The Business Forecasting Deal: Exposing Myths, Eliminating Bad Practices, Providing Practical Solutions
By Michael Gilliland

APPENDIX

Forecasting FAQs

This appendix provides answers to frequently asked questions drawn from conference presentations, webcasts, online discussion forums, and personal correspondence. For additional resources and information, or to submit your own questions, follow this book's accompanying blog *The Business Forecasting Deal* (blogs.sas.com/forecasting).

ACCURACY EXPECTATIONS

Q: How accurate should our forecasts be?
Q: How do we set performance goals?
Q: What is the minimum acceptable level of forecast accuracy?

A: Unfortunately, there is no way to know in advance how accurate our forecasts should be because we don't know the forecastability of the future demand. Therefore, it is wrong to set arbitrary forecasting performance goals, such as "Next year MAPE (mean absolute percent error) must be less than 20%." If demand is not forecastable to this level of accuracy, it will be impossible to achieve the goal. This can lead to a demoralized forecasting staff that just gives up trying—or figures out ways to cheat to reach the goal.

Reasonable goals for your forecasting performance are "to do no worse than a naïve forecast" and to continuously improve the process by making it more efficient. If your process chronically forecasts worse than a naïve model, then there is something terribly wrong with your process. (If this is the case, why not just discontinue the non-value adding process and just use the naïve model for your forecasting?)

You should not put a specific number on your forecasting performance goal. You don't know in advance what accuracy a naïve model will be able to achieve, so you don't know what specific accuracy you are expected to beat. However, you can monitor your process performance against a naïve model each period, and over time determine whether you are forecasting better than the naïve model.

Q: How accurate can our forecasts be?

Q: Is there an upper limit on forecast accuracy?

A: It is easy to determine the lower limit of how accurate a forecast should be: at least as accurate as a naïve model. Determining the upper limit of forecast accuracy is much more difficult. Theoretically, forecast accuracy is limited only by the amount of randomness in the behavior you are forecasting. If you can figure out the "rule" governing the behavior, if that rule doesn't change over time, and if there is no randomness in the behavior, then you should be able to achieve 100% accuracy.

Unfortunately, in real life, we are never sure we have captured the "rule" in our forecast model, we have no guarantee that the rule won't change over time, and we have to deal with some level of randomness. The coin-tossing forecasting contest in Chapter 1 illustrated this point. Randomness determined the limit of forecast accuracy for each process—there was nothing we could do to make the forecast any better.

Q: How do you assess *forecastability*?

A: A naïve forecast can be used to quickly assess the forecastability of the demand pattern by providing a lower bound on the forecast accuracy you should be able to achieve. The coefficient of variation (CV) of a pattern (its standard deviation divided by its mean) can also be indicative of forecastability in many situations (see the "Comet Chart" in Exhibits 1.3 and 3.2). For discussion of several more sophisticated approaches to assessing forecastability, read the series of articles in the Spring 2009 issue of *Foresight: The International Journal of Applied Forecasting*.

Q: No matter what we try, we can't seem to reach the level forecast accuracy the business needs (i.e., what management wants us to achieve). Any suggestions?

A: In some instances significant improvements over a naïve forecast are possible, justifying more elaborate techniques. However, in many cases, the naïve forecast will be difficult (or impossible) to beat, due entirely to the nature of the demand pattern. When elaborate methods *can* do better than a naïve model, the improvement can be disappointingly small, perhaps just a 10% reduction in the error (e.g., naïve method achieves MAPE of 50%, elaborate methods achieve a MAPE of 45%). You may still be unable to achieve the level of forecast accuracy desired. However, you should stay focused on process efficiency, so you aren't wasting resources. Automated forecasting software can often produce forecasts about as accurate as you can ever expect to achieve, and do this very efficiently, without occupying hours of expensive management time for reviews and approvals.

You may have reached the upper limit on accuracy given the nature of your demand patterns, and accuracy improvement may be impossible. In this case, if your forecast is still not good enough to make management happy, you need to seek other ways to solve your business problems. For example, find ways to smooth and shape demand to make it more forecastable (smoother demand is easier to forecast). Or adjust your supply chain capacity and flexibility so you aren't so dependent on accurate forecasts. For an excellent discussion of things to do when you can't forecast as well as you'd like, see the October–December 2009 issue of *International Journal of Forecasting* for a special section on decision making and planning under low levels of predictability, edited by Spyros

Makridakis and Nassim Taleb. As one example, we can't expect to forecast when we'll crash our car or burn down our house, so we buy insurance.

PERFORMANCE BENCHMARKS

Q: What is a "good" MAPE value, and what do you consider "excellent"? I work for a manufacturing company and have always used +/-10% error, three months in advance.

A: What would MAPE be if you used a simple (naïve) forecasting model such as a random walk (using the last known actual your forecast) or moving average? The performance of a naïve forecasting model should be the baseline for determining whether your values are good. It is irresponsible to set arbitrary forecasting performance targets (such as MAPE < 10% is Excellent, MAPE < 20% is Good) without the context of the forecastability of your data.

If you are forecasting worse than a naïve forecast (I would call this "bad"), then clearly your forecasting process needs improvement. If you are forecasting about the same as a naïve forecast, then why bother? Just use the naïve forecast and save yourself a lot of effort. If you are doing somewhat better than a naïve forecast you can consider that good, and if you are doing a lot better than a naïve forecast that could be considered excellent. Naïve forecasts can be surprisingly difficult (and sometimes impossible) to beat. Just avoid the trap of using arbitrary MAPE targets to evaluate your performance. Avoid the other common trap of comparing your performance to industry benchmarks that (among their many problems) fail to take forecastability into consideration. (The company with the lowest error probably has the data that is easiest to forecast.)

Q: Is there a standard MAPE by industry?

A: Comparing the MAPE, or forecast accuracy, among companies or industries may not be meaningful unless you look at the nature of the demand they are trying to forecast. A retailer that does every day low pricing will probably have much smoother (and easier to forecast) demand than a retailer that is constantly promoting and changing prices. Your accuracy expectations should be much different for the two retailers based on the way they do business and impact demand patterns. Industry benchmarks (available from professional organizations, consulting firms, or other sources) should be viewed with extreme skepticism. (See more discussion of this in the Perils of Benchmarking section in Chapter 2.)

Q: We are working on a forecasting process. What do other companies in our industry see as an acceptable accuracy rate?

Q: I am trying to benchmark our forecasting process against another company. Can you suggest a mid-sized company (approximately $1 billion annual revenue) that has about 1,500 active parts?

A: The level of accuracy your company deems acceptable should be based on realistic expectations. Expectations should be based on the nature of the demand patterns you are attempting to forecast. If you have easy-to-forecast demand, your company should expect reasonably accurate forecasts. If you have a lot of new products, short life cycle products, fashion products, lots of pricing and promotional activities, and overall highly erratic demand, then you may never be able to achieve the level of accuracy your company desires, so your expectations for accuracy should be low.

Although it is interesting to know what accuracy other companies are achieving, you should not use this information to establish your own accuracy targets. (A company with highly erratic demand should not be held to the forecast accuracy standard of a company with easy-to-forecast demand.) Instead, the proper approach is to determine the accuracy a naïve forecasting method would achieve with your data. Your accuracy goal is then to beat the naïve forecast. Management should consider it unacceptable for your process to forecast worse than a naïve forecast. (If you can't beat a naïve forecast, then save yourself a lot of time and money by firing the forecasting department and just use the naïve forecast.)

In short, base the acceptability of your accuracy on what a naïve forecast would achieve with your demand patterns. Your goal is to improve upon a naïve forecast. Although benchmark information is interesting to review, do not determine what is acceptable accuracy for your company based on what accuracy other companies are achieving. Their demand patterns may not have the same degree of forecastability.

Q: What is considered an accurate sales forecast? Our company is currently forecasting between 97% and 103% consistently, but our leadership continues to drive for better forecasting.

A: I wasn't aware that you could forecast better than 100% accuracy, but if you are, well, congratulations! What you are probably talking about is more properly described as the *bias* in your forecast (103% means that you typically forecast too high by 3%, 97% means that you typically forecast too low by 3%).

