FORESIGHT

THE INTERNATIONAL JOURNAL OF APPLIED FORECASTING

ARTICLES *by* FORECASTERS *for* FORECASTERS

GOOD AND BAD JUDGMENT IN FORECASTING
LESSONS FROM FOUR COMPANIES

Robert Fildes and Paul Goodwin

PREVIEW

In their ongoing investigation into corporate forecasting practices, Robert Fildes and Paul Goodwin have uncovered evidence of excessive use of judgmental adjustment to statistical forecasts. In this report, they document the extent of the problem within four large companies, explore the motivations that lead business forecasters to this sometimes counter-productive behavior, and offer a series of recommendations to ensure that forecast adjustments are made for the right reasons.

INTRODUCTION

If you are a forecaster in a supply chain company, you probably spend a lot of your working life adjusting the statistical demand forecasts that roll down your computer screen. Like most forecasters, your aim is to improve accuracy. Perhaps your gut feeling is that a statistical forecast just doesn't look right. Or maybe you have good reason to make an adjustment because a product is being promoted next month and you know that the statistical forecast has taken no account of this.

But if you are spending hours trying to explain the latest twist in every sales graph or agonizing over the possible impact of Wal-Mart's forthcoming price cut, is this time well spent? Would it make any difference to forecast accuracy if you halved the number of adjustments you made and spent your newly found free time chatting with colleagues at the water

Robert Fildes is Director of the Lancaster University Centre for Forecasting. He was a founding director of the International Institute of Forecasters and an editor of the *International Journal of Forecasting*. Starting with a PhD in Statistics, he's slowly learned that forecasting is an exciting blend of the technical, the organizational, and the psychological. The work reported here combines all three of these elements. But the results demonstrate how far there is to go in forecasting research to improve organizational forecasts.

Paul Goodwin is Professor of Management Science at the University of Bath in England. He has carried out forecasting projects for a wide range of organizations, including a regional electricity company and the U.K. Department of Health. Paul is Foresight's *Research* Column Editor.

cooler about the Broncos' latest signing, Wayne Rooney's soccer injury, or the best beaches in the Caribbean?

To answer this question, we have carried out an in-depth study of four British-based supply chain companies:

- A nationwide retailer
- A leading international food company
- A subsidiary of a U.S. pharmaceutical company
- A manufacturer of own-label domestic cleaning products.

We collected data on over 60,000 forecasts, interviewed the companies' forecasters, and observed forecast review meetings where managers discussed and approved any adjustments that they thought were necessary. The results allowed us to identify which types of adjustments tend to improve accuracy substantially, which make the forecasts worse, and which make little difference, but simply waste management time. We supplemented this data with survey evidence of 149 (mostly U.S.) forecasters.

KEY POINTS

- From an examination of more than 60,000 forecasts in four supply chain companies, we found that making judgmental adjustments to statistical forecasts is not only a popular activity (75% of the forecasts were adjusted) but one that is far *too* popular for the good of the companies.

- While the forecasters usually felt that they had good justifications for making adjustments, we found them overly confident that their adjustments would improve forecast accuracy.

- Large adjustments did tend to be beneficial, but small adjustments did not materially improve forecast accuracy and sometimes made accuracy worse. And negative (downward) adjustments were more likely to improve forecast accuracy than positive adjustments.

- To make a large adjustment takes some nerve; as a result the larger adjustments are likely to be made only for very good reasons. These are the adjustments that are potentially worth making.

- Over optimism leads to erroneous positive adjustments, while negative adjustments are based on more realistic expectations.

