Making the Best Use of Judgmental Forecasting

BY SHAYNE KAVANAGH AND DANIEL WILLIAMS
Governments use two methods of revenue forecasting — judgmental, which relies on the expert opinion of the forecaster, and quantitative, which uses historical data and statistical techniques. Most local government forecasts rely on forecaster judgment to some extent, in large part because this method is easy to use. There is also value in human judgment, under certain conditions, but judgment is inherently biased. Research has shown judgmental forecasting to be generally inferior to quantitative methods, but in situations where no data exist for statistical techniques, judgmental forecasting can provide useful insights. To determine how judgmental forecasting can best be used, one must consider:

- **Efficacy.** What does forecasting science say about the predictive value of judgmental forecasting?
- **Conditions for Use.** When is it appropriate to use judgmental methods in forecasting?
- **The human side of judgmental forecasting.** How can the user best address the human element of judgmental forecasting?
- **Judgmental forecasting methods.** What techniques make judgmental forecasting most effective?

**Efficacy**

Forecasting science generally recommends against pure judgmental forecasting for two main reasons. First, the process used to construct the forecast exists primarily in the forecaster’s head, which means the forecast is neither transparent nor easily replicated. Hence, the forecast is difficult to fully explain to others, and there would be variations among forecasters, leading to inconsistencies.

Second, a wide body of literature suggests that judgmental forecasts are likely to be of dubious accuracy. One particularly well known example comes from a landmark study by the political scientist, Phillip Tetlock. Over the course of 15 years, Tetlock asked 284 experts to assign probabilities to one of three possible future scenarios for forecast questions that were germane to their fields (e.g., economics, domestic politics, international relations). The three available choices for each scenario covered persistence of the status quo, a change in one direction (e.g., more economic growth in a given country), or a change in the opposite direction (e.g., a downturn in economic growth).

Expert judgment did not perform well in the study. A New Yorker review of Tetlock’s work put it memorably: “The experts performed worse than they would have if they had simply assigned an equal probability to all three outcomes — if they had given each possible future a 33 percent chance of occurring. Human beings who spend their lives studying the state of the world, in other words, are poorer forecasters than dart-throwing monkeys, who would have distributed their picks evenly over the three choices.” In all cases, even rudimentary statistical methods would have provided more predictive power. Further, these disappointing results were consistent across all areas of expertise, experience, or degree of specialization. In other words, greater expertise does not lead to better judgmental forecasts.

**Use Any Quantitative Technique**

As a rule of thumb, a forecaster should always try to apply some quantitative technique — any quantitative technique — before relying solely on expert judgment.

Tetlock’s findings do not mean that expert judgment has no role in forecasting, however — just that it should be judiciously applied. The next section of this article addresses the conditions under which judgment should enter into the forecast.

**Conditions for use of Expert Judgment**

Judgmental forecasting is not universally panned by the research. Under certain conditions, some research suggests that judgmental forecasting can be as good as the best statistical techniques, and may be more consistent. This can be the case when:
The expert has access to information that would not be reflected in the results from a statistical model. This can be a common occurrence in government revenue forecasting because of phenomena such as changes in the tax base or changes in governing law.

The expert has a history of making and learning from similar forecasts, and the environment is relatively stable (i.e., the impact of seasonality or business cycles is low).

Good historical data are limited or unavailable.

A government should exercise judgment carefully even if any of these conditions exist. Techniques that can help improve judgment include the following:

- Start by documenting the organization’s accumulated wisdom about the revenue source. A simple checklist of key factors documents what should be considered before the forecast is made. This helps forecasters stay focused on the most relevant factors, not overlook key factors, and maintain some consistency in forecasting approach.

- Keep records of how forecasts are made and review past performance before making new forecasts. Feedback allows the forecaster to learn. As such, it is advisable to keep records — not only of the forecast itself, but also of judgments about key variables and assumptions behind the forecast. A forecast may ultimately be the product of too many variables to provide useful feedback, but reviewing how the expert’s judgment affected specific key variables may be more instructive than the forecast itself.

- Create graphics of key trends rather than simply studying data in tabular form. Graphical representations of data can help reveal trends that might otherwise be missed. Consider different graphical formats for visualizing data, such as drawing a line of best fit through the data points to better illuminate the trend.

