

Northumbria Research Link

Citation: Goodwin, Paul, Gonul, Sinan and Önköl, Dilek (2022) Commentary on Making Forecasting More Trustworthy. *Foresight: The International Journal of Applied Forecasting* (66). pp. 26-29. ISSN 1555-9068

Published by: International Institute of Forecasters

URL: <https://ideas.repec.org/a/for/ijafaa/y2022i66p26-29.html>
<<https://ideas.repec.org/a/for/ijafaa/y2022i66p26-29.html>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/49512/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Commentary on Making Forecasting More Trustworthy

Paul Goodwin, M.Sinan Gönül and Dilek Önkal

Forty years ago, two analysts at British Gas, a major utility supplier in the UK, wrote a paper describing their experiences of implementing a complex forecasting method in their organisation (Taylor and Thomas, 1982). They had designed their method to forecast the daily demand for natural gas by consumers based on factors such as the day of the week, the previous day's demand, and the following day's temperature forecast. On most days, it proved to be highly accurate. Yet the operational personnel -the intended users of the model - were unconvinced that it was performing well. The occasional large error loomed large in their memories while they tended to forget the model's routine accuracy.

This case study demonstrates the importance of three of the attributes that Simon Spavound and Nikolaos Kourentzes identify in their insightful article as crucial to determining whether a forecasting method is trusted or not. First, the model was not intelligible to its intended users. Second, its reliability was not apparent to them. Third, the model did not align with their objectives: large errors were expensive, so avoiding these was more important to them than achieving modest reductions in smaller errors.

The British Gas model did at least meet the fourth criterion of stability. But the effects of instability on trust can be seen elsewhere -for example, in reports that US President Jimmy Carter repeatedly complained about the inconsistency of forecasts by his economic advisors. He hinted that he'd be better off using a fortune-teller at the Georgia State Fair (Nordhaus, 1987).

Forty years on, the evidence provided by Spavound and Kourentzes suggests that trust in forecasting, has not improved and, indeed, might even have declined. Algorithm aversion

and skepticism has become a major issue in recent years (Dietvorst and colleagues 2015, 2018) -managers in some companies override over ninety per cent of their statistical forecasts (e.g., Fildes and colleagues, 2009). In the media and elsewhere, macro-economic and political forecasts are regularly attacked and even lampooned (Goodwin, 2017, p.4).

Worse still, trust is often misplaced. In one experiment, participants were even prepared to pay for predictions of whether a coin toss would result in heads or tails (Powdthavee and Riyanto, 2015). So-called experts can often convince people, without evidence, that they have special powers of foresight (Armstrong, 1980, Önkal and colleagues, 2017). In some cases, the salience of a single lucky, highly accurate prediction is sufficient to confer credibility on a person's forecasts, despite a general record of inaccuracy (Goodwin, 2017, p.149).

So, what can be done to foster trust where it is merited? How can intelligibility, alignment, reliability, and stability be achieved and demonstrated?

Intelligibility is becoming more challenging as increased computer power permits the application of more complex and more opaque forecasting methods. As a minimum, forecasters need to make their assumptions transparent and declare the information they are using. But additionally, it is often possible to create a non-technical account of how these are turned into forecasts. Providing a narrative alongside the forecasts can be helpful as well. Gönül and colleagues (2006) have found that explanations accompanying forecasts can lead to lower adjustments and a higher acceptance of those predictions. As Spavound and Kourentzes point out, an intuitive understanding of how things work is sufficient for us to trust many technical devices, so why not forecasts?

Alignment means that, in some cases, complexity is not justified anyway. Decisions linked to forecasts don't always require high accuracy, so a simpler and more understandable

approach will suffice. A key element of alignment is the perception of goodwill on the part of the forecast provider – a feeling that they share and understand the user’s objectives (Gönül and colleagues, 2012). This implies the need for a close collaboration between forecasters and their clients. Providers who act to protect their interests, such as in herding or politically influenced forecasts, can compromise trust. Forecasts should be regarded as honest expressions of what is expected to happen in the future -nothing more, nothing less. While they need to be aligned with decisions, they should also be regarded as distinct from decisions. For example, a forecast of the demand for a product is different from a decision on how much stock to hold to achieve a given customer service level. Conflating the two, which sometimes happens (Fildes et al., 2009), can lead to confusion and an erosion of trust.

Even when the distinction is clear, there’s often a lack of understanding of the transformation of forecasts into organizational decisions. A set of generated predictions might be highly accurate, but if they don’t translate to good decisions, and less-than-desirable outcomes are obtained, then the quality of those forecasts is among the first to be blamed. Meticulous attempts should be made in organizations to establish and disseminate a clear connection between forecasts the decisions that depend on them.