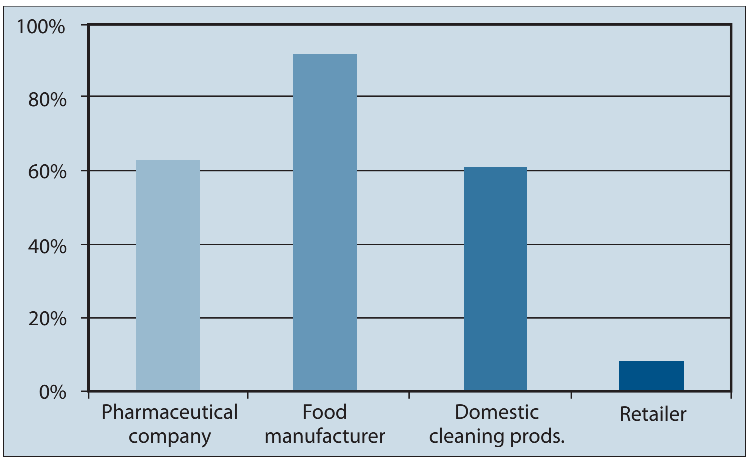
- We also found a bias toward "recency"; that is emphasizing the most recent history while treating the more distant past as bunk. Doing so can undermine the process of statistical forecasting.

ADJUSTMENTS GALORE

Adjusting forecasts is certainly a popular activity in all our companies, as shown in Figure 1. In fact, the forecasters spend so much time making adjustments that they are probably making a significant contribution to world demand for headache tablets.

Those working for the food manufacturer adjusted 91% of the forecasts that had been generated by their expensive and sophisticated forecasting software. The four forecasters employed by the retailer adjusted only about 8% of their forecasts, but then they had over 26,000 forecasts to make each week, so there probably wasn't enough time to put their mark on each forecast. The pharmaceutical company held 17 forecast review meetings every month, tying up about 80 person hours of valuable management time. On average 75% of the statistical forecasts in our companies were adjusted. Our survey of forecasters (Fildes & Goodwin, 2007) tells much the same story, with just 25% of the forecasts based only on a statistical method. Judgment, either used exclusively (25%) or combined with a statistical forecast (50%), was regarded as important or very important by most of the respondents.

Figure 1. Percentage of Company Forecasts That Are Adjusted



What sort of adjustments did the forecasters make? Many of the adjustments were small, and in some cases very small. It was as if forecasters sometimes simply wanted to put their calling card on forecasts by tweaking them slightly to show that they were still doing their job. Indeed, we received anecdotal evidence from a consultant that people at review meetings tend to adjust more of the forecasts that are presented earlier in the meetings, rather than later on. As the meeting progresses they tire and feel that they have already done enough to justify the meeting, so later forecasts are simply waved through.

Tip: Review the important A-class products first.

Of course, showing that they were still alive was not the only reason the forecasters made adjustments. They usually felt that they had good justifications for making them and we found that often this was the case. The problem is that people have a tendency to find a ready explanation for every movement in the sales graphs, including those swings which are really random. And this makes them overconfident that their adjustments will increase accuracy.

"Our customers were stocking up two months ago because they were anticipating a price increase so our sales swung upwards."

"OK, they didn't stock up in the previous year when they knew there was going to be a price increase because interest rates were high and there was a lot of uncertainty about."

We are brilliant at inventing theories for everything we observe. Scott Armstrong (1985, p.54) discusses a case where a Nobel laureate published a hypothesis to explain an oddity in the graph of a macroeconomic variable. Later it was shown that the anomaly was the result of an arithmetic error. At 13.01 on a December day in 2003 after Saddam Hussein had been captured, the price of U.S. Treasuries rose. Half an hour later the price fell. Taleb (2007, p.74) reports that the Bloomberg News channel used the capture of Saddam to explain both price movements. The unfortunate, dull statistical forecast can offer no competition to these colorful, but often groundless, tales and so it gets adjusted.

THE ILLUSION OF CONTROL

All this adjustment behavior can have some odd consequences, according to psychologists. When we engage in activities that involve skill and effort, we normally believe that we have more control over what we are doing. For example, if you develop your skills and invest effort in learning to play a musical instrument, you will make fewer mistakes. The same applies to controlling the ball in a sport like football. But many of the swings in a sales graph are beyond the forecaster's control. They are the result of random, unpredictable events. Yet, because forecasters see making adjustment as a skillful activity, they can develop the false belief that they have control over the demand that they are trying to forecast and hence that they can predict the movements in the sales graph. The phenomenon is called the illusion of control. It's likely to motivate you to make even more adjustments. After all, the more you adjust, the more control you think you have.

