THINKING ABOUT COMPOSITE FORECASTING

By Larry Lapide, Ph.D.

(This is an ongoing column in The Journal, which is intended to give a brief view on a potential topic of interest to practitioners of business forecasting. Suggestions on topics that you would like to see covered should be sent via email to llapide@mit.edu).

I went to a conference about 30 years ago and attended a session by exponential smoothing guru Robert G. Brown. During the question-and-answer period of his talk, someone asked him what he thought about the composite forecasting topic that was being discussed in another room at the conference. I forget what he actually said, but effectively he was not keen on the idea.

Perhaps colored by his view, I had never been a big fan of composite forecasting. I thought that it was not worth the extra effort in business forecasting where we have to generate myriad forecasts. Who has the time to play around with figuring out the best blending of two or more forecasts into a single forecast? It just seemed easier to pick the best one from a bunch of forecasts and go with that. I believe that this is consistent with the thinking of most business forecasters who have to deal with tactical and operational forecasting to support complex business planning needs.

I just finished advising an MIT graduate student on his thesis this past academic year and his findings convinced me that there might be some gains to be gotten from composite forecasting. His Master’s thesis was sponsored by a high-tech company and he was investigating ways to improve its collaborative demand planning process that leverages qualitative sales forecasts from major customers and involves no statistical forecasting. Since his data analysis was not that extensive, his findings cannot be considered quite definitive. However, he was able to demonstrate that top-down and bottom-up statistical forecasts, as well as qualitative sales forecasts had potential value towards improving the company’s forecasting accuracy—sometimes by themselves, and sometimes when incorporated within a composite of two forecasts. (Ratan Jha’s thesis is referenced below for more detail).

WHAT IS COMPOSITE FORECASTING?

To be perfectly honest about composite forecasting, I have to admit that I did not even know the name for the analysis the student was doing during his research. I just advised him to take a look at a weighted average of two forecasts and see how that fared towards improving forecast accuracy when compared against a single, non-blended forecast method. So as research for this article I did a Google search on “composite forecasting” and got 1,180 hits (surprisingly none from Wikipedia). I found this definition of composite forecasting on an SAP-related site:

“What’s the underlying objective is to take advantage of the strengths of each method to create a single “one number” forecast. By combining the forecasts, the business analyst aims to develop the best forecast possible. The composite forecasts of several methods have been proven to out-perform the individual forecasts of any of those methods used to generate the composite.”

(Source: SAP Community Network, https://www.sdn.sap.com)
Taking this definition, composite forecasting generally aims to take two or more forecasts generated in various ways and find the best weighted average of them taken together in order to get a best forecast. Intuitively the weighting factors should be related to the accuracy of each forecast in the composite, so that more accurate forecasts are weighted more than less accurate ones.

**THESIS METHODOLOGY AND APPROACH**

As a background, this thesis looked at the forecasting for a major customer’s demand at product identification (PID) levels that aggregate up into Product Type levels. For this customer all of the PIDs exhibited extremely volatile weekly demand variations and included many PIDS that exhibited “slow-moving” intermittent demand patterns. This type of demand is generally difficult to forecast, so any potential improvement is worth investigating. The analysis followed a two-stage approach along the lines depicted in Figure 1. The thesis involved assessing whether (at a PID-level) statistical forecast methods were more accurate than qualitative sales forecasts, as well as whether single forecasts are more accurate than composite forecasts. A best forecast was one that had the highest accuracy as measured in terms of Root Mean Square Error (RMSE).

In the first stage of the analysis, PID forecasts and Type-level (TD) forecasts were generated by identifying the best Damped Trend smoothing model for the faster-moving items, and the best moving average forecast for items experiencing intermittent demand; historical demand data was used to ascertain the best. Then Bottom-Up (BU) forecasts of Types were calculated by aggregating the corresponding PID forecasts. A composite statistical forecast (CS) was the best weighted average of BU and TD forecasts. The best weighting involved finding the exponential smoothing factor that minimized errors (i.e., the best \( \alpha \), where \( CS = \alpha \times BU + [1-\alpha] \times TD \)). Using hold-out historical data, comparisons were made to see which one of three forecasts, BU, TD, and CS, was the best, and is called the Best Statistical. Analysis of the results among the PIDs was mixed, meaning that the Best Statistical method was PID-dependent. Given the limited amount of data used, no apparent pattern could be discerned from the analysis. However, and most interestingly, the analysis showed that the composite forecast was best for some PIDs.

