International Journal of Forecasting 36 (2020) 224–227

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The M4 competition: Conclusions

1. Introduction

It is exactly 40 years since the publication of the predecessor to the M Competitions, the study by Makridakis and Hibon (1979, MH), with its surprising and hotly contested finding that simple statistical methods were at least as accurate as complex and statistically sophisticated ones. More specifically, the study found that, on average, the mean absolute percentage error (MAPE) of Naïve 2 was 1.4% more accurate than that of the Box-Jenkins methodology for ARIMA models, while single exponential smoothing (SES) was 2.61% more accurate. Clearly, this was far from welcome evidence in regard to the most popular approach to statistical forecasting of that time, and was not received kindly by the statistical academic community (see Hyndman's paper, 2020, "A brief history of forecasting competitions" for details). Nevertheless, the MH study started a long forecasting winter that continued with the M1 (Makridakis et al., 1982) and M2 (Makridakis et al., 1993) Competitions, which provided similar findings. Such findings were partially lifted with the M3 Competition (Makridakis & Hibon, 2000), in which the sMAPE of Theta, still a simple method, was 2.46% more accurate than that of Naïve 2, 1.31% more accurate than that of SES and 1% more accurate than that of the Box-Jenkins method.

The forecasting spring began with the M4 Competition, where a complex hybrid approach combining statistical and ML elements came first, providing a 9.4% improvement in its sMAPE (symmetric MAPE) relative to that of the Comb benchmark, while the top 16 methods achieved sMAPEs that were 4.49% more accurate than that of the Comb benchmark on average. As Gilliland (2020) pointed out in his discussion paper, such results show that the time has come to discard the belief that complex methods are no better than simple ones, and sound the death knell for ARIMA models, which are still in use today despite the empirical evidence that has been piling up for the last 40 years. Spring brings considerable opportunities that can be exploited to advance the theory of forecasting and attract more practitioners who are in search of ways to improve the accuracy of their predictions and assess the future uncertainty more realistically. It is the purpose of this concluding paper to describe the factors that have brought this forecasting spring, briefly reiterate the major achievements of the M Competitions overall and the M4 in particular, and discuss how such factors can be exploited to avoid another winter.

2. Factors that brought the forecasting spring

The factors that, in combination, brought the forecasting spring are discussed below in order of their perceived importance.

A new breed of forecasters: As the field of forecasting has advanced, it has attracted a new breed of forecasters who study in a number of academic institutions that specialize in this field, while at the same time, full-time forecasting experts are being hired by business firms that are interested in improving their predictions and obtaining a better understanding of the uncertainty involved. The major difference between the old bunch of forecasters like Spyros and the new ones like Fotios is that Spyros initially studied statistics, among other things, and specialized in forecasting afterwards, but has also worked on several other projects concurrently. Fotios, on the other hand, has been involved exclusively in forecasting, with his PhD thesis also being in forecasting. Pablo, of the second winning method, has a career that is similar to that of Fotios. He completed his dissertation on forecasting at Monash University where his advisors were statisticians who, like Spyros, specialized in forecasting later in their careers. Pablo's publications are focused entirely on forecasting, as is his thesis. The same is true with Evangellos, the co-organizer of the M4 Competition and co-author of the M4 and several of the other papers included in this special issue. His PhD was on forecasting and he started his research in this field immediately. Practically all of his numerous publications are in the field of forecasting. Furthermore, Fotios, Pablo and Evangelos grew up with computers and are sophisticated programmers, whereas when Spyros started, punched cards were still used to communicate with computers.

https://doi.org/10.1016/j.ijforecast.2019.05.006

On the practitioners' side, Slawek of the first winning method holds an MSc in Physics and has been working full time for Uber doing time series forecasting exclusively. Marciej of the third winning method also holds an MSc in computer science and is interested in neural networks and natural language processing. Slawek and Marciej, like Fotios, Pablo and Evangelos, are computer experts, having written their own advanced programs for their winning methods. Slawek, Marciej and the other practitioners who have contributed papers to this special issue are also working full time in forecasting and hold advanced degrees, including PhDs and Masters. Thus, the difference between the old and new groups of forecasters is fundamental, which explains their significant contribution that we consider to be the single most important factor in the huge accuracy/uncertainty improvements achieved in the M4 Competition. No doubt they and other similar contributors have been responsible for bringing the forecasting spring. The forecasting field has gone from being a part-time career for a group of people who are interested in the field, to a full-time occupation for academics and practitioners alike. Clearly, the last decade has witnessed the emergence of the professional, full-time forecaster in both the academic and business fields.