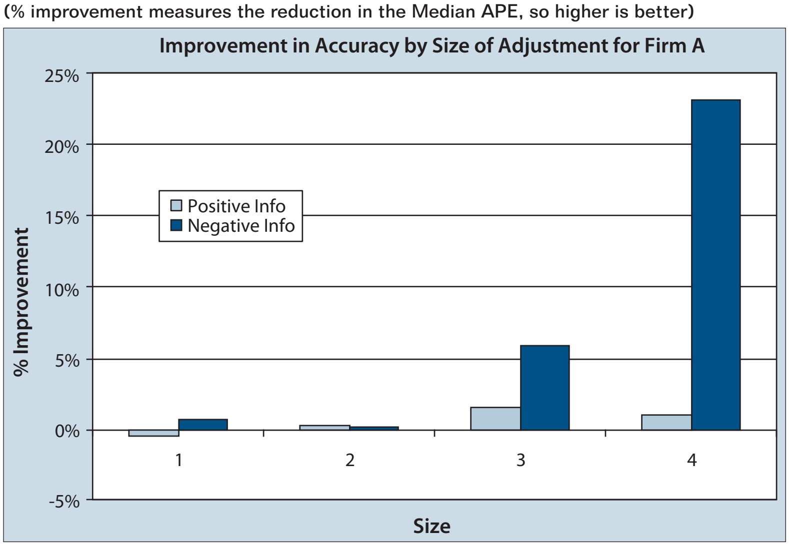
Tip: Don't adjust without an explicit, recorded reason.

WHEN DO ADJUSTMENTS IMPROVE ACCURACY AND WHEN DO THEY NOT?

Despite these concerns, judgmental adjustments to statistical forecasts can still play a useful role in improving accuracy. Our study found that on average they lowered the average percentage error (MAPE) by 3.6 percentage points for all companies except the retailer. But this modest improvement masks considerable variation in the effectiveness of the adjustments. Is it possible to filter out the type of adjustments that are likely to be useless or even damaging to accuracy?

We first examined how the *direction of adjustments* affected forecast accuracy. We contrasted adjustments that increased the forecast (positive adjustments) with those that lowered it (negative adjustments). For one of our four companies, Figure 2 shows the extent to which these adjustments led to improvements. The results are typical of our three non-retail companies.

Figure 2. Effect of Adjustments by Size and Direction
(% improvement measures the reduction in the Median APE, so higher is better)



The graph breaks the size of the adjustments into quartiles: Quartile 25% represents the smallest 25% of the adjustments while Quartile 100% represents the largest quarter of the adjustments. Two results are immediately apparent: 1) larger adjustments tend to improve accuracy and 2) negative adjustments tend to be much more beneficial than positive.

Why are larger adjustments more likely to improve accuracy? To make a large adjustment takes some nerve. Senior managers may be on to you if you make a big adjustment and then things go badly wrong. This means that the larger adjustments are likely to be made for very good reasons. You are likely to have reliable information about some important future events that will cause the statistical forecast to have a large error. In contrast, the smaller adjustments are the tweaks that we mentioned earlier or the result of a forecaster hedging his or her bets because information about a future event is unreliable. The lesson is clear: while small adjustments, by definition, can do relatively little harm to accuracy, they are generally a waste of time. Doodling in your notepad is likely to be more productive and certainly more therapeutic.

Why do the positive adjustments fare so much worse than the negative? Psychologists tell us that people have an innate bias towards optimism. For example, most construction projects usually take longer to complete and cost far more than was originally predicted. Some of this may be a result of deliberate misrepresentation (see Flyvbjerg et al., 2005) to gain contracts, but there is evidence that optimism bias still plays a significant role in these poor estimates. It seems, therefore, that when our company forecasters are asked to estimate the effects of a sales promotion campaign or a price reduction, they cannot resist being overly optimistic. And of course this reflects the enthusiasm of their colleagues in sales or marketing. In contrast, when they make a negative adjustment they are much more realistic in their expectations.

