Introduction to the Special Issue on Judgment in Forecasting

Editors of the Special Issue:

Robert Fildes, Lancaster University, r.fildes@lancaster.ac.uk

Paul Goodwin, University of Bath, p.goodwin@bath.ac.uk

Fergus Bolger, Anglia Ruskin University, fb17@aru.ac.uk

Nigel Harvey, University College London, n.harvey@ucl.ac.uk

Matthias Seifert, IE Business School, matthias.seifert@ie.edu

One of the critical challenges in forecasting relates to the question of how to effectively use human judgment to optimize predictive performance, judgment permeating all aspects of the forecasting process whatever the application area (e.g., Bolger & Harvey, 1998). While past research on the role of judgment in forecasting has frequently focused either on studying the predictive performance of *unaided judgmental predictions* or on the role of *judgmental adjustments* of (linear) model forecasts, there is an emerging necessity to understand the relevance of human intervention more broadly when designing and implementing forecasting systems.

Specifically, in recent years, we have witnessed technological milestone achievements in the domains of artificial intelligence, machine learning, neural networks, cloud computing and more. These have delivered forecasters increasingly sophisticated predictive tools and methodologies to record, store, process and synthesize increasing volumes of data. Organizations are only now beginning to grasp their full potential, which promise to significantly reduce operational costs, for instance, those related to inventory management and delivery optimization (Beasley, 2021).

Yet, harnessing the value of these scientific and technological advancements requires organizations to challenge some of the traditional assumptions regarding the role of human judgment in the forecasting process. In fact, given the increasing complexity of predictive tools and methodologies, there is a need to employ skilled workers to both design models and manage their output, particularly when forecasts result from the collective effort of multiple human actors at various stages of the forecasting process.

From a scholarly perspective, in the past few years, research has increasingly highlighted the variety of circumstances in which human judgment can be either beneficial or detrimental to the performance of modern forecasting systems. For example, while earlier studies have often focused on benchmarking the performance of judgmental forecasts against the outputs of relatively simple, static, linear models, more recent research has investigated issues related to the role of human intervention in AI-driven prediction systems (Revilla et al., 2023), the use of machine learning principles to aggregate probability judgments (Grushka-Cockayne et al., 2017), algorithmic approaches to ferreting out crowd wisdom (Collins et al., 2023), and the role of algorithm aversion in human-machine collaborations (Burton et al., 2023).

Importantly, it appears that the role of human judgment in modern forecasting systems continues to represent a double-edged sword. In some contexts, human forecasters have been shown to be highly proficient in their diagnostic ability to detect abnormal deviations from historical data patterns, as they can rely on information about the forecasting task that is not reflected in the configuration of the model (Lawrence et al., 2006). This can prove to be extremely valuable in situations where disruptive changes in the task environment make constant parameters in models become obsolete. In other situations, however, judgments are likely to suffer from various cognitive biases, for instance, when human forecasters inappropriately anchor their judgment on available cues and make insufficient adjustments from them (Goodwin, Moritz & Siemsen, 2018) in the new circumstances. Moreover, they may systematically under- and overreact to information signals depending on the volatility of the forecasting environment (Kremer, Moritz & Siemsen, 2011), perceive patterns and relationships where none exist (Gaissmaier & Schooler, 2008; Sroginis, Fildes & Kourentzes, 2023) and exhibit the tendency to dampen observed long-term trends in their judgmental forecasts (Harvey & Reimers, 2013). Finally, judgments may also suffer from behavioral biases when generating predictions about probability distributions, in that human forecasters tend to underestimate uncertainty by providing overly narrow prediction intervals (O'Connor & Lawrence, 1989).

Hence, the effectiveness of forecasting systems critically depends on an organization's ability to recognize *when* and *how* human judgment can add value as well as to understand in what manner judgments and model outcomes can be combined to improve forecasts. Addressing this challenge from a scholarly perspective lies at the heart of this special issue.