- Obtain several independent judgments rather than relying on just one person. Research suggests that judgmental forecasts can also be improved by simply averaging the results of multiple independent forecasts. This averages out unsystematic differences in the forecasters.

Even in a less structured approach, getting more views for a judgmental forecast can be helpful — but beware of group forecasting. It may seem intuitive that a team would probably arrive at better judgments than an individual would, but this is not always the case. Group interactions can be dominated by a particularly confident (but not necessarily insightful) participant, or they can devolve into “group think.”

**Decompose the Forecast Problem.** Breaking down the forecast into component problems can help simplify the forecasting task. For example, forecast income from different components of the tax base, not the entire tax base. Only use this technique on problematic areas, however. Keep decomposition to a minimum because research shows that overuse can increase the number of judgments required and thus exhaust the forecaster.

**Require Justification of the Forecast.** Forecasters who are required to verbally justify their forecasts are more consistent; if they know they will have to explain their reasoning to others, they tend to develop more consistent reasoning. A written justification can be used to help evaluate the forecast methodology once results become available.

**Start with an Algorithm.** Judgmental forecasting does not mean pulling a forecast from the ether. Quantitative data
can be used to help develop a starting point for the forecast; a traditional statistical model could fill this role. If that is not possible, even a simple, non-statistical algorithm would likely represent an improvement over a forecast using expert judgment alone. In fact, forecasters should be able to develop at least a simple algorithm for virtually any revenue source. For instance, property tax revenue would be a product of the assessed value multiplied by the tax rate, multiplied by the collection rate. Research suggests that even a simple algorithm that reflects factors that were included in development of the forecast can significantly improve decision making.

**TO ERR IS HUMAN: COGNITION IN FORECASTING**

Thinking back to Tetlock’s study, one might wonder why people are so poor at forecasting. Much of the answer has to do with our cognitive biases — deviations in judgment that are inherent to the way the human mind works and lead people to draw irrational conclusions from their observations. Cognitive biases operate unconsciously, and their influence is often subtle.

Consider, for example, the cognitive bias of overconfidence. Cognitive psychology suggests that about 80 percent of people have an inborn tendency to overestimate their chances of experiencing good events and underestimate their chance of experiencing bad events. In other words, people are more optimistic than realistic. People also regularly overestimate their own capacities. For example, most people believe they are more attractive, personable, honest (and so on) than the average person. While an optimistic outlook on life has many benefits, better financial forecasting is not one of them. In fact, of the many cognitive biases that can afflict forecasting, optimism is thought to be one of the most important. Optimism could lead a forecaster to underestimate the cost of a large project or to be overly confident about the assumptions behind a revenue forecast. Exhibit 1 lists other common biases in forecasting.

**MITIGATING THE BIASES**

Daniel Kahneman is a psychologist who won the Noble Prize for economics for his work relating to decision making, including cognitive biases. Kahneman is not optimistic about

---

**Exhibit 1: Example of Biases in Forecasting**

<table>
<thead>
<tr>
<th>Type of Bias</th>
<th>Description of Bias</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overconfidence</td>
<td>Unjustified belief in the accuracy of forecast assumptions</td>
<td>Being too certain about the potential future behavior of key revenue drivers</td>
</tr>
<tr>
<td>Conservatism</td>
<td>Failing to change one’s mind in light of new information</td>
<td>Ignoring new, less favorable information on the financial impact of a new development project</td>
</tr>
<tr>
<td>“Recency”</td>
<td>Allowing recent events to dominate over events further in the past</td>
<td>Allowing news headlines of recent economic events to obscure larger, long-term trends</td>
</tr>
<tr>
<td>Availability</td>
<td>Relying on specific events that are easily recalled from memory at the expense of other pertinent information</td>
<td>Making issues that are particularly memorable and may have received a lot of political attention the focus of a forecast, even though other issues were more important</td>
</tr>
<tr>
<td>Anchoring</td>
<td>Being unduly influenced by initial information, which is given more weight in the forecasting process</td>
<td>Using a historical high or low as the starting point of discussions for a long-term, multi-year forecast</td>
</tr>
<tr>
<td>Search for Supportive Evidence</td>
<td>Gathering facts that lead to certain conclusions and disregarding contradictory information</td>
<td>Consulting only sources of economic analysis that support the outcome that the forecaster wants</td>
</tr>
<tr>
<td>Underestimating Uncertainty</td>
<td>Underestimating future uncertainty in order to reduce anxiety</td>
<td>Showing a point forecast or single forecast as the only future outcome and not accounting for possible variance</td>
</tr>
<tr>
<td>Selective Perception</td>
<td>Seeing problems in terms of one’s own background and expertise</td>
<td>Considering only the forecast variables that the forecaster knows the most about</td>
</tr>
</tbody>
</table>
the possibilities for eliminating these biases, which are largely hardwired into human cognition. However, Kahneman and other researchers have suggested methods to mitigate the worst impacts of these biases (explained below):

