HOW TO INTEGRATE MANAGEMENT JUDGMENT WITH STATISTICAL FORECASTS by Paul Goodwin

Preview: Many of us make judgmental adjustments to statistical forecasts. But do these improve accuracy? Paul Goodwin explains when you should avoid the temptation to adjust, and shows how the accuracy of your interventions can be improved.

Paul Goodwin is Senior Lecturer in Management Science at the University of Bath in England. Paul's three books and more than 20 journal articles give wide attention to the role of management judgment in forecasting and decision making. He is currently conducting a $300,000 UK government-funded study (with colleagues at other universities) of how supply chain companies use forecasting software and how the design of such software might be enhanced. Paul is an Associate Editor of the International Journal of Forecasting and a Director of the IIF.

- Only adjust statistical forecasts when you have important extra information about forthcoming events.
- Record reasons for your judgmental adjustments.
- Break down the adjustment task into smaller parts if you can.
- Consider using the Delphi method if a group of people is involved in adjusting forecasts.
- Consider averaging your judgmental forecasts with those obtained from a statistical method.

Managers are sometimes apologetic about basing their forecasts on judgment. "I'm afraid that our forecasting is very unsophisticated. We rely heavily on people's judgment." is a typical admission. Perhaps they think that, in this hi-tech world, forecasting should also always be hi-tech. Modern forecasters, it seems, should be relying exclusively on advanced statistical methods and intensive computer algorithms.

Thirty years ago, most researchers would have agreed with this perception. The Nobel laureate Daniel Kahneman and his colleague Amos Tversky had found that people were severely limited in the amount of information that they could handle simultaneously so they employed simplistic mental strategies, or heuristics, to get around the problem. When applied to forecasting and decisions, these heuristics apparently led to such inaccurate judgments that the psychologist Paul Slovic concluded that, "In the face of uncertainty, man may be an intellectual cripple."

More recent research paints a more optimistic picture. Researchers, such as Stephen McNees (McNees, 1990) and Michael Donihue (Donihue, 1993), have found that even technically advanced macroeconomic forecasts tend to be improved by judgmental adjustments that take into account such foreseeable events as the introduction of government incentive programs. Other researchers have found that judgment can bring similar benefits to forecasting in individual companies, as long as it is applied carefully (Sanders and Ritzman, 1992; Mathews and Diamantopoulos, 1990). Moreover, in a series of psychological experiments, Gerd Gigerenzer and his colleagues at the Max Plank Institute in Berlin have shown that forecasts based on very simple judgmental heuristics can outperform those derived from advanced statistical methods under appropriate conditions (Gigerenzer et al., 1999). Nowadays, the question is not should management judgment be used in forecasting, but when should it be used and how can its effectiveness be improved?

Judgment and statistical methods complement each other

One key finding of research is that management judgment and statistical methods have complementary strengths when they are used in forecasting (Blattberg and Hoch, 1990). Unlike judgmental forecasters, statistical methods struggle when past data are scarce. So they have difficulties in handling special events or changes in the environment, such as promotion campaigns or new government policies. Complex statistical methods may also lack transparency, and hence credibility, and deny the manager a sense of owning the forecasts. On the other hand, they can make optimal use of vast quantities of data and handle these data consistently.

Management judgment can bring benefits to forecasts…

- It can take into account special events:
  e.g. - a strike
  - a new type of promotion campaign
  - a government initiative

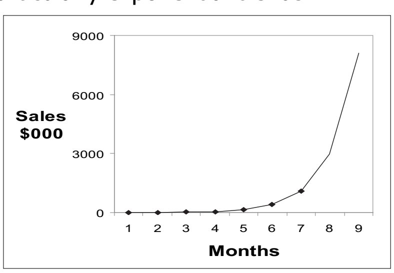
- It can compensate for deficiencies in, or the absence of, historical data.

Despite their potential value, judgmental forecasts are not consistent, nor do they make optimal use of information. Give people the same information on different occasions and they will arrive at different judgments. Judgmental forecasts are also overinfluenced by recent or easily recalled events (Tversky and Kahneman, 1974). For example, rare events – such as an explosion at a chemical plant – that the media have headlined are judged to be more probable than they really are. In addition, people tend to see false systematic patterns in what are really random movements in graphs of past data. They then incorporate these patterns into their forecasts (Harvey, 1995).

Anchoring is another characteristic of judgmental forecasters: when estimating a forecast they often use a starting value, or anchor (for example, last week's sales), and adjust from this to get their forecast. The problem is that the adjustment is usually too small, so the final forecast is too close to the anchor. This may explain the widely observed tendency of judgmental forecasters to underestimate upward trends (for example, sales growth) (Lawrence and Makridakis, 1989). When growth is exponential, the size of this underestimation can be substantial.

Effect on forecasts of anchoring

People underestimate upward trends...
particularly exponential trends.



Even worse things can happen when groups of managers meet to agree on judgmental forecasts. Solomon Asch's experiment in the 1950s showed how group pressures can distort people's judgments. Individuals were even prepared to agree that lines displayed on a screen were of equal length, when this was manifestly untrue, because the rest of the group (who were really actors planted by the experimenter) had claimed this to be the case. In sales or forecasting meetings, domineering individuals or those with high status may seek to impose their judgments on the group, even when they are less accurate or less well founded, than those of other members.

