FORESIGHT

Fall 2020 Issue 59

THE INTERNATIONAL JOURNAL OF APPLIED FORECASTING

The Benefits of Systematic Forecasting for Organizations: The UFO Project

SPYROS MAKRIDAKIS, ELLEN BONNELL, SIMON CLARKE, ROBERT FILDES, MIKE GILLILAND, JIM HOOVER, AND LEN TASHMAN

INTRODUCTION

The purpose of this paper is to provide a realistic assessment of the potential benefits to business organizations that derive from applying systematic forecasting methods, particularly with respect to operational and tactical forecasting problems. Our overall goal is to improve the usage of forecasting in organizations—UFO—while incentivizing the adoption of systematic forecasting in organizations that now employ only ad hoc methods.

We define *systematic forecasting* as the use of appropriate quantitative methods when suitable data are available, while allowing for judgmental inputs and adjustments that are supported by a documented and defensible rationale. Where little or no data are available, such as with new products, our definition encompasses structured management judgment including use of intention surveys, decision aids, Delphi procedures, and others.

The genesis of the UFO project lies in a series of discussions within a group of practitioners and academics about the challenges facing the forecasting field and the need to learn why many organizations do not exploit what have grown to be remarkable advances in forecasting knowledge and technology.

The article seeks to present the advantages as well as the limitations of systematic forecasting methods. We do so to set fair, reasonable expectations of what can and cannot be achieved, considering the uncertainty associated with all predictions. Realistic expectations are key to establishing good forecasting practice.

We also explore the obstacles encountered by companies in the implementation and improvement of their forecasting processes and provide our understanding of how to overcome resistance to process improvement. And for organizations at "ground zero," we offer guideposts on how to get started utilizing systematic forecasting procedures.

We begin with an assessment of the accomplishments achieved in quantitative forecasting methods. As we note below, the many firms that still lack systematic forecasting need to realize that these approaches, whether simple or complex, have enormous potential benefits for their bottom lines and competitive positions.

The many firms that still lack systematic forecasting need to realize that these approaches, whether simple or complex, have enormous potential benefits for their bottom lines and competitive positions.

THE FORECASTING FIELD TODAY

It has been more than 60 years since Robert Brown's pioneering book *Statistical Forecasting for Inventory Control* (1959), which essentially founded the field of business forecasting. Brown's exponential-smoothing methods were simple but effective for forecasting large numbers of items, many down to the SKU/location level, such as those characterizing inventory demand. Yet many statisticians, engineers, and econometricians decried the lack of a theoretical underpinning or statistical/mathematical elegance of these methods, failing to realize their value as practical forecasting tools. Instead, they touted more sophisticated/complex methods. And while there was evidence that the more complex methods proved superior in tracking historical data (the same data used to make the forecasts), there were doubts that they improved the accuracy of forecasting future data (post-sample time periods), at least until the wider utilization of machine-learning (ML) methods.

What distinguished forecasting, however, from other empirical sciences (especially statistics) was and continues to be its emphasis on testing the post-sample accuracy of forecasting methods. In a paper published in the *Journal of the Royal Statistical Society* (JRSS), Makridakis and Hibon (1979) reported two highly surprising findings concerning post-sample forecast accuracy:

- Among the two-dozen methods put to the test, the most accurate results were found using Brown's simple exponential smoothing adjusted for seasonality—a very straightforward, uncomplicated method.
- Second, averaging the forecasts of more than one method improved overall accuracy.

These findings were not well received by the statistical community of that time (Hyndman, 2020), which—taking steady aim at the messengers—often blamed incompetence for the results. In defense, Makridakis organized a study using 1,001 time series (Makridakis and colleagues, 1982). This time, however, anyone could submit forecasts, making this the first *true* forecasting competition.

This first M-competition and the additional competitions and empirical studies to follow provided the forecasting field with the equivalent of the controlled experimentation used in the physical sciences. This fundamentally changed the field of forecasting, separating facts from opinions and folklore, guiding academic research, and abetting the selection and usage of forecasting methods in practice (Hyndman, 2020).

The results of the first M-competition mirrored the findings that statistically sophisticated methods did not produce more accurate forecasts than simpler ones and that combining forecasts would on average improve forecast accuracy. These conclusions, now replicated through other competitions and individual studies, have at last been well accepted by the academic community (Armstrong, 2006).

Armstrong (1978) had concluded that time-series forecasting methods, based only upon the history of the items being forecast, were often more accurate than models using explanatory variables, a counterintuitive finding. In a more recent forecasting competition regarding tourism, Athanasopoulos and colleagues (2011) argued that explanatory variables can be useful, but only under two specific conditions: (1) when the future values of the explanatory variables are known or can be accurately forecast; and (2) when the measured impacts of the explanatory variables are likely to continue into the forecast period. Sometimes both conditions can be satisfied, such as for forecasting electricity demand when temperatures for a few days ahead are predictable, or when certain variables such as promotional activities in retail sales can be controlled. However, neither condition is always satisfied for tourism demand or many other areas of business forecasting.

Recent competitions have upgraded the potential value of sophisticated methods applied to large collections of data (Salinas and colleagues, 2017). The M4 Competition (2018) showed that those sophisticated methods incorporating machine learning (ML) were often more accurate than simple counterparts.