Accuracy is a vague term, and what is considered accurate may depend on context, such as the forecastability of your data. It is fine for leadership to drive for forecasting process improvements, but they cannot expect arbitrary levels of forecast accuracy. Your goal should be to beat the accuracy you would achieve using a naïve forecasting model, such as a random walk. (Don't base goals or expectations on industry benchmarks or other arbitrary targets.) If your accuracy is better than what a naïve forecast would have achieved, you are doing okay—your process is adding value by making the forecast better. If your accuracy is less than what a naïve model achieves, you have very serious problems with your systems and process.

PERFORMANCE MEASUREMENT AND REPORTING

Q: What is *forecast bias*? How is it measured?

A: *Bias* is a measure of whether your forecast is chronically too high or too low. The forecasting process is often a highly politicized part of an organization, and what should be a dispassionate, scientific process gets contaminated with the wants and wishes of process participants. You should measure forecast bias to help bring this to everyone's attention. (See the Glossary, for how to compute bias.)

Bias is based on some aggregation of individual forecasts, not on a single data point. For example, the difference between forecast and actual for an item, at a store, in a particular week is simply the forecast error for that item/store/week. Bias is measured by the summation of these individual errors across the aggregation of items, and/or locations, and/or time periods.

Be aware that the apparent bias in small aggregations can be meaningless, and may be due to chance. Don't conclude that your process is biased just because your forecast has been too high for two weeks in a row! However, if your forecast has been too high every (or nearly every) week for months and months in a row, this is a pretty good indicator that your process is biased.

Q: What does MAPE stand for? How do you measure MAPE?

A: MAPE stands for mean absolute percent error. You can find the formula in any forecasting text (or in the Glossary), but the gist is to compute absolute percent errors (APE) for each observation and average them, where

$$\text{APE} = |\text{Forecast} - \text{Actual}| / \text{Actual}$$

There are commonly used variations of MAPE, including weighted MAPE (WMAPE) and symmetric MAPE (SMAPE). Although suffering from several well-known flaws (such as being undefined when Actual = 0), MAPE (or one of its variations) seems to be the most commonly used forecasting performance metric in business.

Q: From your experience, which has been more reliable: root mean squared error (RMSE) or MAPE?

A: There are dozens of forecasting performance metrics. What is most appropriate for your business depends on its needs and who will be looking at your performance reporting. RMSE may be difficult to interpret by a nonstatistician, and size of the error depends on scale of the data. Since MAPE expresses the error as a percentage, you largely solve the scale problem, but MAPE has its own flaws (such as when actuals are zero or close to zero). For a good discussion of various metrics, see Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman's *Forecasting Methods and Applications*, 3rd Edition (John Wiley & Sons, 1998, pp. 42–45). I would suggest using the forecast accuracy (FA) metric when reporting performance to management. (See Glossary for definition of FA.)

Q: When using MAPE, what is your view on computing error against actuals versus against forecast?

A: The standard, textbook definition of MAPE uses actuals in the denominator. This avoids the issue of denominator management where there is a temptation to purposely bias forecasts on the high side (which makes the denominator higher so the error appears lower). Consider these situations:

Forecast	Actual	Error (F-A)	| F – A|	MAPE (actual in denominator)	MAPE (forecast in denominator)
80	100	–20	20	20%	25%
120	100	20	20	20%	16%

With actuals in the denominator, the forecast error is symmetric in that forecasting 20 units too high gives the same MAPE as forecasting 20 units too low. However, with forecast in the denominator, the MAPE is lower when the forecast is 20 units above the actual, compared to the forecasting being 20 units below the actual. This can create a subtle bias to forecast high to make your performance look better. The issue of what denominator to use was vigorously debated on the International Institute of Forecaster's discussion board, and reported on by Kesten Green and Len Tashman, "Percentage Error: What Denominator?" in the Winter 2009 issue of *Foresight*.

There really isn't a single metric that, by itself, expresses everything you need to know about forecasting performance. Most (if not all) metrics have their flaws—situations (such as division by 0) when they are undefined or give unsuitable answers. Both a bias and an error (or accuracy) metric are necessary. The forecast value added (FVA) metric addresses the whole separate issue of process efficiency and waste.

Q: Why do you prefer the forecast accuracy metric instead of MAPE?

A: I prefer the forecast accuracy (FA) metric for *management reporting of forecasting results* because it is easy to understand and interpret. FA is always scaled 0% to 100% (by definition, FA = 100% when both forecast and actual are zero).

I would not necessarily recommend FA for uses other than management reporting, and some in forecasting soundly (and correctly) criticize FA for its lack of useful mathematical properties. Also, you will get slightly higher FA by biasing your forecasts too high rather than too low. (This ability to impact results by biasing your forecast is common in forecasting metrics. For example, if actual = 100, then if you had forecast 110 (10 units too high) you would get FA = 91% while if you had forecast 90 (10 units too low) you'd only get FA = 90%.)

Q: Some of our analysts use the average of the forecast and the actual as the denominator in the forecast accuracy formula. What do you think of this approach?

A: I think you are referring to a metric called symmetric MAPE (SMAPE). It differs from traditional MAPE by using (F + A)/2 rather than A in the denominator. SMAPE improves on MAPE by avoiding some of the interpretation problems that occur when actuals are close to zero (and the percentage error explodes to very high values). However, like MAPE, SMAPE *by itself* gives us no indication of what forecast error we should be achieving, or whether our forecasting process is efficient and adding value.

If you mean you are using the FA formula but replacing Max(F,A) in the denominator by (F + A)/2, then this could be problematic under some circumstances. For example, if forecast is 2 and actual is 20, then you would have a forecast accuracy of (1 – |F –A|/(F + A)/2) = [1 – 18/11] = –64%. In other words, using (F + A)/2 as the denominator of the FA metric does not scale the values to between 0% and 100% and, therefore, defeats the purpose of the metric.

Q: Regarding forecast accuracy: Can you provide additional references that develop this measure of forecast performance?

A: There is not a lot of literature on the FA metric, but like all metrics, it has both benefits and aspects that users should be aware of. For example, while FA has the major benefit of always being scaled between 0% and 100%, and so is very good for management reporting, the user should be aware that it is nonsymmetric with respect to actuals. This means that you will get a slightly better FA by overforecasting by *x* units versus underforecasting by *x* units.

For example, if actual is 100, then a forecast of 120 would achieve FA = 83%, but a forecast of 80 would achieve FA = 80% even though both forecasts have the same absolute error of 20 units. Forecasters who are aware of this asymmetry may be able to achieve slightly better FA results by biasing their forecasts high. It is a good idea to always include a bias metric when evaluating forecasting performance.

Q: Any advice on monitoring forecast accuracy over time?

A: I'm a big fan of four things when it comes to forecasting performance reporting: forecast accuracy (FA) metric, control charts, coefficient of variation (CV), and forecast value added (FVA) analysis.

FA is hated by many in the forecasting profession, but by always having a value between 0% and 100% it is easy for management to understand. Visualization of performance is crucial, and statistical process control methods can be applied. What is very useful is a control chart of FA, showing multiple periods (e.g., two to three years of monthly or weekly results). Such a chart can be produced at any level of the hierarchy, at any level of granularity (individual items at locations, item/corporate, region/brand, all product/all location total, etc.). It lets the viewer instantly see about how accurately we are forecasting the particular thing (or aggregation of things). You get a sense of the variability in performance, and can readily detect any shifts in the process.

There are other nonstandard ways of reporting forecasting related data, for example:

- The forecast accuracy versus volatility scatterplot or "Comet Chart" (see Chapter 8 for details).
- Conducting FVA analysis (see Chapter 4 for details).

Q: You mentioned measuring the MAPE of your supply chain. How do you do that?

A: Mean absolute percent error (or its variations such as weighted MAPE and symmetric MAPE) can be used to compare planned versus actual behavior of any type. For example, a manufacturer may plan production, by week, three weeks in advance. To measure the production department's *adherence to the plan*, you would compare the production plan for a week to what was actually produced that week. Instead of using forecasted demand and actual demand in the MAPE formula, just use planned production and actual production in the formula.

We often assume that the supply side of an organization is much better behaved than the demand side. We recognize that we usually don't know what future demand will be with a high degree of accuracy, but we assume we know what our future supply is going to be with a high level of accuracy. Until you go through this exercise and determine the "accuracy" of your supply plans, it may be unwise to make this assumption. You may find that there is as much uncertainty in supply as there is in demand!

The uncertainty in supply plans needs to be accounted for in customer service and safety stock calculations. Reduction in the uncertainty of supply plans may have a greater impact on reducing inventories than reduction in supply lead times.