Worse still, despite all of these problems, people generally appear to have unwarranted confidence in the accuracy of their judgmental forecasts relative to that of statistical methods. This confidence persists even in the face of disconfirming evidence. In one experiment (Lim and O'Connor, 1995), people continued to rely on judgment despite receiving such messages as, "Please be aware that you are 18.1 percent LESS ACCURATE than the statistical forecast provided to you."

All of this presents a challenge. How can we draw on the complementary strengths of judgment and statistical methods in forecasting while discarding their weaknesses? The two main approaches to integrating the two types of forecast are (1) judgmental adjustment of statistical forecasts and (2) mechanical integration.

When to adjust statistical forecasts

People tend to over adjust statistical forecasts. Because they tend to read false patterns in the randomness associated with past data, they adjust statistical forecasts, which they think have missed these patterns. All the evidence suggests that, when it comes to identifying systematic underlying patterns in data, there is no contest: statistical methods are superior (O'Connor et al., 1993). These judgmental interventions therefore reduce forecast accuracy and waste the forecaster's time and effort.

Managers should be sparing with their adjustments and should apply them only when they have important information about events that was not available to the statistical forecast. A major promotion campaign, delisting of your product by a large chain store, competitors' actions, a big sporting event on television all may justify adjustments. A 10 percent increase in sales over sales at this time last year probably doesn't. Unless you have good information that the increase is a harbinger of a fundamental change, let the statistical method handle it.

Besides restraint, several actions might improve the quality of judgmental adjustments. Documenting the reasons for adjustments not only helps colleagues to understand the basis for your forecasts; it also reduces the tendency to make gratuitous interventions (Goodwin, 2000a). It can also enable you to learn. By looking back at past cases in which you changed the statistical forecasts, you can assess whether your reasons for applying judgment were justified and hence whether future interventions in similar circumstances are likely to be worthwhile.

Record reasons for your adjustments

- Research suggests that this reduces gratuitous adjustments.
- It allows other people to understand the basis of your forecast.
- It enables you to learn why some adjustments improved forecasts while others did not.

You can make determining the size of adjustments easier if you break the estimation task down into smaller parts. For example, suppose that next month one of your customers is running a buy-one-get-one-free promotion campaign for your product, while another has just finished a campaign and is expecting a postpromotion dip in sales. A third customer has decided to cease stocking the product. Simultaneously estimating the total effect of these events in your head is likely to be difficult and will probably lead to inaccuracies. Instead, it is a good idea to estimate each effect separately and then aggregate the results (Webby et al., in press).

But how should you set about estimating each separate effect? One possibility is to look at analogous events that have occurred in the past and examine the effects they had. Research on this approach is at an early stage, but the idea is that people could employ software that would be designed to search a database for the most similar events (for example, different customer but same time of year, same promotion duration and same type of promotion campaign) and to report on their effects. This might help people to avoid the problems associated with their ability to recall events accurately and their bias towards focusing exclusively on recent events.

What if a group of people have to decide what adjustments to apply to a set of statistical forecasts? One way of avoiding the biases that occur in meetings is to use the Delphi method. Developed by the Rand Corporation in the 1950s, the Delphi method requires participants to provide their forecasts (or adjustments) privately and anonymously to a coordinator, typically via questionnaires. An assessment is then made as to whether there is a consensus, in which case the median estimate is typically used for the forecast. In the absence of a consensus, statistics are used to summarize the group's views, and the results are fed back to the participants, who are then asked to reconsider their estimates. The process is repeated until the participants reach a consensus or few people are changing their views.

The method has many variations; some allow the circulation of anonymous written discussion between rounds. However, while Delphi avoids the biases of face-to-face discussion, this advantage comes at a price. In a meeting, participants can bring new information to the group's attention or they can test important arguments in debate. The scope for providing this level of information exchange in Delphi is severely restricted. Nevertheless, many researchers report successful applications (Rowe and Wright, 1996).

Mechanical integration of judgment and statistical methods

Mechanical integration offers an alternative to the judgmental adjustment of statistical forecasts. It involves using a formula to combine managers' judgmental forecasts with the estimates of statistical methods. Often the formula used is simple. For example, the final forecast might be a straightforward average of the judgmental and statistical forecasts. In fact, this simple averaging approach can be surprisingly effective in achieving a level of accuracy that is superior to that of the individual forecasts (Armstrong, 2001). It works best when the errors of the two forecasts are negatively correlated so that when one method tends to forecast too high, the other tends to forecast too low, and vice versa. People sometimes use more sophisticated combination methods. For example, they might weight the forecasts in the averaging process according to their past accuracy. However, a large amount of past data is needed to estimate these weights reliably, and in a fast-changing business world, the relevant data are often limited.