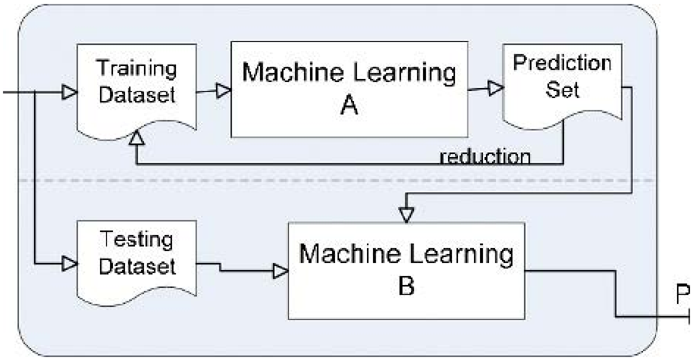
Thus ended a long "forecasting winter chill" against model complexity. The forecasting spring began with the M4 Competition, where a complex hybrid approach combining statistical and ML elements came in first place, while on average the top 16 methods were almost 5% more accurate than that of a common benchmark (Makridakis and Petropoulos, 2020). The top two methods, both hybrids of ML and time-series models, also achieved unforeseen success in estimating the degree of uncertainty in the forecasts, something normally underestimated.

Another important basis for the relative success of ML (in combination with time-series) models is their ability to learn from pooled data. This *cross-learning* results when data from multiple time series are linked in model estimation; for example, modeling groups of products or stores that share common elements of behavior. The top performers in many recent forecasting competitions used cross-learning to improve forecast accuracy over local univariate methods (Boger and Meldgaard, 2020). Applying ML to mobile payment data, Ma and Fildes (2020) report that "by capitalizing on the commonalities in the data across participating retailers, customer flow forecasting based on a large pool of stores from a variety of categories can generate forecasts that are more accurate than those generated by methods based on individual stores" (p.756), a result subsequently confirmed with promotional retail data.

Despite their impressive potential (Li and colleagues, 2020), *ensembles* of different methods remain broadly unadopted. Portfolio segmentation is becoming more widely used, but this has been glacially slow to emerge as a standard way to contend with the challenge of large numbers of time series to forecast with limited resources.

While the findings from the M4 and other recent competitions have elevated the promise of sustained improvements in forecasting methodology, especially through refinements in ML and hybrid models, the accuracy gains come at a huge cost (Gilliland, 2020) in model development and computation time. Gains in forecast accuracy, therefore, must be weighed against the increased costs, knowing that simple methods can achieve respectable levels of accuracy at a small fraction of the resources and effort.

In sum, there has been a slow diffusion of new ideas and approaches among



practitioners. We expect the future to deliver further improvements in both accuracy and estimation of uncertainty, hopefully along with processing efficiencies that increase the value and usefulness of forecasting in organizations and hence receptivity to promising new methods.

FUNDAMENTAL UNDERSTANDINGS

How can we convince decision makers of the benefits of systematic forecasting, while avoiding the formation of unreasonable expectations of what forecasting can deliver?

All organizational leaders know that forecasts are necessary for future-oriented decisions, including budgeting and planning activities. But their first option is often an ad hoc approach, delivering judgmental forecasts when required. Pure judgmental methods are too often aspirational—driven by optimism and the desire to achieve future goals versus assessing the objective reality of what is most likely to happen (Lawrence and colleagues, 2006).

The more challenging alternative is to establish a systematic set of procedures that produces forecasts from proven quantitative methods. Extensive research has shown that usage of systematic forecasting methods results in forecasts less susceptible to bias and superior in accuracy. In addition, some firms, including Johnson & Johnson, have reported major payoffs to improved forecast accuracy; see ***https://www.capgemini.com/us-en/client-story/johnson-johnson-transforms-its-demand-planning-and-external-manufacturing-processes/*** .

Makridakis and Petropolous (2021) summarize the fundamental understandings that decision makers must have.

Executives must understand:

- The forecast is an estimation of a future situation. It is not a target. It is not a plan. Nor is it an inventory decision. We may expect sales of 500 units (a forecast), but decide to stock 600 to minimize the risk of a stock-out to an acceptable level (a decision).
- Forecasting is not crystal-ball gazing. Forecasting methods, from the simplest to the most sophisticated, do not possess prophetic powers. Their predictions are based on identifying and estimating past patterns and/or relationships that are then extrapolated to forecast the future.
- All forecasts come with an error. All forecasts are uncertain with the only certainty being the existence of uncertainty.
- The most important advantage of systematic forecasting is its objectivity. It seeks to (a) identify past patterns and relationships to predict the future in a mathematically optimal manner, and (b) base estimates of the uncertainty in the forecasts on the volatility (variance) in the observed patterns/relationships.
- Forecasting accuracy and uncertainty can be estimated consistently in usual, everyday situations when established patterns/relationships remain fairly constant and can be extrapolated reasonably well.
- During periods of recessions/crises, or when unusual events occur, forecasting accuracy deteriorates—often significantly—while the level of uncertainty increases exponentially and sometimes cannot be measured quantitatively.
- There are trade-offs between the achieved forecasting performance and the respective resources needed (such as data availability, computational cost, and personnel time), so companies need to carefully choose an appropriate balance.
- When possible, forecasts should be assessed in terms of their utility (such as the decrease in the holding cost) instead of their forecasting accuracy.
- Lastly, while technology can substantially improve forecasting accuracy and our understanding of uncertainty, we cannot ignore the value of human judgment in the overall forecasting effort.