Specifically, our objective is to provide an arena for showcasing key scientific advancements on the use of judgment in modern forecasting systems, and to evaluate and compare the various judgmental forecasting methods. In the past, a number of methods have been proposed in the literature related to the elicitation and aggregation of judgmental forecasts. Based on the source of the data, they can be classified as direct or indirect methods. Direct methods use an approach of collecting information from a sample of the target population for which a forecast is required (e.g. consumer intention/ expectation/ probability surveys, role playing, scenario-based forecasting). Indirect methods collect information from individuals who have knowledge about the target population or the variable for which a forecast is required (e.g. unaided judgment, prediction markets, Delphi, structured analysis, judgmental bootstrapping, expert systems, judgmental adjustment of forecasts from statistical models, identification and use of 'superforecasters' in geopolitical forecasting). Both direct and indirect methods have been successful in forecasting future events.

In addition, a second important aspect of this special issue is to present the latest research on how judgment and statistical forecasts interact, and the organizational processes involved, as well as the role of forecasting support systems in enhancing the quality of judgments. A related issue is the use of judgment in the development, calibration and selection of complex forecasting models, which includes decisions regarding the data, variables and logical procedures to be used in the forecasting process and this may be of particular relevance where more complex methods are used (e.g. machine learning/AI-based models).

A final objective of this special issue is to collect the latest insights into the cognitive processes that underlie judgmental forecasting (e.g., the use of heuristics), and the resulting biases that potentially undermine standard models of economic rationality.

Similar to earlier special issues (SIs) on judgmental forecasting in the *International Journal of Forecasting* published in 2007 (23:3), 2013 (29:2) and 2017 (33:1), the present SI intends to bring together empirical evaluations and comparisons of traditional and new methods in a single volume, providing readers with knowledge of these methods, and the opportunity to assess the various advancements made in this area of forecasting in interesting applications.

During the process of identifying the most suitable papers for this special issue, 21 articles have been accepted for publication. Among these, the types of predictions studied by the authors included both judgments about probability distributions as well as future values of variables themselves. Furthermore, the accepted papers draw on a wide range of methodological approaches including laboratory experiments, field studies, meta-analyses of past research, case studies, analyses of consumer survey data, as well as analytical modelling. In terms of forecast applications, the articles in this special issue encompass predictions in the context of COVID-19, election polls, US growth rates and other macroeconomic indices, demand forecasting and planning, general knowledge questions posed in forecast tournaments, predictions regarding existential threats to humanity as well as judgments in the context of artificially generated time series data. The accepted articles also differ in terms of the role that judgments played. In particular, while some studies rely on judgmental inputs to estimate bootstrapping models in environments where no outcome exists, other studies elicit judgments for the purpose of adjusting model outputs, for providing probability distributions, for generating single point estimates and confidence intervals about time series data, for assessing information signals, for applying weighting schemes in election polls to ensure representativeness of underlying population, for expressing preference for human vs machine-generated forecasts, or as the basis for team discussions in collective forecasting.

The topics covered in this special issue are extremely diverse and range from descriptive studies outlining ways in which judgmental forecasts may be biased to prescriptive studies using algorithmic approaches to optimizing (collective) judgmental forecast performance. For instance, Chacon (2024) offers a review and categorization of 162 published papers to examine the potential of algorithms to mitigate judgmental biases. Along similar lines, van der Staak et al. (2024) rely on a case study approach to propose a novel method for improving the effectiveness of judgmental adjustments to model forecasts. Koo, Lee & Seifert (2023) investigate the extent to which human judgment exhibits trend damping when the forecasting task involves cyclic time series data. Similarly, Engler, Hutzler and Hawelka (2024) explore the accuracy and confidence of forecasting judgments regarding time series that exhibit exponential growth. In his study of judgments regarding US growth rates, Pedersen (2024) examines the persistence of judgmental adjustments over time. In terms of individual differences, on a more positive note, Motahhar (2024) investigate how formal training can be used to mitigate biases in probability calibration.

