advantage of a simple process, gentle distributions, and the convenient limit of large numbers. But upper limits to forecast sharpness can also be established in other domains. A forecast implicitly models the parameters of a process. By inspecting that process, we can demarcate the boundary between "possibly achievable" and "out of question." By thought experiments and numerical studies, we can elaborate whether a certain degree of certainty is realistic. Thereby, we can answer this question: What would be, assuming optimal knowledge, the remaining inexorable uncertainty?

Let's become uncertainty-aware, conscious of fundamental limits, and explicit about what we can possibly know and what we cannot. Let's mitigate avoidable ignorance to achieve sharp forecasts, while embracing inevitable uncertainty to use them strategically. Forecast users rightfully expect statements on achievable forecast capabilities and reliable, reproducible performance. After all, forecasters are certainly as capable as power plant, microscope, or car manufacturers.

Malte Tichy has a research background in quantum physics, with a PhD from University of Freiburg. He is a Fellow and Senior Staff Data Scientist at Blue Yonder, working at the interface between customer project implementation, research, and product development to ensure that probabilistic forecasts can unleash their full potential.

malte.tichy@blueyonder.com