With all the focus on technology for the statistical modeling side of forecasting, there is also great opportunity for the augmentation of human judgment using artificial intelligence, ML, and even simple logic and business rules (Van Hove, 2020). For example, ML has shown promise in assisting the demand planner by identifying those forecasts most likely to benefit from adjustment, while also suggesting their direction and magnitude (Chase, 2019).

In the next section, we argue that setting proper expectations about forecast accuracy requires an understanding of the conditions that determine *forecastability*, including the distinction between "normal" vs. extreme behaviors and the inherent element of randomness in all behavior.

Extensive research has shown that usage of systematic forecasting methods results in forecasts less susceptible to bias and superior in accuracy.

AVOIDING UNREASONABLE EXPECTATIONS

Unreasonable expectations lead to disappointment and frustration when unexpected errors are blamed on the inadequacy of forecasting methods and processes. Some of these errors certainly result from the way the forecasts have been generated, but also from the inherent unpredictability of the forces being projected. Even if models could have forecast that the COVID-19 pandemic would occur, it would not have been possible to predict its exact timing and destructive economic impact (Osterholm, 2005), such as skyrocketing unemployment rates and the scarcity of bathroom tissue.

Normal vs Extreme Behavior

Let's consider that toilet-paper shortage. On March 12, 2020, U.S. bathroom-tissue sales ballooned 734% compared with the same day the previous year, becoming the top-selling product at grocery stores by dollars spent. Clearly, forecasts did not predict the huge surge in demand that created panic buying, demand exaggerated once photos of empty store shelves began circulating on social and mass media. Worse, the scarcity lasted for several months even as manufacturers rushed to produce and ship more paper. While this story reveals an aspect of our phenomenal failures during the pandemic, it also confirms the notable success of forecasts for all those normal time periods when toilet paper has been available to buy. We must distinguish the rare from the usual.

Analogy to Forecasting Time to Commute

Consider the challenge of forecasting the time it takes to travel to work in the morning. Most of us know very well how much time to allow to travel from home to work and back and realize that such time varies depending on different factors, such as the day of the week and the time one leaves. It is also evident that the commuting time varies even for the same day of the week and when leaving at the same time for any number of *uncontrollable* factors: a major road accident, highway roadwork, a sudden snowstorm, and so forth.

In the absence of these uncontrollable factors, deviations from the average time it takes to go to work are well behaved, most of the time being small, less often larger, and in rare cases substantial. We usually assume that these deviations follow a normal distribution, allowing the measurement of the variations or uncertainty around the average time it usually takes to travel to work each morning.

The extra time, however, that it will take to get to work in case of accidents, roadwork, and inclement weather is vastly different from the usual commute: it is not only highly uncertain, but cannot be expected to follow normal behavior. Rather, it creates a fat-tail distribution (Taleb, 2020) in which extreme travel times become more frequent, and forecasts cannot be found by simple extrapolation of past patterns, partly because we lack sufficient data on unusual events and also because we can't know whether they will recur and what impacts that will have. The 2020 coronavirus pandemic, which combines health and economic crises, presents a worst-case scenario in terms of the uniqueness of the lockdowns and their economic implications, making forecasting extremely difficult and uncertainty impossible to assess.

Distinguishing the normal from the extreme is of particular importance in how businesses set service levels and safety stock. In normal time, the risk of running short of product versus overstocking can be balanced by considering the costs of lost sales versus the cost of carrying extra inventories. However, during a pandemic or other period of major upheaval, the rules change: the degree of uncertainty is magnified and becomes difficult to assess. Consumers and suppliers aggravate the problem if they overreact, as with toilet paper during COVID-19.

Inherent Randomness

Even in normal time, forecast accuracy is limited by the extent of randomness in behavior. Using the previous example, your travel time to work may also vary with "subliminal factors" such as how tired you feel—you had a bad night's sleep—luck in just missing the change from a green to red light, a call that comes while driving, and so on. The degree of randomness in a variable determines its *forecastability*, and the quality of a forecasting method must be judged in light of the variable's forecastability.

> *If the nature of the demand is so gracious as to allow us to forecast it with 90% accuracy, then with good people, systems, and processes, we should be able to achieve that level of accuracy. But if the nature of the demand does not permit it to be forecast with 90% accuracy, then we never will ... no matter how much time and money and effort and sophistication we apply* (Gilliland, 2010).

The 2020 coronavirus pandemic, which combines health and economic crises, presents a worst-case scenario in terms of the uniqueness of the lockdowns and their economic implications, making forecasting extremely difficult and uncertainty impossible to assess.

There are ways we can reduce randomness, such as by aggregating data into more forecastable groups (e.g. using monthly rather than weekly data) or taking moving averages of volatile variables. Beyond a certain point, however, randomness cannot be reduced further, setting a limit to improvements in accuracy and lowering uncertainty. This is the notion of "unavoidable error" expressed by Morlidge (2013).

