The M4 competition, and a look to the future

Fotios Petropoulos

Mike Gilliland has written a very insightful overview of the latest Makridakis competition (M4), the outcome of which is of great value to practitioners and academics alike. I widely agree with the points that he raised and would like both to reinforce some arguments and highlight several additional elements of the competition.

PREDICTION INTERVALS

Mike correctly notes the welcome evaluation of prediction intervals in the competition. Estimating the uncertainty around the forecasts is invaluable and can inform the preparation of contingency plans. However, I would personally like to see future competitions go one step further and directly link the uncertainty in the forecasts to the setting of safety stocks and proper inventory/utility evaluation of the forecasts. Such an exercise would allow us to gain an empirical understanding of the cost of the forecast error: how 1% in accuracy translates to monetary savings.

TRADEOFF OF ACCURACY AND COST

I also agree that we need to trade-off additional accuracy with the added computational cost of achieving it. A translation of forecasting performance into monetary terms – as noted above – would allow for a direct comparison of the cost of forecast error with the computational cost of the procedures utilized, given the plethora of available-to-rent could computing service such as Amazon Web Services or Microsoft's Azure Cloud Computing.

WINNING FORMULAS

It is worth noting that the top ranked methods share several elements. As Mike correctly notes there are benefits from forecast combinations. Additionally, five of the top six entries in the M4 competition (including my own submission) use at least one variant of the Theta method, which was the winning submission at the M3 competition. (For an introduction to the Theta method, see Petropoulos and Nikolopoulos, 2017). So participants cautiously chose methods that have repeatedly performed well. One notable exception is the winning submission (https://eng.uber.com/m4-forecasting-competition), whose underlying basis is Holt-Winters, a method that did not perform well on its own in previous competitions (at least for longer forecasting horizons).

Another important element of the M4 competition is the use of global models. Instead of focusing on a series-by-series modelling approach, the two top-ranked submissions applied cross-learning, which allows a series to get information from other series, either in terms of selecting the parameters of the model or selecting combination weights. This technique has been widely used in machine learning solutions in several Kaggle competitions, while Bojer and Meldgaard (2020) predict that it will be an important element to be tested in the forthcoming M5 competition.

APPLICABILITY TO FORECASTING PRACTICE

I have heard many comments questioning the applicability of the M4 results to forecasting practice. While I agree that the absence of intermittent data and hierarchical structures is an important limitation—but one that will be addressed by the M5 competition-- the data in the M4 are nevertheless quite representative of business forecasting efforts especially in comparison to datasets from industry-specific Kaggle competitions (see Bojer and Meldgaard, 2020, for more details). Moreover, if practitioners are interested in a single-data frequency (such as monthly or

daily), it is possible to decompose the results of the M4 for that particular frequency and take forward insights from the top performing methods in that cluster.

PERFORMANCE MEASURES

One element about which I'm not entirely in agreement with Mike Gilliland's article is that different error measures result in different rankings. The Spearman's rank correlation coefficient for the overall values of sMAPE and MASE (the two measures used to evaluate forecast accuracy in the M4 competition) is very strong, at 88.5%. Methods that performed well based on the sMAPE are also very likely to perform well with regards to MASE. I understand the limitations when translating these results into practice, with different companies focusing on different KPIs, but I believe that appropriately specifying the cost function to fit the purpose will do the trick. This, for instance, was done by the second-best performing method in the M4 competition who adopted the M4 competition's main cost function (OWA) to estimate the methods' combination weights.

THE VALUE OF PARTICIPATION

I would like to emphasize the importance of participating in forecasting competitions. When a task is challenging, as was the case with the M4 competition where participants had to submit forecasts for 100K series, I view participation as a marathon rather than a sprint. While the rankings are obviously important, participation itself is an achievement that should be celebrated and I hope that more forecasting researchers and practitioners embrace entering future competitions. At the end of the day, we cannot call ourselves forecasters if we are afraid to forecast. We must have skin in the game.

CONCLUSION

Despite some limitations and criticisms, my personal view of the M4 is that we gained valuable insights which advance the theory and practice of forecasting. I am confident that many forecasting researchers and practitioners will benefit from the availability of open access code for most of the submissions in the M4 and use this information to improve upon their own approaches.

Last but not least, I believe that the M4 competition managed to bring together two communities: traditional statistical/econometrics experts with experts from the machine learning world. As demonstrated by the winning entry, exciting things can happen when these two worlds work together.

References

Bojer, C.S. and Meldgaard, J.P. (2020), Learnings from Kaggle's Forecasting Competitions, ResearchGate, 10.13140/RG.2.2.21579.75046

Petropoulos, F. and Nikolopoulos, K. (2017), The Theta Method, Foresight, Issue:46 (Summer 2017) 11-17.

Fotios Petropoulos is Senior Lecturer (Associate Professor) at the School of Management, University of Bath, the Forecasting Support Systems Editor of Foresight and Associate Editor of the International Journal of Forecasting. His research expertise lies in behavioural aspects of forecasting and improving the forecasting process, applied in the context of business and supply chain.