Great advances in ML methods: Luckily, the statistical field of forecasting discovered a close cousin in the ML area that was also interested in pattern recognition, and therefore forecasting. Such a discovery offers significant prospects for further advances in the field and for the creation of an expanded, unified forecasting space. As Januschowski et al. (2020) stated, these two fields *"can learn and benefit from each other's strengths"* as long as they can accept the necessity *"to step outside their comfort zone"* and work together. In our view, this unification will need to happen in the near future, as the complementarity of the two schools of thought is understood better and their advantages and drawbacks become clearer. There is no reason why this artificial distinction should continue to exist in the future, whereas breaking it will allow for a wider choice for satisfying various forecasting needs, without any concern as to whether they are statistical or ML, as long as they can satisfy specific forecasting applications best, achieving the highest performance at the lowest cost. Slawek's wining method is the best example of the advantages to be gained by marrying the two approaches.

Exponential improvements in computer speed and memory: It would have been practically impossible to run the M4 Competition with its 100,000 series even ten years ago, or for some of the participating methods, which require enormous amounts of CPU time, to have been part of such a competition. There is no doubt that major improvements in computer speed have made the M4 Competition possible, especially for those methods that utilize complex machine learning (ML), which require demanding computations. Equally, faster computers have allowed the participants to experiment with various options in order to select the best method(s) to use in the M4 Competition (one participant told us of a huge electricity bill from running his five home computers almost constantly for 4.5 months in order to choose the best method(s)). As improvements in computing speed continue, two things will happen. First, forecasters will be able to experiment with a wider range of options for improving the accuracy and uncertainty; and, second, the use of ML methods for forecasting purposes will increase as computing costs decrease and greater experience in using them is gained.

3. The contributions of the M Competitions and the achievements of the M4

Hyndman's paper (2020) provides an excellent discussion of forecasting competitions, making it possible for us to be brief in this section. The major advantages of the M Competitions are their openness and their objectivity. Their main contribution to the field, and what distinguishes them from others, is their strong emphasis on *learning*, and on using such learning to improve the theory and practice of forecasting. The detailed results of the M Competitions have been published in respectable journals, and moreover they have generated considerable interest in the academic community, which has utilized the datasets for experimenting with new methods. Practitioners have also been interested in these results in their efforts to improve their firms' forecasting performances and to estimate the uncertainty more realistically. Finally, the M Competitions have provided a historical record extending 40 years back that can be used in various ways to experiment in and advance the field.

3.1. Key achievements of the M4 Competition

In our view and that of the organizers, the four key achievements of the M4 are as follows:

- It proved that complex methods could provide accuracies that were well above those of simple ones. More specifically, sixteen methods achieved higher accuracies than the Comb benchmark, while the top six exceeded such a benchmark by more than 5%. The improvement of 19.3% of Smyl's winning method (Smyl, 2020) over Naïve 2 is impressive, unlike the corresponding decrease of 1.4% of the Box-Jenkins found in the MH study.
- It confirmed that combining more than one method improved the forecasting accuracy considerably, while also establishing another form of even more accurate, hybrid combining of statistical *and* ML approaches.
- It demonstrated that the two top methods provided phenomenally precise estimates of the uncertainty (as far as we know, this is the first time that methods have done so) that can be utilized by other methods for estimating the uncertainty realistically.
- Almost half of the methods submitted, including the top nine, have been replicated fully by the organizers of the competition, and the code of each is available for free on GitHub, along with all of the data, for anyone to use in their company or for conducting academic research.

4. Be realistic as to what forecasting can and cannot do, and avoid "overselling" its achievements

The forecasting market is huge and, in our opinion, still virgin. Most business people do not believe or trust forecasting, with some equating it with fortune telling and others with economic predictions that often go wrong. What the M4 Competition has shown is that the forecasting accuracy can be improved considerably over the naïve approaches that are used widely in businesses. In addition, there is a broad choice of alternative methods that can be adapted to all business needs, given some specific forecasting budget and some amount of effort to be devoted to forecasting. This means that an educational effort is required not only to explain the advantages of systematic forecasting, but also to point out that all forecasts are inaccurate, with the only certainty being the extent of such inaccuracy, which should be reported along with point estimates. It will therefore be necessary to assess the potential benefits of the improved accuracy and translate them into dollars and cents to persuade prospective users of the benefits of systematic forecasting for their scheduling, planning and strategic tasks. In addition, the value of estimating the uncertainty realistically must be made clear, as well as its benefits in reducing inventory costs and improving customer satisfaction, among other gains.