Hype

Forecasters are often the target of serious and, at times, legitimate complaints from forecasting users. Some of these surely come from negative experiences in the past and unrealistic expectations of what forecasting can achieve. We frequently hear arguments that if a forecast fails to achieve at least 90% accuracy, either the forecaster or the method used is not believable, this notwithstanding the margins for error reported in the forecast. Alas, consultants and software vendors are prone to exaggeration about the effectiveness of their forecasting toolbox. This is particularly the case with AI solutions and their brethren, which overpromise substantial accuracy improvements and problem-free implementation.

BARRIERS TO FORECAST IMPROVEMENT

There are many practical barriers to forecast usage, both in the initial phases of adoption and post-implementation, that can either stunt progress or render the process redundant. While our principal goal is to foster adoption of systematic forecasting in organizations where it is absent, we can build upon our knowledge of the impediments other organizations have faced to achieving their forecasting goals.

Challenging Preconditions

Preconditions for a systematic forecast methodology include having data sources (such as sales) that are at the appropriate levels, that don't suffer from latency, and that require minimal manipulation to eliminate erroneous or missing values. For some firms, the absence of such data creates a hurdle that requires cross-functional support and investment. Ideally, these data sources should be aligned to the master data used across the organization to provide a bridge to adoption in functional areas outside the one responsible for forecasting.

In the initial phases of adoption, there is often a lack of clear definition of how the forecast will be used—to support an operational process that consumes the data at a high level of frequency and detail, or to support a process that requires output at a higher level of aggregation, perhaps with a longer time horizon? Understanding the specific purposes of the forecasts is a key ingredient in process design. Too little attention can be paid to the units of measure, the time buckets to be used, and the hierarchy elements to include. There is limited understanding of supporting methods, such as clustering, to group similar hierarchical elements to provide the right balance of detail versus scale.

Process Design

Process design is often difficult because it requires cross-functional participation and engagement. It's often far simpler to design a process within a function, but this frequently fails to realize understanding, trust, and ultimately adoption by partners. In many instances, the lack of understanding of which inputs add value and which do not is a major cause of unsatisfactory outcomes.

There can be organizational anxiety about which function "owns" the forecast. In many organizations with an operational forecast output, ownership is in the supply-chain function. This is not to say that it can't also thrive in Sales or Finance. Much attention is paid to this, but little is given to decision rights (Gray, 2019). Who has the final say on the consensus forecast? If not thoughtfully considered, it can render systematic forecasting efforts redundant.

Forecasting Support Systems (FSS)

With the large number of FSS available, numerous selection considerations arise (Entrup and Goetjes, 2018). For those with existing ERP systems, should the tool be an advanced planning tool extension of that? Perhaps a best-in-breed solution is more appropriate? Many fail to follow a structured process of software selection, favoring what is suggested by the IT organization, often with little consideration of gaps or results from a proof-of-concept. Failure to consider these can lead to an unhappy partnership coupled with unfulfilled expectations.

Resistance can also come from the cost and difficulty of implementing an FSS. This is especially true for small and medium-sized firms. SMEs are unlikely to have the skilled staff to implement systematic forecasting nor the databases that these systems rely on. While there are cheap software products designed for such businesses, an additional barrier is that the use of a system may not match the way operations and tactical forecasting are carried out. Costly consultants may be required to ensure proper implementation and to train users.

Finding that small and medium-sized enterprises have lagged behind their larger counterparts in the adoption of suitable forecasting support systems, Matthias Luetke Entrup and Dennis Goetjes (2018) set out a structured process for the SME to identify, select, and implement an FSS that meets the organization's goals.

Metrics

Even when good designs and forecasting support systems are implemented, sustaining success and improvement can be elusive. Managing performance through the "right" metrics and applying improvement efforts specifically against those KPIs is a recipe for success. Too often, however, improvement efforts are applied against the biggest misses without consideration of what improvement is possible, considering the inherent unpredictability of the data.

Organizational Politics

Another source of resistance relates to human nature in the overall forecasting process. Forecasting can be a highly politicized process, with many human touch points. Each touch point becomes an opportunity for bias and personal agendas to contaminate what should be an objective, dispassionate process. Research has repeatedly shown that the more strategic the forecasts, even down to the annual budgeting cycle, the more senior (and inexpert) executives introduce bias and unnecessary inaccuracies.

The key questions then are how firms can achieve the most benefit from systematic forecasting, given that there are a wide selection of methods to choose from, many options for implementation, and a range of considerations in assigning responsibilities for the forecasting function.

GETTING STARTED FROM GROUND ZERO

Need for Historical Data

To initiate a systematic forecasting process, firms must recognize the necessity of developing a historical database. Doing so may require little or no monetary outlay. Initially there will be no need for consultants or expensive software. Instead, they would need to keep detailed information of the number of units sold at each time period of interest. Such data will allow firms to identify and exploit seasonality that contributes the most in improving forecasting accuracy. Later, they can record information about additional factors such as price, advertising, and promotions. These data can be also used for the objective estimation of budgets and cash-flow analysis.