A particularly damaging intervention is called a *wrong-sided adjustment*. For example, this occurs when you adjust the forecast upwards but should have made a negative adjustment. Suppose that the statistical forecast was for 600 units and you adjusted upwards to make a forecast of 650 units. When actual sales turn out to be 580, you'll realize that your adjustment was in the wrong direction. Any wrong-sided adjustment is bound to reduce accuracy. Yet surprisingly, our companies made a large number of these adjustments, particularly when the direction of adjustment was positive. More than a third of the positive adjustments made by the non-retailers were in the wrong direction. If we could remove even 50% of the wrong-sided positive adjustments, accuracy would be improved by 7 percentage points. For negative adjustments the effects were much more limited.

> ***Tip: Beware the enthusiasm of your marketing and sales colleagues!***

We investigated whether the wrong-sided adjustments might be a result of misjudging the timing of promotion effects (e.g. expecting an immediate uplift in sales when the actual increase is delayed) but found no evidence of this. Once again, misplaced optimism seems to be the most likely explanation.

But how can forecasters make fewer wrong-direction mistakes? We've explored some possible solutions. We believe that the first stage in eliminating wrong-sided adjustments is to catalogue the reasons behind each and every adjustment. In our survey 69% of companies claimed to do this. But of the companies we observed, none collected this information effectively. Second, when large errors have occurred, a post-mortem on the reasons has the potential to help the next time similar incidents threaten. And this should be done as part of a *forecast quality improvement program* rather than in an atmosphere of blame. An effective *forecasting support system* can help by encouraging the compilation of the historical record to make life easy for the forecaster to look back at past events (such as promotions) and to reflect on how today's circumstances match with the past record. In our research we showed just how this can be done through the design of effective software that lets the forecaster examine the past record of similar analogous promotions (Lee et al., 2007).

> ***Tip: Collect information on key drivers. learn from large errors.***

THE IMPORTANCE OF DEFINITIONS

So far we have not mentioned the retailer. When we analyzed the accuracy of the retailer's adjustments they looked awful. The positive adjustments its forecasters made more than doubled the MAPE from 32% to 65%. Moreover 83% of these adjustments were either too large or in the wrong direction. Something odd was going on. Why would the forecasters of a major national company be spending so much time and effort making mediocre statistical forecasts so much worse?

Most people would probably consider a forecast to be an estimate of the most likely level of future demand. It turned out that the retail forecasters were estimating a different quantity. Often they were trying to determine the levels of demand that only had a small chance of being exceeded – that is, the level that would limit stock-outs. Determining this level would tell them how much inventory they needed to hold. For example, their statistical forecasting system might provide a demand forecast of 500 units but they would adjust this upwards to 550 units, reasoning that this level of inventory would be sufficient to cover anything but the most extreme level of demand. In an informal way they were forecasting fractiles, as discussed by Goodwin in the Hot New Research Column in *Foresight*, Summer 2007. So our MAPEs were unfairly measuring the effectiveness of the forecasters' adjustment because they were not trying to predict the actual demand.

However, there were still serious problems with the retail forecasters' approach. First, they had never clearly defined what they were forecasting. They simply referred to their adjusted figures as "forecasts," posing the obvious danger that other managers would wrongly interpret these as estimates of the most likely level of demand and then make decisions based on this assumption. Second, their approach was informal. They had never determined what probability of a stock-out was appropriate in order to balance inventory holding costs against the costs of disappointing customers (see Catt, 2007). Nor had they done any analysis to see whether their adjustments were leading to over- or under-stocking for the many products they sold.

Finally, the forecasters were trying to do two jobs at once. They were adjusting the statistical forecasts for special events like promotions and, at the same time, adjusting them to estimate inventory requirements. They may have been taking on too much. The evidence from psychologists is that humans have limited information-processing capacity and that better judgments can be obtained by breaking judgmental estimation down into simpler and easier tasks – a process called *decomposition*.

> ➩ ***Tip: Make your forecast your estimate of most-likely future demand. Then adjust to account for the relative costs of under- and over-forecasting.***

HISTORY IS NOT BUNK

Henry Ford is alleged to have said that history is more or less bunk. Many of the forecasters in our companies had the same philosophy. In review meetings they examined most recent movements in sales graphs with forensic intensity while they often ignored earlier data. In one company the forecasters told us that they never fit their statistical methods to demand data that is more than three years old because "back then the trends were different." Sometimes the software they had bought seemed to share the same attitude – the active data base only went back three years!