Q: Sometimes forecasts are displayed with confidence limits. Can confidence limits help improve forecast accuracy?

A: Confidence limits by themselves don't improve forecast accuracy. However, they do help your organization understand the uncertainty in your forecast. When the limits are narrow, you may have more confidence that the forecast is likely to be reasonably accurate. However, when the limits are extremely wide, this indicates that there is huge uncertainty in the forecast. Confidence limits provide useful information that is not available when the forecast is given as just one number (a single point estimate). Limits help you improve your overall planning process, because they give you a better understanding of the risk of forecast error.

Q: My team forecasts by quarter. Later the forecast is compared to historical data for the same time points. We aim for 10% or closer match-up. What can we do to further optimize our forecasting process?

A: I would first suggest forecasting by month or week, and measuring performance in those time buckets. It is always easier to get more accuracy by forecasting in larger buckets of time, or in aggregated levels of products or locations. But what good is forecasting at 10% error by quarter if you are way off week by week and have high inventories and terrible customer service? Remember, the goal is to run a good operation and make money.

I once heard a presentation by a guy who reported 98% forecast accuracy, and was pushing hard to get accuracy to 99%. (If anyone reports accuracy that high, get suspicious!) As it turns out, he was measuring accuracy of the forecast for the total year, for total company volume. His accuracy where it counted—by product, location, and week—was of course much worse than that, and his company actually carried very high levels of inventory without exceptional customer service. He was focusing on the wrong problem! Who cares if accuracy is 98% at the highest level of aggregation if you have terrible forecasts at the operational level where it counts?

However, if your forecast accuracy is meeting your business needs, then forecasting is not a constraint in your process and you don't need to worry about improving it anymore. In other words, if your forecasts are good enough then focus your efforts on improving something else.

Q: Last month my company started calculating forecast error metrics. We are now in the process of communicating this information to our employees, vendors, and supporting department. What are some good ways to communicate these error metrics?

A: Not sure what particular metrics you are using, but three essentials are some version of accuracy or error, forecast bias, and the value added (comparing your results to a naïve forecast such as moving average or random walk).

Traditional MAPE can be confusing to management since you sometimes get strange looking results (how do you explain a MAPE of 2000%?). For presentation to nonforecasters, the forecast accuracy (FA) metric has the advantage of being scaled between 0% and 100%, which is easy for anyone to understand.

Bias should be reported along with FA, as it indicates chronic problems of over- or underforecasting. Management often uses the forecast to represent what they want to have happen (optimistic) rather than what they really believe (or fear) will happen. It is good to expose chronic biases by reporting this metric, and it can help you avoid excess inventory situations (by being overly optimistic) or service issues (by being overly pessimistic).

The results of your forecasting process should always be compared to doing nothing (i.e., using a naive forecast like moving average or random walk). Presumably, all of your investment in systems and processes, and all of the management time spent on forecasting, are making the forecast better. Unless you measure it, however, you don't know this. In many cases (maybe most), you will find that all your efforts actually make the forecast worse. Tracking and reporting the forecast value added (FVA) metric lets you identify the non-value adding activities or participants, and mercilessly eliminate them from the process. The result can be better forecasts with less effort!

Q: Can we compare forecasting methods using Theil's U Statistic?

A: Theil's U was proposed in 1966. It is a statistic for making *relative* comparisons between your forecasting method and a naïve (random walk) method. It involves squaring of errors, so that large errors carry more weight in the calculation. You probably wouldn't use Theil's U as the metric within an FVA analysis, as it is already an approach to calculating whether a method is any good.

- When U = 1, this means that the naïve model forecasts as well as your forecasting process.
- When U < 1, this means that your process is better than the naïve forecast (and the closer to zero, the better your process compared to the naïve).
- When U > 1, this means that the naïve does better than your forecasting process.

For more discussion and the formula to use, see a good forecasting text such as Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman's *Forecasting Methods and Applications, 3rd Edition* (John Wiley & Sons, 1998).

Q: What is the best way to present monthly accuracy results?

A: It is nice if you have a full year (or better yet, many years) of weekly or monthly accuracy numbers, to get a sense of your organization's capability to forecast, and how much variation there is in the weekly or monthly numbers. Rather than put all this data in a table, it is better to graph it, with accuracy (or error) as your vertical axis, and time across the horizontal axis. This helps you view forecasting as the process that it is.

Q: What forecast accuracy metric have you found most effective for staffing call centers in the financial industry?

A: The metrics we discussed aren't specific to any industry or type of forecasting problem. For call center staffing or any other type of forecasting problem you should use some measure of accuracy or error (such as FA or weighted MAPE), along with a measure of bias (to check whether you are chronically over- or underforecasting).

Also, metrics like coefficient of variation (CV) let you understand the volatility of the patterns you are trying to forecast, and FVA lets you evaluate the overall efficiency of your forecasting efforts. With FVA analysis you can identify those process steps and participants that are failing to make the forecast any better, and that can be safely eliminated from the process.

Q: I work for a call center and I wanted to know how do I chart and interpret the forecast accuracy versus volatility for incoming calls. We forecast incoming call arrival numbers for every half-hour. Most of our clients are 24/7.

A: In the "Comet Chart" shown in this book (Exhibits 1.3 and 3.2), each point represents the volatility and forecast accuracy of an item/location over some time frame (such as 52 weeks or 36 months). There can be several thousand points on the plot, for each of the item/location combinations that is being forecasted and planned for.

In your situation, I would suggest creating the plot with 336 points, each representing a half-hour period on a given day of the week (336 = 7 days × 48 half-hour intervals per day).

Each week, measure the number of incoming calls in each of these 336 time periods, and compare to the forecast for each of these time periods. Over a year you have 336 time series, each with 52 data points (the number of incoming calls in that period each week, and the forecast for that time period each week).

Determine the volatility of the pattern by computing the CV (coefficient of variation) of each series of calls. This will give you the value for the horizontal axis.

You then should compute a forecasting performance metric for each period by comparing the forecasted number of calls in the period to the actual number of calls. I prefer the forecast accuracy (FA) metric that is always scaled 0% to 100% so it is easy for people to interpret. (You could also use MAPE or some other metric of error, but then the scatterplot would tend to slope up (as seen in Exhibit 3.2), with error increasing as volatility increases.)

You end up with 336 dots to put on your scatterplot, and place the dot according to the CV and the accuracy (or other forecasting performance metric). See the Step 1 section of Chapter 8 for more details.

THE NAÏVE FORECAST

Q: What are naïve models?

A: For a good discussion on naïve models and their importance in the forecasting process, see *Forecasting Methods and Applications*, 3rd Editon (John Wiley & Sons, 1998) by Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman. A naïve model is something simple to calculate, without expensive systems or elaborate processes. You can think of it as a forecast you can get for free. If all your forecasting efforts are not beating the naïve model, then they are a waste of time and resources.

The most basic naïve model, called naïve forecast 1 (NF1) in the Makridakis text, is the no change model—a random walk. This means your last observation becomes your forecast for future observations. If you sold 12 units last week, your forecast for future weeks is 12. If you sell 10 units this week, your forecast for future weeks becomes 10. And so on.

There are more sophisticated versions of naïve models that are still essentially free to create, yet may do better at forecasting. An example of a seasonal random walk (NF2 in Makridakis) is when the forecast each week is the actual value from the corresponding week in the prior year. This kind of model may forecast fairly well for seasonal demand. A moving average is also commonly used as a naïve model.

Q: How do I create a naive moving average forecast?

A: In your forecasting software you should be able to create a moving average model, a random walk, a seasonal random walk, or other simple forecasting model. (Consult your vendor if you cannot figure out how to do this, or get better software if it won't let you do it!) For example, for a 10-week moving average, just use your latest 10 weeks of actual demand as your forecast for future demand, recalculating the average every new week when you get the latest actuals.

You should also be able to reconstruct what these naïve models *would have* forecast in the past. With the reconstructed naïve forecasts, you can compare their accuracy to the accuracy of the forecasts your process created. This lets you determine whether you have forecasted better or worse than a naïve model in the past.

Q: What about using a simple time series forecast with no intervention as the naïve forecast?

A: Per Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman's *Forecasting Methods and Applications*, 3rd Edition (John Wiley & Sons, 1998), naïve forecasts are "Forecasts obtained with the minimal amount of effort and data manipulation and based solely on the most recent information available" (p. 607). Purists may argue that the only true naïve forecast is the "no-change" forecast, meaning either a random walk (forecast = last-known actual) or a seasonal random walk.