Data should be captured at the most granular level (such as Item/Store for a retailer, or Item/Ship-to Location for a manufacturer, aggregated to days or weeks) and stored indefinitely (or aiming at least for 5+ years). Granular data can always be aggregated to higher levels based on product, location, or time hierarchies. Orders, sales/shipments, stockouts, and back orders would all be useful variables for constructing a time series approximating "true" customer demand.

For causal models, historical data on potential explanatory variables and other data features such as promotions, sales, and coupons need to be recorded. Implementation of ML algorithms benefits from such features as well as from data on related products.

Even when good designs and forecasting support systems are implemented, sustaining success and improvement can be elusive.

Exploring Naïve Methods and Developing Benchmarks

To evaluate forecasts, a firm needs benchmarks that put bounds on what can be achieved from historical data. A good start to generating benchmark forecasts is to explore several "naïve" forecast methods. The Naïve 1 and seasonal Naïve are two examples: for monthly sales forecasting, Naïve 1 uses the most recent month's sales as the forecast for the next month, while the sNaïve uses the sales of the same month of the previous year as the forecast for the current month. Analogous naïve forecasts can be calculated for data on daily, weekly, quarterly, or any other periodicity. The projections from a Naïve 1 reveal the future of sales if there is no change that increases or decreases sales from the most recent period. The projections from an sNaive extrapolate the seasonal pattern of sales from that in the most recent seasonal cycle.

Many other naïve variations are possible with simple arithmetic extrapolations of the data (e.g. the overall historical mean or median), testing their forecasting accuracy versus those produced within the firm. These simple benchmarks deliver a further benefit: when evaluating a more complex method (such as those proposed by a software house) they show how much of an improvement, if any, could be achieved from a potentially expensive new forecasting method. All too often they may reveal the inadequacy of the in-house forecasting processes: failure to beat the naïve is a damning indictment (Morlidge, 2014b).

When the scale of the data (number of time series) is relatively small, an inexpensive and ubiquitous tool like Excel could allow comparison of naïve forecasts to the internal judgmental or other projections made by the firm. (Larger firms with more time series would require more scalable data management like SAS.) Moving on from monthly forecasts, data can be also collected for weekly and daily sales figures to expand and benefit from the improved accuracy of systematic forecasting methods and the increasing need to plan on a shorter-term basis. Firms can also explore application of the methods to different periodicities (time buckets) such as weekly, monthly, and quarterly. There is good potential in averaging forecasts made from different time buckets (Petropoulos and Korentzes, 2014).

What we want to emphasize in this section is that a simple systematic forecasting system should be introduced step-by-step to test its value before more expensive solutions are adopted. Relative performance is best evaluated in relation to benchmarks, which will often highlight the need to adopt a more formal process of forecasting and evaluation. A common approach is that of calculating forecast value added, or FVA (Gilliland, 2013).

2019), although some learning effort is required to use it effectively. An alternative, Forecasting-as-a-Service (FaaS), is an emerging approach that some vendors are offering, which delivers cheap access to a variety of methods. We see software vendors increasingly offering ML methods—so, in principle, these advanced methods are becoming readily available, even though they cannot be used "out of the box."

As comfort with the software grows (and the historical database lengthens), the firm can begin experimenting with more advanced methods, comparing their effectiveness (and explainability and scalability) to the simpler methods. Available

For firms initiating a forecasting process, applying free and inexpensive software would allow them to see how well systematic forecasting fulfills their forecasting needs and how it can complement their managerial expertise.

Stepping into Forecasting Software

Almost every software package—including spreadsheet add-ins—will offer a set of forecasting procedures known as exponential smoothing. This family of procedure extends the naïve methods by utilizing weighted averages of the most recent historical data. For example, while a Naïve 1 forecast for June would be the actual sales in May, the simplest exponential-smoothing procedure would forecast June sales as a weighted average of May, April, March, and continuing back in time, giving less weight to each month the farther back it is in time. More sophisticated members of the exponential-smoothing family would similarly measure and project any trend and seasonal pattern in the historical data. See Stellwagen (2012) for an introductory tutorial on exponential smoothing.

In addition to spreadsheet add-ins, there are inexpensive commercial packages, most requiring little training to begin usage. Fildes and colleagues (2020) have recently provided a survey of commercial software and their features. An increasingly popular solution that allows the usage of all popular forecasting methods is the free R library (Hyndman,

forecasting methods range from the extremely simple, such as single exponential smoothing, to the highly sophisticated, such as deep learning (DL), which require specialized knowledge and substantial computer resources to run. Both types of methods could be useful; the first when large numbers of forecasts are needed and there are constraints on time and resources to create them, and the second when even small improvements in accuracy/uncertainty are important to save large amounts of money by improving decision making in critical business areas. There are also methods of intermediate complexity. These can be considered by balancing accuracy/uncertainty versus interpretability and ease of use, as well as the computer time required to obtain the forecasts and measures of uncertainty.

There is a considerable body of knowledge to be found, including on the Web, in the many forecasting books, and in journals such as this one. These resources show how various methods work, when they work well, and when they seem to fail.

For firms initiating a forecasting process, applying free and inexpensive software would allow them to see how well systematic forecasting fulfills their forecasting needs and how it can complement their managerial expertise. Many organizations will find that shifting from purely judgmental to systematic methods of forecasting will provide a more reliable basis for their operational decisions.