There was no evidence that they had tested this claim. So great was the bias towards recency that sometimes statistical methods were only fitted to the last six months' data. This did not give these methods much of a chance. As Rob Hyndman and Andrey Kostenko wrote in the Spring 2007 issue of *Foresight,* statistical methods can require quite lengthy periods of data to detect underlying patterns, even when the demand data is well behaved and the level of randomness in the series is relatively low. Moreover, the methods commonly found in business forecasting software are designed so they can adapt to changes in trends or seasonal patterns if these occur. If you restrict the data available to your statistical methods, then you are unlikely to be making judgmental adjustments from a reliable baseline.

> ➩ ***Tip: Use all the data you can lay your hands on. Discard data only with good reason.***

CONCLUSIONS

Judgmental adjustment of statistical forecasts is a crucial part of the forecasting process in most companies. It is often not practical to use statistical methods to model the effect of forthcoming events that you know are likely to have a big impact on demand. Management judgment then has to step in to bridge this gap and, if applied correctly, it can bring great benefits to forecasts. However, our study has shown that these potential benefits are largely negated by excessive intervention and overoptimism. Indeed, had our non-retail forecasters been banned from making positive adjustments to their forecasts, but still been allowed to make negative adjustments, their judgmental adjustments would have improved the MAPE by over 20 percentage points, rather than the mediocre 3.6 points that we reported earlier.

In most companies, however, banning all positive adjustments would not be a realistic strategy. The answer is to make these adjustments with more care and only on the basis of better market information. In the long run, software enhancements might be helpful here.

Our study also emphasizes the importance of having a clear definition of what you are forecasting. It's not good for morale when a colleague complains you've over-forecasted demand by 40% when that's not what you were trying to predict.

Finally, we leave you with these recommendations on your adjustment policy.

- Accept that many of the movements in your sales graph are random. You have no control over them and they cannot be predicted.
- Small adjustments are likely to waste time and effort and may damage accuracy.
- Positive adjustments (moving the statistical forecast upwards) should only be made with care. Be especially cautious about being too optimistic.
- Give statistical forecasting methods a chance: they need plenty of data to detect underlying patterns in demand.
- Define clearly what you are forecasting.

[Ed. Note: For a good summary of the research on judgmental adjustments, see the special feature section in *Foresight,* Issue 1 (June 2005), "When and How Should Statistical Forecasts Be Judgmentally Adjusted?"]

REFERENCES

Armstrong, J.S. (1985). *Long-Range Forecasting* (2nd ed.), New York: Wiley.

Catt, P.M. (2007). Assessing the cost of forecast error: A practical example, *Foresight: The International Journal of Applied Forecasting*, Issue 7, 5-10.

Fildes, R. & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting, forthcoming in *Interfaces*.

Flyvbjerg, B., Skamris Holm, M.K. & Buhl, S.L. (2005). How (in)accurate are demand forecasts in public works projects? The case of transportation, *Journal of the American Planning Association*, 71(2), 131-146.

Goodwin, P. (2007). Supermarket forecasting – check out three new approaches*, Foresight: The International Journal of Applied Forecasting,* Issue 7, 53-55.

Hyndman, R.J. & Kostenko, A.V. (2007). Minimum sample size requirements for seasonal forecasting models, *Foresight: The International Journal of Applied Forecasting*, Issue 6, 12-15.

Lee, W.Y., Goodwin, P., Fildes, R., Nikolopoulos, K. & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks, *International Journal of Forecasting,* 23, 377-390.

Taleb, N.N. (2007). *The Black Swan*, London: Allen Lane.

ACKNOWLEDGEMENTS: This research was supported by Engineering and Physical Sciences Research Council (EPSRC) grants GR/60198/01 and GR/60181/01. Michael Lawrence, Kostantinos Nikolopoulos, and Alastair Robertson contributed substantially to the data analysis.

CONTACT (*corresponding author)
Robert Fildes
Lancaster University
R.Fildes@lancaster.ac.uk

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