Gilliland (2020) in his discussion paper talks about *"the shocking disappointment of real-life business forecasting"* that falls well short of its theoretical potential, and suggests that research is needed to determine the causes that stop firms from using more accurate methods and what can be done to persuade business people of the benefits of using more appropriate forecasting techniques. He suggests the use of a forecast value added (FVA) analysis to identify bad practices in order to avoid them and determine the potential savings from using better methods. Now that big tech firms, including Amazon, Google, Microsoft and SAP, among others, are offering ML forecasting services, this will probably open the business forecasting market, thus increasing the attractiveness of forecasting as a field and creating substantial opportunities for research to further improve its value, leading to more accurate predictions. Moreover, business firms like Uber use forecasting heavily in their day-to-day operations, which highlights its benefits and provides an example for other firms to follow.

5. Closing remarks

This M4 Competition special issue has brought together high caliber academics and top-level practitioners to comment on, discuss and criticize the M4 Competition, as well as the challenges facing the field and the best way forward. Their suggestions have been invaluable and have established a direct line of communication between the academic and business communities that we hope will continue to grow and strengthen. Similarly, we expect that the integration of the statistical and ML groups will be successful, providing a common effort to further advance the field. As a part of the wider data science field, forecasting is bound to play a critical role in future in identifying patterns in data and consequently forecasting as accurately as possible, while also providing realistic estimates of the uncertainty. We should add that this special issue contains detailed descriptions of the winning methods by their authors that can serve as starting points for using them, conducting additional academic research, and improving the practice of forecasting in business firms, while making sure that practitioners fully understand both its benefits and its limitations, as well as the fact that all future predictions are uncertain.

Once this special issue is finalized, planning for the M5 Competition will accelerate. The major difference between the M5 and the M4 will be the inclusion of explanatory and exogenous variables, in order to determine whether doing so will improve the accuracy of pure time series predictions. As in the M4 Competition, several difficult choices will have to be made, and when making these choices, we will take into account the feedback and numerous suggestions received from the forecasting community in response to the M4. We are hopeful that progress will continue and that future competitions will further improve the forecasting accuracy/uncertainty and help to persuade more business users of the advantages and practical benefits of utilizing systemic predictions as an integral part of scheduling and planning, as well as for formulating strategies. We believe that continuous progress and a realistic view of what forecasting can and cannot do will avoid another forecasting winter.

References

Gilliland, M. (2020). The value added by machine learning approaches in forecasting. *International Journal of Forecasting*, *36*(1), 161–166.

Hyndman, R. (2020). A brief history of forecasting competitions. *International Journal of Forecasting*, *36*(1), 7–14.

Januschowski, T., Gasthaus, J., Flunkert, V., Wang, B., Bohlke-Schneider, M., Salinas, D., et al. (2020). Criteria for classifying forecasting methods. *International Journal of Forecasting*, *36*(1), 167–177.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*(2), 111–153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5–22.

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society. Series A*, *142*(2), 97–145.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.

Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, *36*(1), 75–85.

Spyros Makridakis was until recently the Rector of the Neapolis University of Pafos and an Emeritus Professor at INSEAD. He is now taking on the role as Director of the Institute for The Future (IFF) at the University of Nicosia in Cyprus. He has held teaching and research positions with several institutions: as a research fellow with IIM Berlin, an ICAME fellow at Stanford and a visiting scholar at MIT and Harvard. Mr. Makridakis has authored, or co-authored, 24 books including Forecasting, Planning and Strategy for the 21st Century (The Free Press), Forecasting: Methods and Applications, 3rd ed. and Forecasting

Methods For Management, 5th ed. He has also published more than 120 articles and book chapters and was the founding chief editor of the Journal of Forecasting and the International Journal of Forecasting. He has been the Chairman of the Board of Lamda Development and the Vice Chairman and board member of more than a dozen companies.

Fotios Petropoulos is Associate Professor at the School of Management of the University of Bath, Associate Editor of the International Journal of Forecasting and the Forecasting Support Systems Editor of Foresight. His research expertise lies in behavioural aspects of forecasting and improving the forecasting process, applied in the context of business and supply chain. He is the co-founder of the Forecasting Society (www.forsoc.net).

Spyros Makridakis
Institute for the Future (IFF), University of Nicosia, Cyprus

Fotios Petropoulos *
School of Management, University of Bath, UK
E-mail address: f.petropoulos@bath.ac.uk.

* Corresponding editor.