The topic of *forecasting skill* also represents the focal point in Karger et. al's (2024) study on probability judgments in the context of existential threats to humanity, where he finds that specialists were more pessimistic than generalists in their forecasts, particularly when it came to generating estimates regarding the long-run risk of AI. Moreover, Comerford (2024) studies how cognitive reflection, arithmetic ability and financial literacy are likely to influence perceptions of inflation expectations. In their study of cue-based versus time-series forecasting, Harvey and De Baets (2024) reconcile previously conflicting findings regarding forecasters' preference for human versus algorithmic predictions. Fahimnia, Tan and Tahirov (2024) demonstrate how forecasters are likely to overestimate sales as a result of their failure

to distinguish between sales forecasts and demand plans. Finally, Motahhar, Gruca and Tavakoli (2024) rely on a textual analysis of traders' justifications for buying contracts to analyze the role of emotions in political prediction markets.

Among those papers that take a more prescriptive approach to studying judgment in forecasting, Atanasov et al. (2024) examine algorithmic crowd prediction systems for eliciting and aggregating judgmental forecasts, finding that small, elite crowds outperform larger ones. In a similar vein, Ho, Budescu & Himmelstein (2024) propose a measure of coherence in probability judgments to identify superior forecasters in crowd prediction problems. Moreover, Peker and Wilkening (2024) develop a two-step algorithm to counteract the conservatism bias in judgmental probability forecasts. Abolghasemi (2024) investigates the forecasting accuracy of human experts versus large language models (LLMs) in the retail sector during standard and promotional sales periods. The findings indicate that LLMs do not consistently outperform humans in forecasting accuracy and that advanced statistical forecasting models do not uniformly enhance the performance of either human forecasters or LLMs. Adopting the lens of a Forecast Value Added analysis, Fildes, Goodwin and De Baets (2024) rely on a large scale data set from diverse companies to analyze when forecasters are likely to make judgmental adjustments to model forecasts and propose a debiasing method accordingly. The results show the limitations of judgmental adjustment but also the potential for irradicating systematic errors. And Fritsch, Haupt and Schnurbus (2024) uses data from German federal state elections to study the efficiency of forecasting systems that rely on election polls as a measure of voter sentiment. Using a pre-registered experiment, Hardy et al. (2024) propose and test a novel method called "model-assisted judgmental bootstrapping" for making predictions in domains where outcome data are unavailable. Comerford & Soll (2024) questioned over 2000 panel members in the USA to obtain estimates of economic and political trends. The study suggested these estimates should be made indirectly by eliciting judgments of levels at different points in time rather than obtaining direct estimates which may be biased by political allegiances and a tendency to answer a simpler question than the one posed. Lastly, Khosrowabadi, Hoberg and Lee (2024) use behavioral nudges to guide supervisors in AI-enabled forecasting contexts.

We are confident that the selected papers are all based on rigorously conducted research and that they offer many important contributions to the field of judgmental forecasting. Processing the paper submissions to this special issue would not have been possible without the hard work of the many expert referees, who provided constructive feedback to the authors and significantly contributed to shaping the final versions of the accepted manuscripts. Considering this, we would like to thank the following scholars for their efforts to review and improve the submitted manuscripts:

Hal Arkes, Pavel Atanasov, Rob Basten, Bekki Brau, Muhammed Bulutay, Jason Burton, Alok Choudhary, Michael Clements, David Comerford, Miguel Cruz-Ramirez, Shari de Baets, James Derbyshire, Cuneyt Eroglu, Florian Federspiel, Spencer Fox, Andreas Fügener, Luigi Gifuni, Sinan Gönül, Kesten Green, Ross Gruetzemacher, Sreyaa Guhaa, Mohsen Hamoudia, Emily Ho, Kai Hoberg, Ali Kabiri, Matteo Kalchschmidt, Ezra Karger, Konstantinos Katsikopoulos, Edgar Kausel, John Kenny, Stephan Kolassa, Shijith Kumar PM, Yun Shin Lee, Andrew Meyer, Gilberto Montibeller, Brent Moritz, Prana Narayanan, Konstantinos Nikolopoulos, Xiaoxiao Niu, Dilek Önkal, Asa Palley, Michael Pedersen, H Niles Perara, Pierre Pinson, Stian Reimers, Thomas Rietz, Fabio Rumler, Nada Sanders, Ville Satopää, Martin Schonger, Jack Soll, Matthew Spaniol, Anna Sroginis, Norman Swanson, Susan Thorp, Juan-Ramon Trapero, Maximilian Voigt, Michael Weber, George Wright.

