Practical Considerations in Forecast Value Added (FVA) Analysis
HARPAL SINGH, SANJIV RAMAN, AND ERIC WILSON

PREVIEW A business forecast is normally produced by applying overrides and modifications to an initial statistical forecast. Whether these adjustments to the statistical forecast improve forecast performance is always an important question – one to which the Forecast Value Added (FVA) concept can be applied. In this article, Harpal, Sanjiv, and Eric identify some of the practical issues in applying FVA analysis within a demand-management process.

FVA ANALYSIS

In a typical forecasting process, a set of statistical forecasts is generated using historical demand data. These initial predictions are then modified by input from the demand planner, the sales organization, marketing, and management, and finally passed down to the Sales and Operations Planning (S&OP) process.

The question is which of these steps, if any, in developing a final forecast actually improves the result. In his Foresight article ”FVA: A Reality Check on Forecasting Practices” (2013), Michael Gilliland posed the key issue: The typical business forecasting process consumes large amounts of management time – but is it “adding value” by making the forecast more accurate and less biased?

After all, there doesn’t seem to be any point in gathering a lot of data from folks if their inputs don’t improve accuracy.

“Good” and “Bad” Judgments
Numerous articles suggest that management overrides do not systematically improve forecast accuracy. One frequently cited analysis is provided by Robert Fildes and Paul Goodwin in their 2007 article in this journal. They report there is “good” and “bad” judgment, and recommend that overrides be avoided in many common situations.

Gilliland defines FVA as “the change in a forecasting performance metric that can be attributed to a particular step or participant in the forecasting process” (p.15). While the definition is a useful starting point, this article addresses some practical difficulties in the gathering of inputs and FVA evaluation.

Collaboration
The statistical model used to generate the initial forecast tries to capture the systematic behavior of the underlying demand pattern. The assumption is that observed demand is the sum of an underlying pattern that can be approximated by a statistical model and of random variability. Past data are used to identify the parameters of the underlying pattern as well as the extent of the randomness.

The statistical forecast is typically passed on to collaborators with little or no guidance other than to “improve on it if you can” or “check it for reasonableness.” This sets the folks responsible for improving the forecast in direct competition with statistical forecasting. They see their role as correcting the forecast, rather than improving it. A better practice would be to view the statistical forecast as one of the collaborators, not the competition.

The purpose of gathering inputs is not to validate the statistical model or calculations, but rather to include selective information that may be available but not reflected in historical data. The input can either change a forecast number or the underlying statistical model. Making sure that the statistical
Key Points

■ The Forecast Value Added (FVA) concept is designed to determine which if any steps in the forecasting process—particularly those steps that impose judgmental overrides to statistical forecasts—improve forecast accuracy, and which do not.

■ The statistical forecast is typically passed on to collaborators given little or no guidance other than to “improve on it if you can” or “check it for reasonableness.” This sets the folks responsible for improving the forecast in direct competition with statistical forecasting. A better practice would be to view the statistical forecast as one of the collaborators, not the competition.

■ When there are multiple levels of overrides—from sales, marketing, management—averages or weighted averages of these can be taken to create a consensus forecast. FVA analysis can determine if the consensus forecast proves to be better, not just whether the individual inputs improve the forecast. In fact, it is possible that each individual collaborative input can make things worse, while the consensus forecast shows overall improvement.

■ The right question, then, is not whether each of these inputs adds value, but whether each of these inputs can be combined to effectively integrate analytics and planner expertise into a better forecast.

■ An adaptive weighting scheme is suggested for calculating consensus forecasts. In this scheme, the weights assigned to the collaborators change over time to reflect changes in their relative forecasting performance.

SALES AND MANAGEMENT OVERRIDES

Salesperson

The salesperson is often the only individual who has information on imminent changes in future demand, such as a customer plant outage or change in a customer’s planned orders. The salesperson can improve the forecast by making selective adjustments based on information not immediately reflected in history or that represents a departure from past history.

Independent Adjustments

While the statistical forecast is the starting point, these adjustments are captured independently so that their value can be measured over time. Many organizations use tools that directly overwrite the statistical forecast with the salesperson’s best estimate. This is not good practice, since it no longer maintains the integrity of each forecast input independently, and as a result dilutes the efficacy of these inputs in the forecast process.

Importance of Communication

While imminent changes should translate into manual overrides, longer-term shifts in demand need to be communicated to the demand planner (or forecast model builder) so that they can be applied to adjust the underlying statistical model. For example, a sudden but temporary surge in demand due to a weather event represents a departure from the forecast, but should not necessarily modify the underlying statistical model.