- Promote discussion of biases and their impact on forecasting.
- Actively question your assumptions.
- Use reference case forecasting.
- Be careful about the forecast starting point.
- Stick with a simple representation of future revenues.

Participants in the forecasting process can use the information in Exhibit 1 to make themselves aware of potential cognitive biases and, ideally, help each other recognize when the forecasting process may be approaching areas that are prone to biases. Implicit in this recommendation is that having more than one person involved in the forecasting process is advisable. People are better at recognizing biases in others than in themselves, and having multiple people involved slows the forecasting process down, providing more time for rational thinking. (Going too fast invites the human brain to rely on its biased intuition.) It also provides a quality check, as the forecaster will need to justify his or her assumptions.

Forecasters should take conscious steps to critically examine their own forecast assumptions, which are often significantly affected by biases. One simple method is to write down all important forecast assumptions, along with two reasons why that assumption might be wrong, thus prompting introspection. Kahneman recommends a “pre-mortem” — once a forecast is reached, but before it is made public, the forecasters should hold a brief meeting at which they imagine that, a year in the future, the forecast was horribly wrong. Writing a brief history of what happened — the pre-mortem — helps temper uncritical optimism and forces people to consider ways in which their initial estimates might be wrong.

Before committing to a forecast, forecasters should examine similar forecasts or cases from other organizations or instances to see how they turned out, and if these experiences are consistent the forecasters’ expectations for their own work. For example, if the organization needs to forecast revenues for a new tax or fee, the forecasters should find out if similar governments have done so, and what their experience was.
Forecasters are often tempted to use the last actual observation as the starting point for the forecast, but the most recently available actual number is not necessarily the appropriate starting point. If the most recent observation is unusually high or low, the entire forecast would be biased in that direction (this is an example of the “anchoring” cognitive bias from Exhibit 1). Instead of the most recently available actual number, forecasters should consider using an average of recent observations to pick their starting point.

It can be tempting to add ups and downs to the forecast, in an attempt to simulate business cycles — especially when making long-term forecasts. But while this might make the forecast seem more “realistic,” it will likely introduce additional inaccuracies.

**Foxes and Hedgehogs.** Cognitive biases do much to explain the limitations of judgmental forecasts, but Tetlock’s work suggests an additional explanation. While Tetlock did not find that experience or degree of specialization made a difference in an expert’s forecast accuracy, he did find one variable that seemed to make a difference. Tetlock categorized the forecasters into one of two cognitive styles: “hedgehogs” and “foxes.” In short, hedgehogs subscribe to one or a few clear overriding ideas or approaches to a question, while foxes take more of a multidisciplinary approach, using many ideas and changing approaches as circumstances suggest. Foxes are thought to be better forecasters because the way they think better suits the forecasting task, and indeed, Tetlock’s experiments demonstrated that foxes outperform hedgehogs by a significant margin. In fact, the performance of hedgehogs was worse than random guessing. The best foxes approached the accuracy of some quantitative extrapolation models. Exhibit 2, adapted from the work of noted statistician and forecaster Nate Silver, illustrates the differences in attitudes between the two in more detail.

**JUDGMENTAL FORECASTING METHODS**

This section addresses two specific methods of applying expert judgment to the forecast. First is the “Delphi” method, which relies solely on expert judgment to produce a forecast. Second is a discussion of non-statistical algorithms. While statistical models are preferred for forecasting, a non-statistical algorithm is better than no algorithm at all.