In addition, we editors actively reviewed the manuscripts and consolidated their reports.

In sum, this special issue has sought to provide a comprehensive overview of the latest research addressing the role of human judgment in forecasting, whether through modern forecasting systems or unaided. However, some important topics slipped through the submission and refereeing process. In particular, the SI is lacking further discussions on, for example, the intersection of scenarios with forecasting (Goodwin, Gönül and Önkal, 2019) the role of judgment in model specification (Petropoulos, Kourentzes, Nikolopoulos & Siemsen, 2018), the refinement of forecasts by interacting forecasters (Bolger & Wright, 2017), or the organizational design and support systems through which forecasts are delivered (Oliva & Watson, 2009): readers will have their own views on omissions. Nevertheless, we hope that the ideas and research methods included in this special issue are thought-provoking and stimulating and that they will generate further ideas for carrying out future studies on this important topic.

References

Abolghasemi, M., Ganbold, O., & Rotaru, K. (2023). Humans vs Large Language Models: Judgmental forecasting in an era of advanced AI*. International Journal of Forecasting*, Special issue on Judgment in Forecasting.

Atanasov, P., Witkowski, J., Mellers, B., & Tetlock, P. (2024). Crowd prediction systems: Markets, polls, and elite forecasters*. International Journal of Forecasting*, Special issue on Judgment in Forecasting.

Beasley, K. (2021). Unlocking the power of predictive analytics with AI. Forbes Technology Council. Accessed at https://www.forbes.com/sites/forbestechcouncil/2021/08/11/unlocking-the-power-of-predictive-analytics-with-ai/?sh=17fc929b6b2a.

Bolger, F., & Harvey, N. (1998). Heuristics and biases in judgmental forecasting. In P. Goodwin and G. Wright (Eds.), *Forecasting with Judgment*. Wiley.

Bolger, F. & Wright, G. (2017). Use of expert knowledge to anticipate the future: Issues, analysis and directions. *International Journal of Forecasting*, 33(1), 230–243. DOI 10.1016/j.ijforecast.2016.11.001

Burton, J. W., Stein, M. K., & Jensen, T. B. (2023). Beyond algorithm aversion in human-machine decision-making. In *Judgment in Predictive Analytics* (pp. 3-26). Cham: Springer International Publishing.

Chacon, A. (2024). An overview of the effects of algorithm use in biases affecting forecasting, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Collins, R. N., Mandel, D. R., & Budescu, D. V. (2023). Performance-weighted aggregation: Ferreting out wisdom within the crowd. In *Judgment in predictive analytics* (pp. 185-214). Cham: Springer International Publishing.

Comerford, D. (2024). Cognitive reflection, arithmetic ability and financial literacy independently predict inflation expectations: Implications for forecasting, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Comerford, D. & Soll, J. (2024). Partisan bias, attribute substitution and the benefits of an indirect format for eliciting forecasts and judgments of trend, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Engler, B.H., Hutzler, F. & Hawelka, S. (2024). An extended logarithmic visualization improves forecasting exponentially growing numbers, but residual difficulties remain, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Fahimnia, B., Tan T. & Tahirov, T. (2024). Service level anchoring in sales forecasting: the moderating impact of retail promotions and product perishability, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Fildes, R., Goodwin, P., & De Baets, S. (2024). Forecast Value Added in demand planning, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Fritsch, M., Haupt, H. & Schnurbus, J. (2024). The complementary effects of model assistance and judgmental bootstrapping for out-of-population prediction, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Gaissmaier, W., & Schooler, L. J. (2008). The smart potential behind probability matching. *Cognition*, 109(3), 416-422.