Thus it is important to support the process of gathering overrides with communications on the basis and need for them. Fildes and Goodwin recommend that “the first stage...is to catalogue the reasons behind each and every adjustment” (p.8). This is daunting and impractical, however, given that the forecast is manually input over a majority of the entire portfolio. It is more effective to capture the strategically important reasons behind those overrides that exceed a certain threshold of the statistical forecast or that pertain to key customers or product families. Mandating that these key reasons be provided for these overrides can ensure that the forecast inputs are given due consideration and also serve to deter frivolous inputs.
The old maxim “Bad news is best delivered early” holds true in the forecasting process as well. Fildes and Goodwin suggest that negative adjustments to the statistical forecast tend to be more effective than positive input in improving the forecast accuracy. This result may be attributed to the tendency of loss aversion among human beings, suggesting that incentives need to be provided that align with the goal of improving forecasts rather than just increasing sales.

**Management of Adjustments**

Adjustments to a statistical forecast are not all provided at the same aggregation level. Those from individual salespersons will normally be at the product/customer level of detail, while sales management generally provides input at an aggregated regional level. In some organizations, this is complicated by customers that span different geographical regions so that sales managers with key customers may have responsibilities that intersect with regional managers.

**Figure 1. Sales Manager Overrides**

<table>
<thead>
<tr>
<th></th>
<th>Product A</th>
<th>Product B</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Forecast</td>
<td>40</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Salesperson Override</td>
<td>50</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Sales Manager</td>
<td>50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the example of **Figure 1**, the initial statistical forecasts are 40 and 60 for Product A and Product B, respectively. The salesperson changes only Product A – from 40 to 50 – but the manager reduces the combined forecast to 90.

**Reconciliation Options**

There are at least a couple of ways that the combined forecast can be rationalized:

1. Attempt to reconcile the twin overrides. Assuming that the salesperson has some critical insight, the final forecast for Product A could be 50, so the forecast of Product B could be changed to 40 to accommodate the management override.

2. Apply the overrides in sequence. Initially change the forecast to 50 for Product A while keeping 60 for Product B, but then adjust both downward to sum to 90.

The point is that it is not possible to determine if the process involving the collective inputs of the salesperson and the sales manager has added value without first establishing how the transition occurred between the statistical and final forecasts (100 vs. 90). You have to precisely define forecasting process steps (and any combining or reconciliation methods), and each component of the override must be captured independently.

**Moving Up the Ladder**

Marketing decisions including promotions are normally planned well in advance of the marketing initiative. More often than not, the initial forecast presented to the salesperson for adjustment incorporates marketing overrides that were entered months ago. The goal of marketing is to make certain that the impact of new marketing initiatives is reflected in the tactical forecast. In many forecasting engines, this information is not in the historical data because the aberrations caused by previous marketing initiatives are deliberately removed to make the history “forecastable.”

Some management overrides apply to distant periods in an attempt to align strategic business goals with the tactical forecast. They deal more with shaping the overall demand across different segments, industries, and regions. These changes ought to be incorporated into the forecasts presented to the sales organization.

**Figure 2** depicts the typical time sequence of forecast adjustments.

**Figure 2. Time Sequence of Overrides**

**COMBINING FORECASTS**

The typical forecasting process starts with an input forecast based on a statistical model, and results in an output forecast that takes all the forecast adjustments into account. As noted, these adjustments are
provided at different levels of aggregation and for different time horizons. The resulting forecast is forwarded to the next process step, typically the S&OP meetings.

**Input Combinations**

In order to determine if a forecasting process step adds value, it is not sufficient to simply look at one input (the sales override forecast, say). Rather, we should consider intelligent combinations of inputs. These can be different inputs combined or the same input aggregated differently. Each combination can be thought of as a different forecast.

In their paper "Improving Forecast Accuracy by Combination," Feng Zhang and Robin Roundy (2003) proposed an analytical framework that combines multiple forecasts into a single and hopefully more accurate forecast. Their method creates weights for the different forecasts and combines them in an "optimal" way to generate a consensus forecast.

Figures 3a–3c illustrate the combination process they propose.

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Forecast</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>Sales Overrides</td>
<td>45</td>
<td>55</td>
<td>65</td>
<td>75</td>
<td>75</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>Marketing Input</td>
<td>40</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Management Adjustment</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
</tbody>
</table>

If historical analysis indicates that the salesperson’s input in the first period is very reliable, but reliability progressively decreases in later periods, we would assign a weight closer to 1.0 to the sales overrides and closer to 0 to the rest. In the other periods, the weights would be assigned based on how the different forecasts have performed in the past. An example assignment of weights might be:

<table>
<thead>
<tr>
<th>Period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Forecast</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Sales Overrides</td>
<td>1</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Marketing Input</td>
<td>0</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.5</td>
<td>0.4</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Management Adjustment</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>

By multiplying the weights and the individual forecasts, we come up with a consensus forecast:

**Individual Inputs vs. the Consensus**

An ultimate objective of FVA is to determine if the consensus forecast proves to be better, and not just whether the individual inputs improve the forecast.

Determining the appropriate weightings may be a nontrivial problem, particularly for items that do not have long enough histories to make a proper assessment of the performance of the individual inputs. For example, assessing comparative performance based on just one year of data may not be very reliable. On the other hand, simply averaging the various inputs may provide improved results with little extra effort.