In forecast value added (FVA) analysis, our purpose is determining whether all our forecasting efforts are making the forecast any better—whether our elaborate forecasting systems and processes are adding value. For this objective, it is perfectly acceptable to use something more sophisticated than a random walk as the naïve forecast as long as what we use is something that requires no effort on our part and can run automatically. The key is comparing costly and heroic forecasting efforts to forecasts created by doing the minimum amount of work. Does the extra cost and effort make a meaningful improvement in the forecast? If not, then the cost and effort probably aren't worth it.

Q: Can a naïve model be accurate?

Q: Should an analyst's forecast always be able to beat a naïve forecast? Why or why not?

A: Naïve models can be reasonably accurate, and in some situations they can be surprisingly difficult to beat! The goal of your forecasting process should always be to beat (or at least do no worse) than a naïve model. If you can't consistently beat a naïve model, then why bother?

When you do forecast modeling, you assume there is some underlying systematic structure or "rule" governing the behavior, and you hope that your model captures that rule. You then assume that same rule governs future behavior. The forecast analyst never knows for sure whether he or she has captured the rule, or how much randomness there is in the behavior about the rule. Even worse, there is no guarantee that the behavior will continue following that rule in the future.

A naïve model, such as a random walk (where the forecast is given by the last-known actual value), makes no assumptions about the future. When you use a naïve model you are admitting that you really don't know about the future, and that using the last-known observation as your forecast is as good as anything. Sometimes we just over-think, and create more elaborate processes than the forecasting effort merits.

Q: You mentioned that organizations often forecast worse than a naive forecast—why is that?

A: There are lots of reasons, the three primary ones being randomness in the data being forecast, lousy forecasting models, and process politics.

Randomness limits forecastability. If your job is to forecast heads or tails in the toss of a fair coin, you will never consistently be correct other than 50% of the time. There is nothing you can do to forecast better (or worse) than this, so a fancy statistical model will be of no help. If the behavior you are trying to forecast has a lot of randomness and very little systematic structure, then a random walk or moving average may forecast just as well as the most sophisticated model you can create. It is important to understand randomness and avoid making excessive (and costly) efforts to forecast things that just can't be forecast that well.

Shoddy software, or forecast modelers lacking good skills, are other ways that the naïve can do better. In Chapter 2, we discussed the worst practice of "overfitting" the statistical model to history. In overfitting we see patterns in what is really just randomness. Poorly designed forecast software packages enable this by their ill-designed "pick-best" or "best-fit" or "focus forecasting" options that will select a model based solely on its fit to history over some recent time period.

Political contamination of the forecasting process is another way that elaborately constructed forecasts are frequently beaten by naïve forecasts. Elaborate processes can create too many human touch points, each giving the opportunity for someone to bias the forecast with their own personal interests. (Some might drive the forecast high, to create plenty of inventory. Others might drive the forecasts low, to reduce carrying costs but at the risk of missing orders.) When an organization fails to forecast better than a naïve model, it is often due to this sort of political contamination—using the forecast to represent what they want to have happen, rather than what they really believe is going to happen.

FORECAST VALUE ADDED ANALYSIS

Q: Measuring FVA sounds like a lot of work. Are there any tools available to do this?

A: While there is not an off-the-shelf package for conducting FVA analysis, it is possible to do a small-scale, one-time analysis in a spreadsheet. However, for conducting ongoing FVA analysis on an enterprise scale, spreadsheets won't do. Software such as SAS® Analytics Pro or SAS® Visual Data Discovery are ideally suited for this type of work. Such software can help automate the collection and reporting of the data. They have the scalability to capture all the steps and participants in a forecasting process, for all items and locations, for each time bucket. All of these data points are recorded, and then reports are generated to compare the forecasts to actuals and compute FVA.

Q: Is there something other than MAPE I can use in my FVA calculations?

A: Yes, you are free to choose the particular metric to use in your FVA calculation. FVA is defined as "The change in a forecasting performance metric that can be attributed to a particular step or participant in the forecasting process."Any of the usual forecasting metrics such as MAPE, weighted MAPE, forecast accuracy, bias, and so forth are suitable. It is good practice to utilize both an accuracy/error metric, and a bias metric (the former tells you the size of your errors, the latter indicates whether you are chronically forecasting too high or too low).

Q: Is FVA a quality control methodology like Six Sigma?

Q: Does lean methodology and Six Sigma relate to FVA? If so, how?

A: FVA is a metric for applying a lean methodology to the forecasting process—identification and elimination of process waste. Statistical process control methods, such as those in Donald Wheeler's *Understanding Variation: The Key to Managing Chaos*, 2nd Edition (SPC Press, 2000), are an important element of this approach. Management must understand variation in its demand and sales processes, and react appropriately. Unfortunately, organizations too often overreact to random behavior, wasting resources and degrading performance.

Q: How would this work in utility sales/demand forecasts that project annual numbers with a 20-year horizon?

Q: Comment on FVA in long-range forecasting applications.

A: To compute FVA you need to know your actuals, and what you forecasted. In order to compute the FVA of your long-range forecasting process, you'd need to wait until the forecasted time period had passed, so you'll have your actuals.

Very long-range (more than five years) forecasts are needed in some industries that make large, long lead time capital investments. Construction of a new power plant is one example. Unfortunately, you cannot evaluate the effectiveness of your current process using FVA until you have the actuals to compare to.

If you happen to keep records from the distant past, you may be able to make some judgments about your ability to forecast. One large technology manufacturer had good data systems and was able to go back six years to review forecasts they had made. These long-ago forecasts were evaluated against the known actuals, and performance could also be compared to what a naïve model would have achieved. This company found that they would have forecasted nearly as well using a simple model, so that an elaborate process may have been unnecessary.

Q: How would you compute forecast value added when forecasting a perishable product given the effect of last-minute price decreases?

A: I assume you are talking about any situation when a product is marked down or sold on clearance, which is common for perishable goods and fashion/apparel items, or any items at the end of their selling season. By lowering prices you can generate additional sales, but should these really count as "true demand" for the product, and be part of forecast accuracy calculations? And should you use this markdown demand in your history for generating forecasts next year?

Answering this last issue first, you should probably exclude markdown sales from the demand history you provide your statistical forecasting system. You make markdowns when you have unsold inventory that is at risk of going obsolete (or will cost too much to hold until the next selling season). Presumably, if your forecast had been correct, you wouldn't be left with all this excess that you have to mark down to sell. Since you probably don't want to repeat the error of forecasting too high, most companies will exclude the markdown sales from the demand history that goes into their statistical forecasting model.

Regarding FVA, you can decide in advance whether your forecast is for total sales (including off-price sales), or is only for full-price sales. Then measure performance accordingly. You can still use FVA in this situation to evaluate your forecasting efforts. Just be clear exactly what you are forecasting (all sales, or just full price sales).

Q: If you find that a process step or participant is failing to add value, you say to eliminate that step or participant. But shouldn't you try to improve their performance first?

A: This is not intended to bring a wholesale slash and burn and firing of forecasting process participants. The point of the FVA method is to bring attention to an area where attention maybe hasn't been focused in the past, to measure the value added by different steps and participants, rather than just assume they are all making the forecast better. There are many situations when long hours and heroic efforts by an overworked forecasting staff are not delivering payback. The first step should always be to find opportunities for improvement, and provide training and good systems for the forecasting participants to work with. But don't be afraid to eliminate process steps that have proven incapable of improving the forecast. And don't be afraid to reassign process participants, redirecting them to tasks that can benefit the company if they prove unable to benefit forecasting. It may be better to have your sales force out playing golf with customers—building relationships and making sales—than spending their time providing forecasts if those forecasts aren't any good. You can't just assume that more work and more investment in the process will deliver better results.

Q: The examples you cite are from manufacturing and retail companies. I work in financial services—can you apply FVA in services companies too?

Q: What would be the importance and usability of FVA in the finance industry?

Q: I am working on the jail population forecast. Do you think the FVA can be used on the population forecast?

A: FVA can be applied anywhere you do forecasting. Service organizations must forecast many things, including staffing requirements, revenues, office supplies, call center volume, cash requirements for ATMs, insurance claims, and yes, even jail populations. Wherever you do forecasting, it is important to get the best results you possibly can and do this as efficiently as possible. FVA applies in service companies because it helps identify and eliminate wasted efforts, and helps develop better forecasts.