Judgmentally Adjusting Statistical Forecasts

Forecasting methods are accurate if established patterns/relationships do not change during the forecasting period. This means that any changes such as a large order from a new customer, a major new promotional campaign, a significant price reduction, or a competitor going out of business will not be included in the forecast model, and thus will have to be incorporated into the final predictions judgmentally. A novel promotion would probably justify judgmental intervention, but in some cases we may have a sufficient record of the effectiveness of past promotions or price reductions to justify statistical modeling of their effects. Equally importantly, we should not let the optimism about the potential success of the promotion unduly influence its forecast.

Judgmental adjustments present a major management challenge. Advice in the forecasting literature on how to manage adjustments include:

- Avoid small adjustments to the forecast—even if directionally correct, they have at best a small impact on forecast accuracy and have little effect on decision making. Rather, concentrate on large adjustments that will impact the future by requiring changes to existing plans.
- Recognize and attempt to minimize optimistic biases in judgmental adjustments of statistical forecasts.
- Keep track of and document the reasons for the adjustments. Doing so reduces gratuitous adjustments and enables us to determine their forecast value added (Gilliland, 2013)—which adjustments are justified and which aren't.

Judgmental adjustment of statistical forecasts is attractive to executives for many reasons; it offers the forecaster control and allows the incorporation of myriad factors not included in the model. Particularly with complex ML methods, managers are "algorithm averse": they prefer to rely on their own judgments rather than on incomprehensible models. While research has shown the need to improve the effective incorporation of judgment into the statistical forecasts, for many companies this has proved difficult. The online Appendix [***https://foresight.forecasters.org/wp-content/uploads/UFOAPPENDIX_Aug26-2020.pdf***] summarizes key studies about the desirability and impact of judgmental adjustments and the manner in which they should be implemented.

OFFERING GUIDELINES

To organizations endeavoring to create systematic forecasting, we have few guidelines at present to offer that demonstrate an awareness and understanding of what constitutes best practices in the field. Some attempts at such guidelines include Morlidge (2010) and Smith and Clark (2011). Lacking such guidelines, companies may seek role models in other firms, and surveys of similar size organizations that have successful forecasting functions should be valuable. A useful preliminary to these surveys is holding direct interviews to identify successful firms and understand how they are utilizing forecasting.

We need also to conduct interviews with firms that do not use formal forecasting, to determine what information and motivation they would require to initiate a systematic forecasting process. To support this initiative, the Makridakis Open Forecasting Center (MOFC) at the University of Nicosia will sponsor a project of interviews and questionnaires, with *Foresight* serving as co-sponsor and forum for publication of results. Producing a set of guidelines for proper forecasting usage, as well as an inventory of best practices, will provide a valuable service to the field and increase the use of systematic forecasting. It may also help to identify "bad" practices, make firms aware of their negative consequences, and offer recommendations on how to do better.

CONCLUSIONS

The field of forecasting has advanced a great deal in recent years, while data availability and computer power have seen spectacular increases. The more apparent benefits of systematic forecasting should make adoption of such a process much more advantageous to organizations that have not yet "seen the light."

Producing a set of guidelines for proper forecasting usage, as well as an inventory of best practices, will provide a valuable service to the field and increase the use of systematic forecasting.

A key challenge is that of persuading more organizations of the considerable benefits from systematic forecasting. The central argument is the gain in business efficiency, accountability, and profitability that firms stand to realize utilizing systematic forecasting methods versus those with ad hoc judgment. Ultimately, the challenge is how to demonstrate to skeptics that a scientific/statistical approach to forecasting, while imperfect, still works better than the alternatives.

While we have focused our remarks on operational and tactical forecasting, even with strategic analysis some major components will depend on analytical methods. To establish credibility, this requires acknowledging to practitioners—and skeptical management—that a scientific/statistical approach often does not work very well because of the inherent limitations on forecastability. It also requires recognition, by all parties, of the difficult and challenging dilemma in which the forecaster is placed: having to show confidence about his or her predictions to management, while at the same time providing management with what can amount to a wide range of uncertainty around the forecasts.

Of one thing we are certain, however: forecasting skeptics are so used to the hype and overpromises of consultants and vendors that they are reluctant to believe anything. This can only be addressed, and must be addressed, with a refreshing dose of candor.

REFERENCES

Armstrong, J.S. (2006). Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error, *International Journal of Forecasting*, 22(3), 583–598.

Armstrong, J.S. (1978). Forecasting with Econometric Methods: Folklore versus Fact, *Journal of Business*, 51(4), 549-564.

Athanasopoulos, G., Hyndman, R.J., Song, H. & Wu, D.C. (2011). The Tourism Forecasting Competition, *International Journal of Forecasting*, 27(3), 822–844.

Boger, C. & Meldgaard, J. (2020). The M5: A Preview from Prior Competitions, *Foresight*, Issue 58 (Summer), 17-23.

Brown, R.G. (1959). Statistical *Forecasting for Inventory Control*, McGraw-Hill, New York.

Chase, C. (2019). Assisted Demand Planning Using Machine Learning for CPG and Retail, SAS whitepaper.