During the second stage, another composite forecast, the best weighted average Qualitative-Quantitative forecast (QQ), that incorporated the Best Statistical and the Sales forecasts, was generated by finding the best smoothing factor (i.e., the best \( \alpha \), where \( QQ = \alpha \times \text{Best Statistical} + [1-\alpha] \times \text{Sales} \)) that minimized the RMSE. Again using the hold-out historical data, comparisons were made to see which of three forecasts, Sales, Best Statistical, and QQ, was most accurate. Once again the analysis of the results was mixed and the best forecast was PID-dependent. Importantly, once again the composite forecast was best for some PIDs.

While the data used in the thesis described above were not extensive enough to make a business case for moving forward with the concepts investigated, they demonstrated that there is potential to use composite forecasting to improve...
the ways collaborative sales forecasts are leveraged. The sponsor company will be doing more analysis to further evaluate the concepts for their own use.

LESSONS FOR FORECASTERS

There are a variety of lessons to be learned from the thesis for forecasters who wish to incorporate collaborative information gleaned from customers, as well as for those using top-down and bottom-up forecasting. Here are some:

- Typically when top-down bottom-up forecasting is done, one expects the top-level forecast to have less relative error due to the benefits of compensating errors when aggregating lower-level demand variations. So a forecaster typically assumes the top-level forecast is more accurate and adjusts the lower-levels forecasts so that their sum equals the top-level forecast. Since this thesis had some cases where the BU forecast was more accurate, forecasters should first check to see whether the top-level forecast is more accurate before adjusting the lower level forecasts. In some cases the top-level should be adjusted to equal the BU forecast.

- In some cases the composite statistical forecast (CS) was more accurate than either the BU and TD forecasts taken separately. In these cases the top-level and lower levels forecasts should not be adjusted to equal each other, but rather adjusted to equal the CS forecast instead.

- In a collaborative planning environment, the customer Sales forecast is usually used as the forecast since it comes from the customer, and is assumed to be more accurate than any that could be generated by the supplier. However, results from the thesis show that often enough the Best Statistical forecasts were more accurate than the Sales forecasts. This demonstrates that statistical forecasting can add value in improving accuracy in collaborative planning. As a caveat, however, one cannot assume that just because it is quantitative it will always be more accurate than a qualitative forecast.

In addition to this, since the qualitative-quantitative composite forecast (QQ) was most accurate for some PIDS, this supports the fact that neither statistical nor qualitative forecasting alone can provide the most accurate forecast. Each can be used to provide information that can help forecast future demand more accurately, and thus composite forecasting can help extract the information from each. (An interesting benefit from the use of composite forecasting is that the \( \alpha \) (smoothing factor) tells one how much confidence one has in each forecast. For example, if \( \alpha = 0.3 \) then one has 30% confidence in the statistical forecast and 70% confidence in the Sales forecast. When updating \( \alpha \), any change can be used to gauge changes in the confidence levels of the two types of forecasts.)

These lessons have convinced me that forecasters should look into ways to use collaborative planning and composite forecasting to better draw out value-added information from statistical, as well as qualitative forecasts. To implement the thesis’ approach, forecasters would need to routinely generate six forecasts—the BU, TD, CS, Best Statistical, Sales and QQ forecasts—and keep track of their accuracies relative to each other. They can update the forecasts by following the process depicted in Figure 1. This, of course, means forecasting systems will get more complex, and should only be done if accuracy improvements justify the increased complexity.

In summary, while for thirty years I haven’t been an advocate of composite forecasting, the MIT thesis just completed has convinced me otherwise. Given the computer power we now have at our disposal, this added complexity could be easily automated to improve forecast accuracy. I wonder, however, whether Robert G. Brown’s mind would be changed with this analysis?

REFERENCE