Q: Do you know any FVA analysis application in clinical trial management?

A: One large pharmaceutical company has spoken publicly (a conference presentation) on their use of FVA as part of their demand forecasting process (see the case study in Chapter 4), but I have not heard of anyone using FVA in clinical trial management. FVA analysis can be used anywhere you do forecasting, so if there were some aspect of your clinical trial you were trying to forecast, you could use FVA there to determine the efficacy of your forecasting process.

Q: When was FVA analysis first used?

Q: What kinds of companies are using FVA analysis, and what have they found?

A: The concept of relative forecasting errors has been around for many years, but FVA analysis was first used at a CPG company around 1990. Forecasts made at each step of the forecasting process were stored, and then compared to actual sales. In addition, a naïve forecast (moving average) was computed. Each step in the process could then be compared to prior steps and to the naïve forecast. We found several interesting things that allowed us to streamline our forecasting process. First, we found that the statistical forecast was adding value by beating the naïve forecast. But we also found that management overrides generally made the forecast worse, so we urged analysts not to make overrides unless they had compelling reasons to make any change. We also found that input from field sales had no impact on forecast accuracy, so we were able to "excuse" them from the forecasting process. We ended up with a process that created good quality forecasts that were adding value, while using very little management time and effort.

The first mention of FVA analysis in the literature was my article "Is Forecasting a Waste of Time?" in the July–August 2002 issue of *Supply Chain Management Review*. Since that time, FVA concepts have been adopted at a number of major organizations across a wide range of industries (including retail, transportation, technology, manufacturing, and pharmaceuticals), and several case studies on their results appear in Chapter 4.

In some situations, organizations have found that a naïve model would have forecasted better than what they are doing. While this may seem a little embarrassing, it is actually very good news—they can start getting better forecasts just by switching to a naïve forecasting model!

Other organizations have identified specific process steps or participants that are or aren't adding value. In one instance, a consumer products company found that its statistical forecasting software was generating forecasts that were better than a naïve model—so the software was adding value. However, manual overrides made in a consensus forecasting process were making the forecasts slightly worse.

When a forecasting process is failing to add value, this is often the result of the process being contaminated by politics and personal agendas of the participants. The forecast should represent an *unbiased best guess* at what is really going to happen. But some process participants, including executive management, may try to use the forecast for other purposes, such as to express a goal they are pushing the organization to achieve.

Even when value is being added (and your process is beating a naïve model), you may not be adding as much as you'd expect. The future tends to be a lot more uncertain than we think it is, and in many situations even elaborate efforts may not make the forecast much better than a naïve model. The key is to focus on process efficiency, and not waste organizational resources trying to achieve levels of forecast accuracy that are not achievable given the nature of your demand patterns.

Q: Can you provide some examples on non-value adding activities you have seen?

A: The first thing to recognize is that any time your forecasts are good enough to meet your business requirements, any extra work you do on improving accuracy can be considered waste. If your forecasts are good enough to meet business needs, then forecasting is not the constraint on your business performance, so focus your efforts somewhere else.

If your forecasts aren't yet good enough, the most common non-value adding activities are overly elaborate processes involving too much management involvement, allowing each participant's biases to contaminate the process. We make assumptions that these elaborate processes work and that more effort is better. But we have to be cautious about this, make the measurements, and only do these things when we can demonstrate they are improving the forecast. Non-value adding activities typically involve manual overrides to a statistical forecast. For example, an apparel manufacturer found that for mature products (with at least two years of sales history), a moving average forecast would have been 25% more accurate than what their forecasting process generated. Their process allowed executive management to have final approval over all forecasts, and this was found to contaminate the process by adding personal biases and making the forecast worse. Other examples are available in the case studies section of Chapter 4.

In general, if you have good statistical software that is forecasting reasonably well, you should be very wary of manual overrides to the statistical forecast. Manual overrides consume management time, and they must be tracked to determine whether they are truly making the forecast any better. If not, they are just process waste.

Be particularly wary of an executive approval step in your forecasting process. The forecast should be an unbiased best guess of what is really going to happen in the future. The forecast should reflect the voice of the marketplace, and provide a heads-up when there are gaps between plans and what is really happening. When the forecast exposes a gap between reality and the plan, action can be taken to address the gap (e.g., promote products falling below plan, or increase the plan for successful products). However, it is often easier for management to simply bury the gap, by changing the forecast to match the plan.

Q: How would one implement the FVA analysis in the marketing world where there are seasonal differences over the year with different types of products?

Q: If the hourly predictions of a seven-day horizon can drift as the series moves throughout the year, would something like seasonal moving average be an example of suitable naïve model?

A: With seasonal data, you can use a seasonal random walk as your naïve model. (For example, use actuals from the corresponding week last year as your forecast for that week this year. Or use actuals from the corresponding hour last week, as your forecast for that hour this week.) A seasonalized naïve model will probably give better forecasts than a random walk or plain moving average, and be more difficult for your process to outperform.

Q: What do you think will be the FA and FVA of naïve model (NF1, NF2, and/or composite) and statistically developed models (with lot of changing variables)?

A: Forecast accuracy depends on the nature of what you are trying to forecast, more than on the particular model you are using. Thus, if you have easy-to-forecast data, you should get good forecasts no matter what method you choose. If you have erratic data with lots of random behavior, no method may be able to deliver the accuracy you desire.

I cannot say in general what will be the FA and FVA of a naïve model, because it depends on the data being forecast. If it is relatively smooth, without a lot of ups and downs, then the random walk (NF1) or a moving average may do quite well, and you may not be able to get significant improvement with a more statistically complex model. If the pattern is seasonal, NF1 may not do very well, but NF2 (the seasonal random walk) may be much better.

More complex statistical models have the advantage when there is truly structure in the data that can be understood and incorporated in the model. One situation when complex statistical models may not perform better than a naïve model is when the data truly has no underlying structure, and is basically just random.

Q: How can you conduct an FVA when predictions are adjusted on a weekly basis? Wouldn't this complicate the projections since they may be adjusted by management and/or the statistics so frequently?

A: Although it involves more data collection, there is nothing wrong with collecting forecasts adjusted each week, and comparing whether they improve as you get closer to the time period being forecast. Normally in FVA we use the forecasts "locked" at some time ahead of the period being forecast. How far in advance you lock the forecast usually depends on supply lead times. Thus, if you can react quickly to demand changes, you might measure FVA based on the forecast one week prior to the period being forecast. If you have long lead times, such as for apparel manufactured in Asia for sale in North America or Europe, you might measure FVA based on forecasts two to six months in advance of the period in which you will sell them.

We tend to assume that we can forecast better as we get closer to the period being forecast. This assumption is worth testing, as it doesn't always hold true. If you find that the forecast gets worse as you get closer to the period being forecast, this is a good indication of a politically contaminated forecasting process, where misses are rolled ahead into future forecasts, making them even worse.

Q: At what level of the hierarchy do you apply FVA analysis? SKU level? Product family level? The error varies highly depending at what level the forecast is.

A: Absolute errors should be calculated at the most granular level of detail, which is typically SKU at a location, per time bucket (usually week or month). Performance can be reported at higher levels (such as product family) as an aggregation of these absolute errors. However, if you aggregate the data first, and then compute the error, you are really reporting bias, not forecast error.

It is mathematically impossible to forecast better at more granular levels than at higher levels of aggregation. Many unscrupulous forecasters use this fact to make their performance look better than it is. (They report errors at higher levels of aggregation after first aggregating the data, rather than first computing the absolute errors at the most granular level.) But then, management often sets unrealistic (even impossible) forecast accuracy objectives that are unachievable without such gamesmanship, and just encourages cheating.

FVA can be applied at any level, using whatever metric (error, accuracy, bias, etc.) you choose. FVA is defined as the "change in a performance metric that can be attributed to a particular step or participant" in the forecasting process. With FVA, you only care about the change in the metric you choose, but it is still necessary that the chosen metric be calculated properly.

FORECAST MODELING

Q: How to address forecasting in a heavily promotions-driven retailer?

A: In this situation, the first question to ask is whether historical data on pricing and promotional activities has been tracked. In most places this information hasn't been maintained, or is not readily accessible. For example, the data may only be available in spreadsheets maintained by individual product managers or salespeople. The first thing a company needs to do is start tracking this kind of information—both historical activities and future plans. It may also be possible to reconstruct the history by gathering up all the individual spreadsheets and putting the data into a common form.