Goodwin, P., Moritz, B., & Siemsen, E. (2018). Forecast decisions. In: Donohue K, Katok E and Leider S (eds) *The Handbook of Behavioral Operations*. Wiley, 433-458.

Grushka-Cockayne, Y., Jose, V. R. R., & Lichtendahl Jr, K. C. (2017). Ensembles of overfit and overconfident forecasts. *Management Science*, 63(4), 1110-1130.

Hardy, M.D. , Zhang, S., Hullman, J., Hofman, J.M., & Goldstein. D.G. (2024). The complementary effects of model assistance and judgmental bootstrapping for out-of-population prediction, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Harvey, N. & De Baets, S. (2024). Factors affecting preferences between judgmental and algorithmic forecasts: Feedback, guidance and labeling effects, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Harvey, N., & Reimers, S. (2013). Trend damping: Under-adjustment, experimental artifact, or adaptation to features of the natural environment?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 589.

Ho, E., Budescu, D.V, & Himmelstein, M. (2024). A measure of probabilistic coherence to identify superior forecasters, *International Journal of Forecasting, Special Issue on Judgment in Forecasting*.

Karger, E., Rosenberg, J., Jacobs, Z., Hickman, M. & Tetlock, P.E. (2024). Subjective probability forecasts of existential risk: initial results from a hybrid persuasion-forecasting tournament, *International Journal of Forecasting, Special Issue on Judgment in Forecasting*.

Motahhar, V.K. (2024). How does training improve individual forecasts? Modeling differences in compensatory and non-compensatory biases in geopolitical Forecasts, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Motahhar, V. K., Gruca, T. S., & Tavakoli, M. H. (2024). Emotions and the status quo: The anti-incumbency bias in political prediction markets. *International Journal of Forecasting*, Special issue on Judgment in Forecasting.

Khosrowabadi, N., Hoberg, K. & Lee, Y.S. (2024). Guiding supervisors in AI-enabled forecasting: Understanding the Impact of salience and detail on Decision-Making, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Koo, M. S., Lee, Y. S., & Seifert, M. (2023). Investigating laypeople's short-and long-term forecasts of COVID-19 infection cycles. *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, 57(10), 1827-1843.

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of forecasting*, 22(3), 493-518.

O'Connor, M., & Lawrence, M. (1989). An examination of the accuracy of judgemental confidence intervals in time series forecasting. *Journal of Forecasting*, 8(2), 141-155.

Oliva, R., & Watson, N. (2009). Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and Operations Management*, *18*(2), 138-151. https://doi.org/10.3401/poms.1080.01003

Pedersen, M. (2024). Judgment in macroeconomic output growth predictions: Efficiency, accuracy and persistence. *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Peker, C., & Wilkening, T. (2024). Robust recalibration of aggregate probability forecasts using meta-beliefs, *International Journal of Forecasting, Special issue on Judgment in Forecasting*.

Petropoulos, F., Kourentzes, N., Nikolopoulos, K., & Siemsen, E. (2018). Judgmental selection of forecasting models. *Journal of Operations Management*, *60*, 34-46. https://doi.org/https://doi.org/10.1016/j.jom.2018.05.005

Revilla, E., Saenz, M. J., Seifert, M., & Ma, Y. (2023). Human–Artificial Intelligence Collaboration in Prediction: A Field Experiment in the Retail Industry. *Journal of Management Information Systems*, 40(4), 1071-1098.

Seifert, M. (Ed.). (2023). *Judgment in Predictive Analytics*. Springer International Publishing.

Sroginis, A., Fildes, R., & Kourentzes, N. (2023). Use of contextual and model-based information in adjusting promotional forecasts. *European Journal of Operational Research*, 307(3), 1177-1191. https://doi.org/https://doi.org/10.1016/j.ejor.2022.10.005

van der Staak, B. B. J. P. J., Basten, R. J. I., van de Calseyde, P. P. F. M., Demerouti, E., & de Kok, A. G. (2024). Light-touch forecasting: A novel method to combine human judgment with statistical algorithms. *International Journal of Forecasting, Special issue on Judgment in Forecasting*.