An alternative mechanical approach is to correct judgmental forecasts for bias. Suppose that a judgmental forecaster tends to underestimate the sales of a product so that his average forecast error is 3,000 units. If the forecaster continues to exhibit this bias, then adding 3,000 units to his forecasts will improve their accuracy. The late Henri Theil proposed a more sophisticated method for correcting forecasts. He suggested first obtaining past data on a person's forecasts and the actual outcomes that occurred and then fitting a regression line to these data. For example, we might find that the line of best fit is Actual Sales = 8 + 0.9 x Judgmental Forecast. This implies that, when the person makes a judgmental forecast, we should multiply it by 0.9 and then add eight units to improve its accuracy. Thus we would transform a judgmental forecast of 200 units into a corrected forecast of 188 units.

Theil's Correction

Regression model:

Actual sales = 8 + 0.9 Judgmental Forecast

so:

Corrected forecast = 8 + 0.9 Judgmental Forecast

If, for example, the judgmental forecast is 200, then the

Corrected forecast = 8 + 0.9 (200) = 188

But be careful. Biases can change over time.

Theil designed his method to remove two types of bias from forecasts, a tendency to forecast too high or too low and a more subtle type of bias, which he called regression bias. Regression bias occurs when the forecasts systematically fail to track the pattern of the variable being forecast. For example, it would be present if a person tended to forecast too high when the actual sales were low, but too low when the actual sales were high, or vice versa. Of course, the method assumes that biases that have been observed in the past are still prevalent; otherwise we might be correcting for problems that no longer exist. Nevertheless, I have found that Theil's correction can be effective in improving forecasts in a wide range of situations (Goodwin, 2000b).

Mechanical integration methods are likely to be most effective when employed by forecast users, rather than by judgmental forecasters themselves. Otherwise, if a judgmental forecaster uses averaging and is aware of the statistical forecast before forming a judgment, he or she might anchor on the statistical forecast. If the forecaster uses correction, he or she might try to preempt the changes that will be applied to the forecast or put limited effort into the task knowing that correction will be applied anyway. However, if these problems can be avoided, the advantage of mechanical methods of integration is that the final forecast is obtained through an objective process that cannot be readily manipulated for political purposes. These methods also reduce some of the inconsistencies inherent in judgment.

Conclusions

Management judgment can play a valuable role in enhancing statistical forecasts. But, given the biases and inconsistencies associated with judgment, it needs to be applied with care and discipline. In particular, judgmental interventions should be made only when the forecaster has important new information that is not available to the statistical method. Moreover, the underlying rationale for such interventions should be recorded, and if possible, the estimation task should be structured by breaking it down into smaller, easier tasks. Special care should be taken when judgments are obtained from groups of people. The employment of methods that control the dynamics of group interaction, such as Delphi, should be considered. Users of forecasts, under the right conditions, can apply such mechanical integration methods as averaging and correction, to improve accuracy.

Follow this advice and you'll have no reason to apologize for using judgment in your forecasts. You'll be making the best use of a crucial resource.

References

Armstrong, J.S. (2001). Combining forecasts. In: J.S. Armstrong, (Ed.), *Principles of Forecasting*. Norwell, MA: Kluwer Academic Publishers, pp. 417-439.

Blattberg, R.C. & Hoch, S.J. (1990). Database models and managerial intuition: 50% model + 50% manager, *Management Science*, 36, 887-899.

Donihue, M.R. (1993). Evaluating the role judgment plays in forecast accuracy, *Journal of Forecasting*, 12, 81-92.

Gigerenzer, G., Todd, P.M. & the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*, New York: Oxford University Press.

Goodwin, P. (2000a). Improving the voluntary integration of statistical forecasts and judgment, *International Journal of Forecasting*, 16, 85-99.

Goodwin, P. (2000b). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods, *International Journal of Forecasting*, 16, 261-275.

Harvey, N. (1995). Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes*, 63, 247-263.

Lawrence, M.J. & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals, *Organisational Behavior and Human Decision Processes*, 42, 172-187.

Lim, J.S. & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: Its effectiveness and biases, *Journal of Behavioral Decision Making*, 8, 149-168.

McNees, S.K. (1990). The role of judgment in macroeconomic forecasting accuracy, *International Journal of Forecasting*, 6, 287-299.

Mathews, B.P. & Diamantopoulos, A. (1990). Judgemental revision of sales forecasts: Effectiveness of forecast selection. *Journal of Forecasting*, 9, 407-415.

O'Connor, M., Remus W. & Griggs, K. (1993). Judgemental forecasting in times of change. *International Journal of Forecasting*, 9, 163-172.

Rowe, G. & Wright, G. (1996). The impact of task characteristics on the performance of structured group forecasting techniques, *International Journal of Forecasting*, 12, 73-89.

Sanders, N.R. & Ritzman, L.P. (1992). The need for contextual and technical knowledge in judgmental forecasting, *Journal of Behavioral Decision Making*, 5, 39-52.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases, *Science*, 185, 1124-1131.

Webby, R., O'Connor, M. & Edmundson, B. (In Press). Forecasting support systems for the incorporation of event information: An empirical investigation, *International Journal of Forecasting*.

Contact Info:
Paul Goodwin
The Management School
University of Bath
mnspg@management.bath.ac.uk