Once you have this information, it can be used in software such as SAS® Forecast Server and incorporated in the statistical forecasting models. In Forecast Server you can import the data on pricing, promotion, or other "events" and it will estimate the "lift" or change in the underlying forecast from doing these events.

One final note is that heavily promoted demand is generally more volatile than unpromoted demand and, therefore, more difficult to forecast accurately. Switching from promotions driven to an "everyday low pricing" approach will most likely stabilize demand patterns and make them more forecastable.

Q: What do you recommend to adjusting outliers?

A: First, be sure that the data point is truly an outlier, and not just an instance of the randomness or extreme variability of the data. I've seen some software that lets you essentially truncate any data points a specified distance from the mean. The problem with this approach is that it gives you a false sense of security. Your models will fit better on the modified (less volatile) history, and your forecasts probably won't jump around as much. But will the forecast be any better? And more important, will your confidence in the forecast accuracy be unjustifiably increased? I'm just trying to warn against becoming too aggressive in adjusting your historical data. Highly volatile data is telling you something—that it may be pretty hard to forecast accurately, and you should be aware of the risks.

There are several ways of handling extreme data points. These include: drop the point (exclude it from the history), smooth the point (swap it for a new value that is halfway between the points before and after it, or else forecast the point based on the surrounding points), or shift the point to some prespecified limit (e.g., three standard deviations from the mean).

A true outlier is something like a keypunch error (where it is simply the wrong number in there, so we need to replace it with something that is approximately correct). Or, it might be a correct number that was caused by some extraneous one-time event (e.g., a tornado strikes a store so that its sales go to zero for two weeks while it is repaired). When you need to modify history because of something like this, you can probably judge which method of filling in the history is most appropriate.

Q: How much historical data is reasonable for statistical forecasting?

A: As far as time-series models are concerned, the more historical data you have available the better (even though you may not use it all). The issue is that companies often keep only one or two years of granular level data. Technically, you can still generate a statistical forecast as long as you have one data point (e.g., by using a random walk model, or using the mean of a small number of data points).

If you think there is a seasonal pattern in the data, a common rule of thumb is to have at least three full cycles of observations to detect or estimate the seasonal pattern. (Thus three full years of data for calculating seasonality within a year, or three weeks of data to calculate the seasonality of days within a week.)

Q: If a forecast model has a low MAPE in its fit to history, will it forecast well?

A: If a model has low MAPE in the historical data ("in sample"), it may forecast well—or it may not. You can almost guarantee that the error of the future forecasts will be worse than the error of the fit to history. Sometimes the forecasts will be much worse. You can always construct a model that fits the history very well, even perfectly, but this is no guarantee that such a model will be any good at forecasting. Remember that the goal is to generate good forecasts, not to fit history.

There is a danger in focusing too much attention on MAPE of the model fit to history—"overfitting" the model (see Chapter 2) by confusing random ups and downs with structure. Rudimentary "pick best" functionality in many software packages makes this mistake—selecting models purely for their fit to history, not for their appropriateness for forecasting the future. For a much more thorough discussion of this issue, see Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman, *Forecasting Methods and Applications*, 3rd Edition (John Wiley & Sons, 1998).

Q: We use a MAPE for our statistic of fit for model selection. How do we set a threshold for MAPE that will signal bad forecasts?

A: There is no such threshold value. MAPE in the fit region ("in sample") is not a reliable indicator of accuracy of the future forecasts.

Q: How often should you revisit/calibrate your forecast (assuming you are utilizing a regression model)?

A: If the model is forecasting well enough to meet your needs, then just set it and forget it! In the pursuit of efficiency, don't waste your time tweaking something that is performing well enough for your needs—it isn't constraining your organizational performance.

However, if you find that the model is no longer forecasting as well as you really want it to, it may be time to rediagnose the history and recalibrate the model. Just be aware that wanting a certain level of forecast accuracy is no guarantee that you can achieve that level of accuracy no matter how much effort you put into generating the forecast. Some behavior is just not forecastable to the degree of accuracy desired.

Since diagnosing history, identifying appropriate model types, and fitting model parameters is more time consuming than just generating a forecast from an existing model, you only want to do these things when necessary. Some companies will go through the full diagnosis and modeling process on a periodic basis, such as every six months or year.

If you find that your model does not forecast well, and you are constantly recalibrating to try to improve performance, then either the type of model you are using is just not appropriate, or else the behavior is just not forecastable to the degree of accuracy desired. Some behavior is constantly changing, so a good model one period is completely inappropriate the next period. This is just the reality we have to deal with, and accuracy expectations should be managed accordingly.

POLITICS AND PRACTICES OF FORECASTING

Q: If you determine that a group is making the forecast worse, do you have any suggestions on removing them from the process?

A: If you suspect a particular individual or group is contaminating the forecasting process and making the forecast worse, your first step is to gather the data and conduct forecast value added (FVA) analysis to prove that this is true. That may be the easy part! The next step is to convince those in power within your organization that the process needs to be changed.

If you find, for example, that inputs from field sales is just making the forecast worse, you probably won't get any complaints from the sales reps themselves if you tell them they no longer have to forecast. Few people want to have to forecast, or be held accountable for their numbers. You might be able to suggest a better use of their time—perhaps asking that the sales force communicate when there are *significant* things impacting customer demand. But there is no need for them to communicate all the little ups and downs that tend to cancel each other out, and their time is better spent building relationships with customers and making sales!

Nobody is likely to change anything without solid data and analysis. But upper management should be willing to accept facts, and ultimately, it is in their interest to get better forecasts. It is also in everyone's interest to have the sales force out making more sales, rather than spending time making the forecast worse.

If the individual or group causing the problem is in executive management, this may be more of a challenge. Many organizations have a final step of executive approval for all forecasts, and those with approval power may not want to give it up. But again, if you have the data and analysis to show that this final step is making the forecast worse, you need to bring it to management's attention. Just be courteous and tactful when presenting the information, and don't use it to ambush anybody or publicly embarrass them. Remember, they probably think they are doing the right thing by overseeing the forecasting process and approving the numbers. But if you can demonstrate that this final approval process is simply a waste of effort and is actually detrimental to company performance, they might be willing to step aside—or at least be more willing to trust the forecasts that the forecasting professionals have delivered to them.

Q: From a forecasting department point of view, do you have any tips for dealing with executive politics regarding manual overrides?

A: If you find that executive management is intervening in the forecasting process, such as by requiring executive approval for all forecasts, then it is important to measure the impact of executive intervention. Is this approval step making the forecast better or making it worse? Executive intervention often makes the forecast worse, because the forecasting process is contaminated by what management wants to see, and the forecast gets used to express targets or plans rather than expressing the realities of the marketplace. Also, an executive approval step is costly in terms of the high-cost management time that is occupied in reviewing and approving the forecasts.

If you have evidence (FVA analysis) that executive intervention is making the forecast worse, you still have to be careful (and tactful!) in communicating this information. Most people are willing to change their behavior when shown their behavior has ill effects. However, some executives may prefer to maintain control over the forecast, and will continue to intervene even when you can demonstrate they are making the forecast worse.

Q: My company's management does not have a history of using objective forecasts. What is the best way to communicate to them that manual overrides simply add variability to forecasts and are not making them more accurate? Must I allow us to go through the "growing pains" of learning this lesson before we "lean out" the forecasting process?

A: The communication has to start with data. You can voice your opinion to management until you are blue in the face, but they may not be receptive to it and may have their own opinion. However, if you have gathered and analyzed your own company data, then you have the objective and scientific basis to justify your views.

If you don't yet have the data about your own company, you can utilize some of the ongoing research on this topic. You can read about methods that have been tried at other companies, and the results that were obtained, and use this to bolster your argument. Meantime, start tracking your own data and within three to six months, you should have enough evidence to bring to management on the impact of excessive overrides.

Research reported by Robert Fildes, Paul Goodwin, Michael Lawrence, and Konstantinos Nikolopoulos (see the Fall 2007 issue of *Foresight*, or the January–March 2009 issue of *International Journal of Forecasting*) suggests that it makes most sense to make a manual override when there is a big change you want to make—when there is some significant piece of information you want to incorporate in the forecast that wasn't included in the statistical model. Minor tweaking of the statistical forecast is generally not worth the effort, and tends to make the forecast worse. Lots of minor tweaking may just be overreaction to randomness in the data.

