Professor KOSTAS NIKOLOPOULOS, DIMITRIOS D. THOMAKOS and KONSTANTIA LITSIOU provided lessons from forecasting competitions…

Summer 2019

P44-45

Page title

Bangor University

Draft header

Tricks of the forecasting trade

Standfirst

Professor Kostas Nikolopoulos, Dimitrios D. Thomakos and Konstantia Litsiou provide lessons from forecasting competitions.

Body copy

Time series dominate the information bankers use when they need a meaningful estimate of the future. This is complemented often by narrative or even richer information. So, for example, a statement and probably a press conference will accompany a balance sheet of a major corporation. Nevertheless, a longitudinal sequence of meaningful numerical observations (for example number of loans per period, a stock index, an asset price, etc), and their respective projections in the future, drive our decisions in most decision-making contexts.

Whatever works

To that end, one question comes to mind: what time series forecasting method should we use? Nowadays the forecasting arsenal is so rich, ranging from exponential smoothing approaches (dating back to the 1950s) to AI, machine- and deep-learning methods. Nevertheless, despite theoretical advances in analytics, statistics and econometrics, still as of today no method fits all! No method can forecast consistently better in all contexts: there are horses for courses, and even these unfortunately change over time.

Therefore, to play safe we really need to continuously perform empirical forecasting competitions. We must compete different forecasting models over a holdout of our data (usually the last 20% of our data), each time new data become available, in order to pick a 'winner' to be used to prepare forecasts and respective prediction intervals. If in doubt, we can always combine, so if we feel the 'winner' is a first-time-ever or just not our cup of (forecasting) tea – some bankers absolute hate 'black-box' methods – then we can always use the combination of our top three or top five winners. That way, we can create an ensemble that might be less accurate but more robust for out-of-sample extrapolation. And the cycle goes on: new data, (empirical) forecasting competition, new forecasts etc. So, as they say in UK, the proof of the pudding is in the eating. You are never sure what is the best method for your data; you have to empirically test many methods and rely on the one that would do the trick for your data. In other words, do whatever works!

We do, however, need a starting point. If we don't have enough data to empirically decide on the best method to use, then we need an a priori selection protocol. Previously published empirical forecasting competition results can be very helpful; these give you an idea of what methods are expected to perform well for different contexts. Even more importantly, published forecasting completion results highlight benchmarks to beat. It is very often the case, in practice, for forecasters to strongly argue for a new 'wonder method' they have developed in-house, but fail to recall to test it against simile and computationally cheap time series forecasting benchmarks. Methods such as the random walk, exponential smoothing and multiple regression are so hard to beat in the long run!

The Makridakis (M) Forecasting Competitions

Book excluded, there are more than 18,000 citations in Professor Spyros Makridakis' work to date; the most are around M1, a competition among 1,001 time series, and M3 – done in 2000 – and testing models over 3,003 time series. But the impact of the former at its time was huge. You can arguably claim that the whole discipline for forecasting as we know it is an offspring of M1 and that very publication. The main lessons learned from these competitions are:

- *Statistically sophisticated methods do not necessarily provide more accurate forecasts than simpler ones*

- *The relative performance of methods varies according to the accuracy measure being used*
- *The accuracy when methods are combined outperforms, on average, the individual methods being combined*
- *The accuracy of methods depends upon the length of the forecasting.*

In the latest iteration of these studies – The M4 competition – 100K+ time series had to be forecasted and for the first time we had a wide variety of innovations. These included industry participations, many well-performing combinations, machine learning methods, testing of prediction intervals, transparency and replicability of black-box methods. Uber forecasting team won the competitions via a hybrid machine learning and statistical method. It was state-of-the-art technically and also intuitively appealing as it used and exploited properties of the entire dataset every time it forecast an individual time series. Forecast pro was the top-performing commercial software, while the Theta Method was the top statistical benchmark.

The ones to beat

We have seen so many forecasting reports that advocate for wonder-forecasting-methods but which can only outperform Naïve, a potential moving average or just ARIMA. This is fundamentally and methodologically wrong and we should eliminate it as bad practice. It has been obvious for the last two decades that there exists a series of accurate methods – methods that do run very fast and are free to use. These include computationally cheap benchmarks available in R or Python, such as Hyndman's forecast package, for example. The Theta method, ARIMA, ETS, Damped Exponential Smoothing and many combinations in between them perform well consistently and as such should always be used as benchmarks in practice.

For any method you want to try in-house to be considered rationally employed, it should be on par, or better than, these fast and cheap benchmarks. It is a hard fact of forecasting!