Entrup, M. & Goetjes, D. (2018). A Blueprint for Selecting and Implementing a Forecasting Support System, *Foresight*, Issue 50, 10-18 & Issue 51, 8-15.

Fildes, R. & Goodwin, P. (2007a). Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting, *Interfaces*, 37(6), 570–576.

Fildes, R. & Goodwin, P. (2007b). Good and Bad Judgment in Forecasting: Lessons from Four Companies, *Foresight*, Issue 8 (Fall) , 5-10.

Gilliland, M., Tashman, L. & Sglavo, U. *Business Forecasting (Vol 2): The Emerging Role of AI and Machine Learning* (Wiley).

Gilliland, M. (2020). The M4 Forecasting Competition – Takeaways for the Practitioner, *Foresight*, Issue 57 (Spring), 5-10.

Gilliland, M. (2013). FVA: A Reality Check on Forecasting Practices, *Foresight*, Issue 29 (Spring), 14-18.

Gilliland, M. (2010). *The Business Forecasting Deal*, Hoboken, NJ: John Wiley & Sons.

Gray. C. (2019). Why Is It So Hard to Hold Anyone Accountable for the Sales Forecast? *Foresight*, Issue 54 (Summer), 38-43.

Hyndman R.J. (2020). A Brief History of Forecasting Competitions, *International Journal of Forecasting*, Issue 36 (1), 7-14.

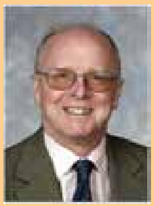
Hyndman, R.J. (2019). Forecasting Functions for Time Series and Linear Models, *RDocumentation* ***https://www.rdocumentation.org/packages/forecast/versions/8.10***

Hyndman, R.J. & Khandakar, Y. (2008). Automatic Time Series Forecasting: The Forecast Package for R, *Journal of Statistical Software*, 27(3), 1–22.

Spyros Makridakis is Professor and Director of the Institute for the Future at the University of Nicosia.

Ellen Bonnell is a consultant and author of "How to Get Good Forecasts from Bad Data" (*Foresight* Issue 7, Summer 2007).

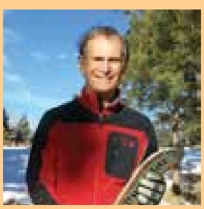
Simon Clarke is a Principal of Crimson & Co and formerly Group Director of Forecasting at Coca-Cola.



Robert Fildes is Distinguished Professor of Management Science at Lancaster University and Director of the Lancaster Centre for Forecasting.

Mike Gilliland is Marketing Manager for SAS forecasting software and *Foresight* Associate Editor.

Jim Hoover is Director of the Business Analytics program at the University of Florida and Chairman of the *Foresight* Advisory Board.



Len Tashman is *Foresight* Editor.

Lawrence, M., Goodwin, P., O'Connor, M. & Onkal, D. (2006). Judgmental Forecasting: A Review of Progress Over the Last 25 Years, *International Journal of Forecasting*, 22(3), 493-518. ***https://doi.org/10.1016/j.ijforecast.2006.03.007***

Li, Y., Berry, D. & Lee, J. (2020). How to Choose Among Three Forecasting Methods: Machine Learning, Statistical Models, and Judgmental Forecasts, *Foresight*, Issue 58 (Summer), 7-14.

Ma, S. & Fildes, R. (2020). Forecasting Third-Party Mobile Payments with Implications for Customer Flow Prediction, *International Journal of Forecasting*, 36:3, 739-760.

Makridakis, S. & Petropoulos, F. (2020). The M4 competition: Conclusions, *International Journal of Forecasting*, 36(1), 224–227.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V. (2020). The M4 Competition: 100,000 Time Series and 61 Forecasting Methods, *International Journal of Forecasting* ***http://dx.doi.org/10.1016/j.ijforecast.2019.04.014.***

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.G. & Hibon, M. (1979). Accuracy of Forecasting: An Empirical Investigation (with Discussion), *Journal of the Royal Statistical Society*, Series A, 142, 97–145.

Morlidge, S. (2014a). Do Forecasting Methods Reduce Avoidable Error? Evidence from Forecasting Competitions, *Foresight*, Issue 32 (Winter), 34–39.

Morlidge, S. (2014b). Forecast Quality in the Supply Chain, *Foresight*, Issue 33 (Spring), 26–31.

Osterholm, M.T. (2005). Preparing for the Next Pandemic, *Foreign Affairs* ***https://www.foreignaffairs.com/articles/2005-07-01/preparing-next-pandemic***

Petropoulos, F. & Kourentzes, N. (2014). Improving Forecasting via Multiple Temporal Aggregation, *Foresight*, Issue 34, 12-17.

Salinas, D., Flunkert, V., & Gasthaus, J. (2017). DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks, arXiv preprint arXiv:1704.04110.

Smith, J. & Clarke, S. (2011). Who Should Own the Business Forecasting Function? *Foresight*, Issue 20, 4-7.

Stellwagen, E. (2012). Exponential Smoothing: The Workhorse of Business Forecasting, *Foresight*, Issue 27 (Fall), 23-28.