Q: Isn't it misleading to say management approval that increases the error is bad since an optimistic forecast will change behavior and can result in a net gain in profit?

A: The forecast should represent an unbiased best guess at what is really going to happen. There is nothing wrong with management setting aggressive targets and trying to push the organization to better performance. It is also management's prerogative to authorize extra capital investment for capacity, extra inventory, or extra advertising to help drive revenue growth. If there is a reasonable expectation that this activity and investment will increase sales, then the forecast should reflect that. Again, the forecast should reflect what you *really believe* is going to happen.

However, to simply raise a forecast on the hope and wish that you'll hit the number seems unlikely to result in increased profit. Do you really want to build a lot of extra inventory nobody believes you will be able to sell? Do you really want to promise Wall Street a revenue number that nobody believes you will hit?

The main concern is management authorizing numbers that have no reasonable expectation of being achieved, and calling this the forecast. This is just an invitation to trouble. However, if management wants to push for more aggressive revenue numbers, and authorizes actions to achieve this (adding salespeople, increasing advertising and promotion, pursuing new customers or markets, etc.) *and* it is reasonable to expect these activities to succeed, then a new higher forecast would make sense.

DEMAND VOLATILITY

Q: How do you get a company to change its practices to reduce volatility?

A: First, you need to determine whether reducing volatility is a good thing. The goal in business is to make money, and you can do this by reducing costs as well as by increasing revenue.

It should be fairly easy to demonstrate that reducing volatility will reduce costs. With lower volatility you need less capacity, you can service customers with less inventory, you can run operations more smoothly, and so on. However, the big question is the impact on revenue generation.

Customers across many industries have been trained to delay purchases until they can get the best deal, and this is often at a period end when vendors are scrambling to hit their short-term sales and financial targets. While some retailers employ EDLP (every day low pricing) policies, others maintain higher prices punctuated by promotions and sales, which leads to increased sales volatility.

It may be a nontrivial exercise to determine whether an EDLP policy is better for your organization from the revenue perspective. However, anything that reduces sales volatility will almost certainly reduce your cost of operations.

Think of an extreme situation, with highly volatile demand yet the ability to forecast it perfectly. Even though you have perfect forecasts, you have a lot of extra costs in maintaining the capacity to supply your product during the peaks, or building inventory in advance, or securing suppliers and a distribution network to handle the peaks. You then have considerable unused capacity (and unhappily underutilized suppliers) during the low periods.

Contrast the high volatility/perfect forecast scenario with very stable demand that you can't forecast perfectly. Since demand is stable, you can pretty much plan to supply the average, and not worry about the small random ups and downs that occur each week. Even though you can't forecast perfectly, you can still operate with much lower costs because of the demand stability.

Q: In a retail environment, I am not sure how one avoids the volatility of promotions. For example, Halloween is coming up, then Thanksgiving, Christmas, and year-end. These seem to be opportunities not to be missed rather than something to avoid.

A: There is demand volatility inherent in many types of products. Easter egg painting kits, Halloween costumes, and Christmas trees are good examples. Demand naturally spikes at a certain time of the year for these items. Unfortunately, most types of promotional activity make the volatility even worse, because the promotional activity is designed to increase the spike!

Increasing demand volatility tends to increase costs and makes it more difficult to forecast. Does it really make good sense to drive spikes even higher by cutting prices or spending promotional dollars, which will likely also increase your costs in fulfilling the demand? The answer to this question will depend on the situation, but it is at least worth asking. Wouldn't it make more sense to find ways to smooth and stabilize demand, which will simultaneously reduce your costs and increase your ability to forecast? Why not keep prices high (and promotional activity low) during the peak seasons, while cutting prices and increasing promotion during the low season? Or at least consider this approach.

Some things have zero demand in the office season, such as Christmas trees. In these cases, it makes no sense to promote them during the off season. How you approach this issue may ultimately depend on whether your objective is to maximize revenue, maximize profit, maximize market share, or something else.

Q: Is the coefficient of variation (CV) of point-of-sale (POS data) a clean measure of *inherent volatility?* Clearly, POS data is affected by customer promotions.

A: We defined *inherent volatility* as variation in *natural* (or *normal*) *consumption* of a product or service. As this question illustrates, the POS data at a retailer may not be a perfect representation of "normal" consumption if the retailer is running a promotion. The consumer may stock up on the product during the promotion and then either consume more of it than normal, or not need to repurchase it again until some later date. In these circumstances, the CV of POS data may not accurately represent inherent volatility.

If we look at this situation from the supplier's perspective, *artificial volatility* is still the difference between the CV of shipments to the retailer and the CV of the retailer's POS data. It would still be a legitimate question from the supplier's perspective whether their shipments were more volatile than POS consumption (and if so, then why?).

This question is valuable because we may never know the true inherent volatility of consumption. But as a practical matter—from the supplier's perspective—we can focus on the retailer's POS (whether it is "natural" or due to promotional activities is not our primary concern).

Looking at the issue from the retailer's perspective, however, it should be recognized that promotional activities tend to add volatility, and volatility tends to add costs. Sometimes promotional activities are used to "smooth" demand—usually in industries with practical limitations on short-term supply capacity (such as hotels, airlines, and utility companies). But most retail and consumer product promotion is designed to add volatility (and the costs of such volatility need to be fully recognized by those designing the promotions).

FORECASTING PROCESS

Q: How do you factor in customer returns, replacement shipments, and product substitutions into your forecast?

A: *True demand* is a nebulous concept, and there is probably no perfect operational definition. Ideally, you would start the statistical forecasting process with an historical time series of "true demand" and then project this into the future. As the question points out, there are a lot of things that happen to contaminate the pattern of true demand. Many of these are discussed in Chapter 1.

If you have good records of returns, substitutions, and the like, then you might be able to approximate true demand by making the appropriate adjustments to your historical data. The term *perfected history* has been used to describe historical data that has been somehow re-created to be more representative of what customers really wanted and when they wanted it.

If you generally have good customer service, then it is probably not necessary to bother worrying about these issues. Forecast error is generally quite high compared to imperfections in the historical true demand stream, and so it isn't even worth worrying about. (Fixing a few percentage points of error in the historical demand will not result in significantly better forecasts.) However, if returns, stockouts, and poor customer service are severe problems accounting for a high percentage of demand, then it may make sense to try to account for them in your historical data stream. Don't worry about trying to be perfect—just get the historical data as good as reasonably possible with the least amount of effort.

Q: What is your stance on forecasting based on shipments rather than on orders?

A: In well-performing organizations with high customer service levels (e.g., filling orders at 97% or above) then it doesn't really matter which you use, orders or shipments. Remember that the magnitude of forecast error is often 25%, 50% or more. Given that errors are typically this high, it doesn't really matter if the historical data stream you are using to generate the forecasts is off by a few percentage points. It is not worth the trouble trying to create history that perfectly represents true demand.

In situations when customer service is poor, then orders and shipments can be a long way apart. Unfilled orders may be cancelled, or rolled ahead to future time periods. During periods of chronic shortages, orders representing the same true demand may appear over and over again until finally shipped. Also, when customers or salespeople hear that service is low, they may purposely increase orders hoping to get a larger share of the allocation of available product.

Because of the amount of gamesmanship that can go into orders, you need to be very cautious is using orders for forecasting. Make sure you understand how unfilled orders are handled (cancelled, rolled ahead, or even filled with substitute items). There may be fewer opportunities to manipulate shipment or sales numbers, so I tend to prefer them as my historical data for forecasting. But shipments can also be far from perfect, as when shipments are made weeks after the original order was requested, or if orders are never shipped at all due to lack of supply. This is also very obvious in retail; when there is a stockout on the store shelf, you never really know how many more you could have sold.

You need to determine what is most appropriate in your situation, being aware of the pitfalls of each approach. You should also weigh the amount of effort you put into creating perfect history versus the payback. How much better are your forecasts going to be? If a lot of effort is not going to make the forecasts significantly better, then focus your efforts in areas that have more of an impact.

Q: We forecast over 100,000 time series with many causal factors changing. Limited staff requires handling this manually, so the result is using same model many times for all kinds of different products. Is there a better way, or any way to automate this modeling effort?