**The Delphi Method.** The Delphi method is a traditional way of producing a forecast using pure expert judgment, and it is widely accepted by forecast science. The Delphi method relies on a panel of five to 20 experts who supply forecasts over two or three rounds. After each round, a facilitator provides an anonymous summary of the experts’ forecasts from the previous round, along with the reasoning behind their judgments. Experts are thus encouraged to revise their earlier answers in light of the replies given by other members of the panel. During this process, the range of the answers is expected to decrease, and the group is expected to converge toward the optimal answer. The process ends
after a predefined stop criterion is met (e.g., number of rounds, achievement of consensus, stability of results), and the mean or median scores of the final rounds determine the final forecast. Generally, three rounds are thought to be sufficient.

Delphi works because the experts are anonymous, so the forecasts are not influenced by social pressures, and the forecasters have the opportunity to change their opinions over the rounds without embarrassment. In addition, the feedback is controlled so a dominant personality cannot take over, and the responses are averaged in the end. The following principles will help forecasters make the best use of the Delphi method:

- Use experts with appropriate knowledge of the revenue being forecast. Non-experts might be prone to changing their views to match the opinions of others, or their estimates may be the product of little more than guessing.
- Use experts with different perspectives. This way, the combined knowledge of the experts should cover a more comprehensive scope of the factors that affect revenue yield. For example, a forecast for the property tax might include people from the finance, assessor, collector, and community development departments.
- Provide information about all results to the group. This feedback must include a narrative rationale from each of the forecasters; otherwise, forecasters are responding to what appears to be a series of random numbers. Supply forecasters with the rationales for all of the forecasts, not just some of them.

**Non-Statistical Algorithms.** Forecasters are not always able to build statistical models, perhaps because they lack statistical expertise or because there is a lack of quality historical data. Rather than creating a forecast from pure judgment, forecasters in this situation should use their knowledge of the factors that drive revenue yields to construct an algorithm to guide the forecast.

The algorithm should be developed using a spreadsheet or other software. It can take many forms, and some basic principles to observe are listed below. While this section was written with non-statistical algorithms in mind, many of the same principles can apply to statistical models as well.

**The Software Has a Clear Purpose.** A forecasting spreadsheet or other piece of software might have multiple purposes — the most obvious being to develop and use a forecast algorithm. The software might also be used to create graphical presentations of the forecast or to demonstrate possible financial scenarios, building on the baseline forecast. This is a situation where one-stop shopping isn’t ideal, however; it is often better to use separate software for distinct purposes (especially when using Excel-based programs) to keep each program as streamlined as possible. For example, a dedicated forecasting algorithm will often focus on one particularly important and/or complex revenue source, and another model might be used to present the forecast. (For more information, see “Forecasting Technology: The State of the Market” on page 54.)

**The Elements of the Algorithm Are Transparent.** A revenue forecast algorithm will comprise a number of component parts, each of which should be obvious to a user. This makes the algorithm more credible and makes it easier to transfer knowledge of how to use the algorithm from one person to another. At a minimum, a forecasting algorithm should make obvious the following elements:

- **The forecast.** The algorithm will produce a forecast which should be clearly shown, along with any adjustments that the forecaster makes.
- **Decompositions.** An algorithm may decompose a revenue source into smaller parts for easier forecasting. For example, a property tax algorithm might have residential and commercial components, or a user fee algorithm might take customer segments into account. These decompositions clarify the domain knowledge the forecaster put into building the algorithm.

- **Historical data.** All historical data used in the algorithm should be labeled as such. Besides the date of the data, an algorithm should note any important adjustments made for modeling purposes. For example, an Excel “comment” could be used to note where an outlier was removed.

- **Explicit assumptions made.** An algorithm will often make a number of explicit assumptions that have a direct mathematical role in the algorithm. To illustrate, if a water utility billing forecast is made by multiplying an assumed price per gallon by the number of gallons sold, then the assumed price per gallon should be clear. In Excel, the price per gallon might be contained in a conspicuous area of the worksheet — its own cell. Formulas in the workbook would then reference that cell, rather than building the price directly into the formula.

**Implicitly Made Assumptions are Transparent.** Because an algorithm is a simplification of reality, there will always be some assumptions made about the social, technological, environmental, economic, and political forces that could impact the revenue. Most of these assumptions will not be expressed mathematically in the algorithm, so the modeler should highlight the most critical ones. For example, sales taxes in most jurisdictions are strongly influenced by business cycles. In spite of this, a long-term sales tax forecast algorithm would not likely attempt to model the impact of future business cycles, which are impossible to predict with any degree of accuracy. Hence, an implicit assumption in the algorithm is that the forecast period does not experience dramatic economic downturns or expansions. Forecasters should always be forthcoming about such simplifications of reality and be able to explain why the model is still useful.