Taleb, N.N. (2020). Statistical Consequences of Fat Tails: Real World Preasymptotics, Epistemology, and Applications, ***https://arxiv.org/abs/2001.10488***

Van Hove, N. (2020). Technology Support in Business Planning: Automation, Augmentation, and Human Centricity, *Foresight*, Issue 58 (Summer), 43-48.

UFO ARTICLE APPENDIX

Judgmental Forecasts and Adjustments to Statistical Forecasts

As early as 1986, Mathews and Diamantopoulos (1986, 1992) documented improvements in forecast accuracy through judgmental overrides of baseline forecasts. Their more recent study reported only marginal improvements and recommended a formal system for reviewing overrides. Goodwin and Wright (1993) argued for a much greater understanding of the cognitive processes adopted in judgmental forecasting tasks.

Judgmental Forecasting: A Review of Progress Over the Last 25 Years (Lawrence and colleagues, 2006) Reviewing more than 200 studies, they concluded that human judgment can be of significant benefit but is also subject to significant biases. They suggest steps to overcome judgmental biases and recommend keeping track of the judgmental forecasts for feedback and evaluation.

Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting (Fildes and Goodwin, 2007a, 2007b) Fildes and Goodwin collected data on more than 60,000 forecasts in four companies, finding examples of good forecasting practice but frequent failures to follow basic principles. They recommend limiting the frequency of judgmental adjustments of statistical forecasts by requiring managers to justify such adjustments in writing, avoiding small adjustments, and recognizing a bias toward optimism.

Forecast Quality in the Supply Chain and ***Do Forecasting Methods Reduce Avoidable Error?*** (Morlidge, 2014b and 2014a) Morlidge's analysis of the M3 data found that forecasts produced by experts under controlled conditions, with no difficult-to-forecast series, still failed to beat a naive forecast 30% of the time. Moreover, his study of nine datasets covering 17,500 products over an average of 29 (weekly or monthly) periods reported that 52% of forecasts made did not improve upon naive projections. He also found that only 5% of the 17,500 products had errors on average less than half those of naïve forecasts, positing this as a reasonable estimate of the practical lower limit for forecast error.

Judgmental Forecast Adjustments Over Different Time Horizons (Van den Broeke and colleagues, 2019) This paper analyzed over 300,000 forecasts to determine how the size, direction, and accuracy of judgmental adjustments changed across different time horizons. They found that as the point of sales is approached, adjustments become larger and more positive. These shifts can put pressure on operations and lead to conflicting signals for the time-phased production planning decisions, increasing production costs that could be avoided with stable forecasts.

M2 Competition: A Direct Comparison of the Accuracy of Statistical Methods and Human Forecasters (Makridakis and colleagues, 1980)

The M2 competition compared the forecasting accuracy of statistical methods with those of five human forecasters for predicting the monthly budget figures of four companies on a real-time basis for two consecutive years. Overall, average accuracy of the three exponential-smoothing methods was superior to that of the five human forecasters, suggesting that extrapolating historical patterns with exponential smoothing methods can produce more accurate budget forecasts than can forecasters who spend time studying the data.

In summary, all studies agree on the potential value of judgmental adjustments to include new information and domain knowledge not represented in the statistical models. At the same time, all studies agree that judgment is subject to biases, most often on the optimistic side with the frequent result that forecast accuracy deteriorates. The challenge is to be able to incorporate judgment in a systematic way that ensures objectivity and avoids undue optimism.

References

Fildes, R. & Goodwin, P. (2007a). Against Your Better Judgment? How Organizations Can Improve Their Use of Management Judgment in Forecasting, *Interfaces*, 37(6), 570–576.

Fildes, R. & Goodwin, P. (2007b). Good and Bad Judgment in Forecasting: Lessons from Four Companies, *Foresight,* Issue 8, 5-10.

Lawrence, M., Goodwin, P., O'Connor, M. & Onkal, D. (2006). Judgmental Forecasting: A Review of Progress Over the Last 25 Years, *International Journal of Forecasting*, 22(3), 493-518. https://doi.org/10.1016/j.ijforecast.2006.03.007

Makridakis, S. (1993). The M2 Competition: A Real Time Judgmentally Based Forecasting Study, *International Journal of Forecasting*, 9 (1), 5-22.

Mathews, B.P. & Diamantopoulos, A. (1992). Judgmental Revision of Sales Forecasts: The Relative Performance of Judgmentally Revised versus Non-revised Forecasts, *International Journal of Forecasting*, 11(6), 569-576.

Mathews, B. P., & Diamantopoulos, A. (1986). Managerial intervention in forecasting. An empirical investigation of forecast manipulation, *International Journal of Research in Marketing, 3(1), 3-10.* https://doi.org/1 0.1016/0167-8116(86)90038-8

Morlidge, S. (2014a). Do Forecasting Methods Reduce Avoidable Error? Evidence from Forecasting Competitions, *Foresight*, Issue 32, 34–39.

Morlidge, S. (2014b). Forecast Quality in the Supply Chain, *Foresight*, Issue 33, 26–31.

Van den Broeke, M., De Baets. S., Vereecke, A., Baecke, P., & Vanderheyden. (2019). Judgmental Forecast Adjustments Over Different Time Horizons, *Omega*, 87, 34–45.