Reliability is often regarded as being synonymous with accuracy measurement. But, as Spavound and Kourentzes indicate, accuracy metrics can hide the occurrence of rare but impactful, large errors or ignore the need to avoid bias. Many decision analysts argue that we should judge the quality of a decision by the process that produced it, not its outcome. A good outcome does not necessarily imply a good decision and vice versa because decision outcomes can be subject to luck. The same perspective can be embraced when assessing the reliability of a forecast. Were appropriate and cost-effective data employed? Was there an underlying rationale for the method? Were the needs of the decision-maker addressed? Were the underlying assumptions plausible, and did they survive challenges? Was the process free

of political interference? The disclosure of this process to forecast users is more likely to engender trust in the reliability of forecasts than an abstruse accuracy metric, which might be out-of-date anyway if underlying conditions have changed.

Trust based on reliability also requires an acceptance of uncertainty on the part of users. Some factors are inherently unpredictable, so forecast errors are inevitable, but some senior managers, in particular, are known to be intolerant of forecasts that miss the mark. While forecasts, such as prediction intervals, do indicate uncertainty, managers may discount them if they are too wide, even when they accurately reflect the true level of uncertainty. It has even been suggested that forecasters should artificially narrow intervals to increase the likelihood that managers will accept that at least some uncertainty is present -though, in the long run, such dishonesty is unlikely to engender trust! Education appears to be the way forward here.

Achieving stability can require a careful balancing act from forecasters. A German study found that economic forecasters were unwilling to alter their forecasts even when new information suggested they needed to be changed (Kirchgässner and Müller, 2006). They feared that people would see such changes as a sign of incompetence and their reputation would suffer. Other studies have suggested that people are too eager to make changes (Van den Broeke and colleagues, 2019). Again, transparency of the underlying process is likely to be the answer. The reasons for any changes can be subject to scrutiny, and where they appear to be justified, they are more likely to be trusted.

All these dimensions point to how behavioural factors lie at the very core of achieving trust in forecasting. As also noted by Spavound and Kourentzes, while educational focus has been on developing the algorithmic and analytical knowledge, training to enhance behavioural insights appear to have been largely neglected. At precisely the time when

human-AI interactions are taking center-stage in forecasting and decision-making, we need to develop the behavioural forecasting toolbox to achieve trustworthy predictions that translate to winning decisions (Önkal and colleagues, 2019).

In conclusion, Simon Spavound and Nikolaos Kourentzes's article is timely. There is a danger that recent improvements in forecasting methods, particularly those based on algorithms, will not be exploited because they are misunderstood and distrusted by users. By identifying the key attributes underlying trust, their discussion provides a set of valuable indications of how we might avoid this danger.

References

- Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, June/July, 16-24.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155-1170.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, 25, 3-23.
- Gönül, M. S., Önkal, D., & Goodwin, P. (2012). Why should I trust your forecasts? *Foresight: The International Journal of Applied Forecasting*, 27, 5-9.
- Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, 42(3), 1481-1493.
- Goodwin, P. (2017). *Forewarned: A Sceptic's Guide to Prediction*, Lindon: Biteback Publications.

- Kirchgässner, G., & Müller, U. K. (2006). Are forecasters reluctant to revise their predictions? Some German evidence. *Journal of Forecasting*, 25, 401–413.
- Nordhaus, W. D. (1987). Forecasting efficiency: concepts and applications, *The Review of Economics and Statistics*, November, 667–674.
- Önkal, D., Gönül, M. S., Goodwin, P., Thomson, M., & Öz, E. (2017). Evaluating expert advice in forecasting: Users' reactions to presumed vs. experienced credibility. *International Journal of Forecasting*, 33, 280-297.
- Önkal, D., Gönül, M.S. & DeBaets, S. (2019). Trusting forecasts. *Futures & Foresight Science*, 1, 3-4, 10 p., e19.
- Powdthavee, N., & Riyanto, Y. E. (2015). Would you pay for transparently useless advice? A test of boundaries of beliefs in the folly of predictions, *Review of Economics and Statistics*, 97, pp. 257–272.
- Taylor, P. F., & Thomas, M. E. (1982). Short term forecasting: horses for courses. *Journal of the Operational Research Society*, 33, 685-694.
- Van den Broeke, M., De Baets, S., Vereecke, A., Baecke, P., & Vanderheyden, K. (2019). Judgmental forecast adjustments over different time horizons. *Omega*, 87, 34-45.