**The Algorithm Is as Simple as Possible.** Algorithms should be easy to understand and not overly complex. In addition to the general rule of trying not to accomplish too much in a single algorithm, other spreadsheet practices that should be observed include separating data from formulas. Data should never be typed into formulas; instead, they should be kept in their own area of the spreadsheet and incorporated into formulas via cell references. This makes it possible to clearly label the data and update the algorithm if the data change.

**Test and Validate the Model.** Just like statistical techniques, a non-statistical algorithm should be tested for accuracy before its results are put to use. Try simulating the algorithm’s results. Also, the Excel “formula audit” tools can be used to visually map the relationship between the cells that are referenced in a given formula, allowing the forecaster to verify that there is a logical flow in the way information is laid out.

**The Algorithm is Reusable.** The algorithm should be able to evolve. Providing room to add more historical data (without losing older data), using modular designs (rather than monolithic), and adhering to the principles of algorithm simplicity (as described above) all contribute to the reusability of an algorithm. As previously mentioned, one key Excel practice is using a cell range as an argument in a formula. The “sum” formula is a well-known example of this — a range of cells can be entered into the formula rather than adding each individual cell together. In this way, cells can be easily added and subtracted from the model without having to rewrite the formula. The “sumproduct” formula can be particularly useful for an algorithm-builder; it allows the forecaster to use ranges to sum up the products of multiple ranges of cells. “What-if” capabilities are also important for allowing one algo-

---

Forecasters who are required to verbally justify their forecasts are more consistent; if they know they will have to explain their reasoning to others, they tend to develop more consistent reasoning.

“All models are wrong, but some are useful.”

— George E.P. Box, statistician and member of the American Academy of Arts and Sciences

---
algorithm to answer multiple questions. This is done by separating key variables from formulas, labeling them, and then inviting the user to change the variable and observe the result.

**Discard Bad Algorithms Early.** There is a good chance that you won’t get the algorithm exactly right on the first try. Avoid throwing good time and energy after bad by recognizing when the best thing to do is just start over.

### SUMMARY

Expert judgment is an important part of forecasting, but it can also be an area of substantial risk to the forecasting process. Good judgment is needed to build solid forecast models, to document information that is not reflected in quantitative models, and to compensate for problems in the historical data set. Judgment, however, is subject to a number of potential weaknesses such as a lack of replicability and transparency, dubious accuracy, and vulnerability to a number of cognitive biases. In cases where expert judgment is an appropriate method of forecasting, this article has identified a number of strategies for mitigating the weaknesses of judgmental forecasts, including documenting forecasts, not relying on a single forecaster, understanding cognitive biases and taking steps to mitigate them, and using some type of algorithm.

### Notes


8. Lawrence, et al., “Judgmental Forecasting.”

9. Harvey, *Principles of Forecasting*.


14. Table adapted from Spyros Makridakis, et al., *Forecasting Methods and Applications*, 3rd ed. (Hoboken, New Jersey: John Wiley and Sons, 1998). GFOA then reviewed Makridakis’ list with a group of public managers and against other sources to select the biases to highlight in this article.


16. Tetlock originally borrowed the basic construct from Isaiah Berlin.

17. Silver is best known for his high level of accuracy in predicting the outcomes of elections. For example, in 2012, he correctly predicted the presidential winner of all 50 states and the District of Columbia. Silver’s predictions of U.S. Senate races in 2012 were correct in 31 of 33 states.

18. Adapted from Silver, *The Signal and the Noise*.


22. These modeling principles are derived from the work of Frits Willem Vaandrager, Principal Investigator, Institute for Computing and Information Sciences, Department of Model-Based System Development, Radboud University Nijmegen, Netherlands.

---

**SHAYNE KAVANAGH** is senior manager of research in the GFOA’s Research and Consulting Center, in Chicago, Illinois. He can be reached at skavanagh@gfoa.org. **DANIEL WILLIAMS** is an associate professor at the Baruch College School of Public Affairs, where he’s taught since 1995, after nearly 20 years with the Virginia Medicaid program, at the state and local levels. Before that, Williams was the budget director for the Department of Medical Assistance Services. He can be contacted at Daniel.Williams@Baruch.CUNY.edu.