By tracking the individual inputs as well as the consensus forecast, it is possible to determine the FVA for each. It is also possible to compare performance of the Zhang and Roundy "optimal" combination method with a simple unweighted average of the inputs.

Evaluating a consensus forecast is important, because it is possible this blended consensus will improve accuracy even if the individual inputs by themselves fail to improve upon the statistical forecast. For example, suppose in a given period the statistical forecast is 100, the marketing input is 122, the sales override is 90, the management adjustment is 130, and the actual turns out to be 110. Each of the individual inputs is worse than the statistical forecast (absolute errors of 12, 20, and 20 compared to just 10 for the statistical forecast), yet a simple average of the three inputs (114) has an absolute error of just 4.

This example illustrates the potential value of a consensus process for combining inputs. Applying FVA simply to the individual inputs would suggest they are all non-value-adding and should be
eliminated from the forecasting process. But their true value to the process may be in that they add additional perspectives to incorporate into the consensus.

**Segmenting the Portfolio**

Another use of FVA is to identify segments of the portfolio – products or customers – where collaboration can improve the forecast accuracy. In some situations, such as with extremely smooth and stable demand, a simple statistical model may forecast just fine, and fiddling with it won’t make it any better. By segmenting out products with these behaviors and leaving them untouched, you only make adjustments where there is good opportunity for improvement.

We have found that, in the near term, collaborative input adds value—a positive FVA over the statistical forecast. By contrast, in the outer periods, the statistical forecast engine does a much better job at forecasting, with collaborative input often making matters worse.

**WEIGHTING FORECAST INPUTS**

**Averaging Inputs**

If the different forecast inputs are independent, they can be blended by using a simple average to improve the forecast accuracy.

However, the different inputs in the forecasting process are typically related or hierarchical, such as salesperson ➔ sales manager ➔ management. Each forecast input is shared with the next group of collaborators as a reference point. In such situations, the biases and assumptions held by one group tend to permeate the forecasting process, so a simple averaging method may no longer be sufficient.

Hence, to blend the forecast inputs to arrive at a consensus forecast, forecasting processes that involve hierarchical inputs may need to employ more sophisticated techniques. The Zhang and Roundy weighting technique provides a useful framework for blending using different weights for different time horizons. Advanced planning systems should support the capability to archive past forecast inputs so that the efficacy of each input can be measured over time.

**Adaptive Weights**

In the example above, we did not focus on the analysis that generated the weights used to combine different forecasts. Zhang and Roundy (2003) compare many schemes that have been proposed, and identify several procedures that produce robust results over a variety of forecasting situations. While the actual combination method is not that critical, they assert it is important that the weights adapt as business conditions and the forecasting process change. For example, the statistical forecast may improve over time so that sales modifications no longer add value.

This notion that forecast inputs can add value over some portions of the horizon and not others is called Time-Phased Forecast Value Added (TPFVA). Over time, the weights that are calculated to combine the forecasts change. If the weights associated with a particular input get too low, that input can be effectively eliminated. Conversely, an improvement in the process of gathering inputs from, say, marketing may result in more accurate marketing numbers. If this accuracy continues, the method of calculating the weights will begin to assign a higher weight to the marketing inputs, effectively recognizing the process change.

In general, we have found that, in the near term, collaborative input adds value—a positive FVA over the statistical forecast. By contrast, in the outer periods, the statistical forecast engine does a much better job at forecasting, with collaborative input often making matters worse.

We also found that the weights would be attribute-dependent; for example, the weights could be different for different product families or different lines of business. The results can also be industry-specific: in industries where product life cycles are on the short side, the input from
marketing and business executives consistently provided the most accurate forecast and thus got the highest weights.

Using an adaptive weighting scheme like this has an additional advantage. It can distinguish between individuals who provide accurate information and those who do not. If a particular person’s input is highly inaccurate, the weights assigned to it will be small, and his/her overrides will be effectively eliminated. But if this forecast improves, the weights will increase, and these overrides will have a greater impact on the final forecast.

A one-size-fits-all solution to lock out specific collaborative inputs over specific horizons can exacerbate forecast accuracy and affect the spirit of the collaborative process. The adaptive weighting scheme establishes an efficient temporal segmentation of the forecasting process to help improve the accuracy across different time horizons. Still, it is important to evaluate this method versus the simpler method of blending the inputs using simple averages.

**MAKING THE FORECASTS BETTER**

At the end of the day, our goal is to make a forecast more accurate and reliable so that it adds business value to the planning process. Increasing the accuracy of the forecast is not an end in itself. It is important only if it helps to improve the rest of the planning process.

Our experience indicates that sales inputs should be applied sparingly. A good indication that collaborative overrides are justified is that there is an accompanying cause or reason provided. It is also our experience that collaborative inputs should not be gathered too far into the future. This ends up “overriding the overrides” without any significant improvement.

The right question is not whether each of these inputs adds value, but whether each of these inputs can be combined to effectively integrate analytics and planner expertise into a better forecast.

**REFERENCES**