A: An issue for many companies is having too many series to forecast with limited forecasting staff. Out of necessity, they may have to reuse the same model across all time series, yet the model may not be appropriate for all series. Good software can actually build custom models for you automatically, for each series. For example, SAS® Forecast Server can look at all the causal factors you have provided, detect which ones are helpful in the forecast, and build custom models for each series based on the unique characteristics of the series. Forecast Server will also test whether transformations of the causal variables are useful. By having a custom model automatically built for each series, you have a much better chance of an appropriate model being used to forecast each series, which should improve forecast accuracy. Another major benefit is that a large proportion of your 100,000 time series can be forecast using automation alone, allowing your analysts to focus their attention on the more important or high value forecasts for the company.

Q: Assuming forecasting overrides typically make the forecast worse, why not just use the statistical forecast?

A: If you have conducted FVA analysis and determined that overrides typically make the forecast worse, then you may have very good reason to just use the statistical forecast. Researchers have found that overall there appears to be a slight net benefit to overrides. However, the benefits are mainly in large overrides when the analyst knows of significant things that are not being accounted for in the statistical forecast. They find that small overrides generally have no net benefit and are often making the forecast worse. Overrides consume costly management time, so they should be made only when it really matters.

If your demand patterns are fairly well behaved and amenable to statistical forecasting, then it makes sense to rely largely on the output of good forecasting software. Rather than wasting time on small adjustments that probably won't result in any improvements, analysts can focus their attention on those situations when the statistical model is missing something important (such as an upcoming event, an increase or decrease in geographical or channel distribution). When a large adjustment to the statistical forecast is needed, it probably makes sense to do that. However, the analyst should remain vigilant in tracking performance of statistical versus override, to make sure the override efforts are adding value and making the forecast better. It is certainly faster and cheaper to rely on the statistical forecast, rather than have to manually update every single forecast.

Q: What do you suggest if the cost associated with forecasting errors is not symmetric? For example, underforecasting costs more than overforecasting.

A: The forecast should still represent an unbiased best guess of what is really going to happen. If the cost of forecast error is not symmetric, you would take this into consideration in your planning process. For example, if there is a severe penalty for failing to ship an order, then you might authorize carrying extra inventory. Or, if inventory carrying costs are high, there is risk of product obsolescence, and customers can readily find substitutes if you go out of stock on a particular item, then you may plan to carry low inventory. In either situation, you still forecast what you really think is going to happen—what you are really going to sell—but you compensate for asymmetric error costs in your planning decisions.

Q: Should granular forecasts always add up to aggregate forecasts in a hierarchy?

A: Yes, the forecasting hierarchy should be reconciled, meaning lower levels add up to the higher levels. The forecast should represent your unbiased best guess at what is really going to happen. It would make no sense to believe that granular level forecasts will add up to one number, but the aggregate level forecast is another number.

There are situations when companies do not reconcile their hierarchies, but this in the realm of "planning" or "goal setting"—even though they may still call it forecasting. For example, management may make a conscious decision to have lots of inventory and not risk missing any sales. For the supply chain planning systems, they provide a so-called "forecast" for each item that is higher than they really believe they are going to sell. This triggers the supply chain organization to build to that higher number. However, for financial planning they authorize a forecast that is lower than the sum of the item level forecasts, which would be closer to what they really expect to sell.

You might play the same game with sales quotas. You give every salesperson a "stretch target" but don't really believe they all will hit it, so your territory or region has a target that is less than the sum of the individual targets. Again, the hierarchy doesn't reconcile, but this is in the realm of planning, not forecasting.

For planning purposes, your organization can do whatever it wants. There may be situations when the cost of excess supply is high (such as when your product has shelf-life limitations) and you are willing to miss some sales opportunities to avoid being stuck with excess supply.

Q: How can you get from unconstrained to constrained forecasts without any planning taking place? Am I mixing something up? Is the term *constrained forecast* somewhat misleading? Should it instead be called a *constrained plan*? Is the difference between the unconstrained and constrained forecast just a net calculation and not planning?

A: We start the process trying to figure out what customers want and when they want it (a guess at unconstrained true demand). Then we identify constraints related to supply (can we make or procure that much?) or other financial or budgetary constraints. At some point we need to determine what we really think is going to happen, and this I have called the *constrained forecast*. It represents our best guess at what is really going to happen—how much we are really going to sell. You need this number to make appropriate revenue estimates. (The revenue forecast should be based on what you really expect to sell, *not* on demand that is known in advance you cannot fulfill.)

The plan represents a specific course of action that the company decides to undertake. For example, it may plan to build enough inventory to sell 10,000 units, or it may plan for profits of $5000 on revenue of $50,000. There are sometimes reasons why your forecast does not equal your plan. For example, you might create a production plan to build 10,000 units and you might forecast (truly believe) that you will sell all 10,000 units, but to be safe you build a financial plan assuming sales of 9,000 units and a profit of $4500 on revenue of $45,000.

Measuring forecasting performance should be based on the constrained forecast of what you really think is going to happen. The reason for this is that you never really know what true unconstrained demand is because there is no good operational definition of true demand. So you cannot *with certainty* measure whether your unconstrained forecast is good or bad, because you cannot observe what unconstrained true demand really was. This was discussed in Chapter 1.

Q: When do you stop trying to improve forecast accuracy?

A: Some rules of thumb:

1. Is your forecast accuracy good enough to meet your business needs? If so, don't waste any more time building fancier models or developing a more elaborate process.

1a. What are the consequences of a less-than-perfect forecast? If the costs and consequences are small, don't waste your time trying to get great forecasts. Focus your efforts on those forecasts that have the most impact on your business.

2. Compare your forecasting performance (MAPE or whatever metric) to what you would have achieved with a naïve forecast (such as random walk). If you are forecasting worse than a naïve model, then you need to keep tuning your process since there is plenty of room for improvement. Your goal is to do no *worse* than a naïive model.

3. Once you are beating a naïve model, you need to consider how much additional improvement is possible, and whether the improvement is worth the effort. Naïve models can be surprisingly difficult to beat by a large amount. If your error is 10% or 20% less than the error of a naïve model, that may be about the best you can ever do, so don't waste time trying to reach perfection.

4. There is some research on the theoretical limits of forecast accuracy. Basically, the best you can do is discover the "rule" or "structure" governing the behavior and hope that demand continues to follow that same rule into the future. (Of course, there are no guarantees that the rule won't

change in the future—and most likely it will.) Once you have discovered the rule, your accuracy is limited by the amount of randomness in the behavior about the rule. This was illustrated with the coin tossing contest in Chapter 1.

In short, if you are already beating a naive forecast by a good amount, you are probably forecasting about the best that can be expected. Unless these are very high value forecasts with a lot of impact on the company, there is no need to waste your time trying to improve them further. Focus on the things that really matter to the business. And if you aren't beating a naïve forecast, than stop whatever it is you are doing and switch to a naïve model—you'll get better forecasts for free.

Finally, anything you can do to smooth demand patterns will give you better forecasts for free. If you have organizational policies and practices (such as lots of sales promotions, price changes, quarter-end pushes, and so on) you are creating incentives for your customers to buy in erratic patterns, and that is making demand harder to forecast. If you can encourage smooth and stable demand for your products, that is the surest way to get better forecasts.

JUDGMENT

Q: A hands-off approach is great when the statistical models work well, but what are some common factors that require an analyst to manually apply judgment?

A: An ideal state is when the statistical forecasting models are doing well enough on their own that the analysts don't have to touch them. However, there are plenty of situations when a model can be blindsided by new information, or information that is not made available to the model. So there will always remain a place for human intervention in the forecasting process.

For example, if you are forecasting fashion apparel, or any item that has color choices, how will the computer know what colors are going to be in fashion next season? (Are you going to be modeling a time series of the numerical wavelengths of popular colors over time, and then forecast the next popular wavelength? Probably not.) In any kind of new product forecasting, judgment will be involved, and can be assisted by analytical methods such as the structured analogy approach (see Chapter 5). Known changes in organizational policies and practices can cause fundamental shifts in demand behavior. An example of this is a retailer putting an item on promotion that hadn't been promoted before. Until this information is incorporated in the forecasting models, it will have to be applied to forecasts through manual overrides.

FORECASTING ORGANIZATION

Q: Where should the forecasting function reside in an organization?

Q: Typically, who takes the stewardship for forecasting in an organization? If it is marketing, how do you guard against forecasts based on anecdotal data they get from sales?

A: The typical places for forecasting are in sales, marketing, finance, operations, or in an independent forecasting department. There is a good article on this topic "Where Should the Forecasting Function Reside?" by Larry Lapide in the Winter 2002–2003 issue of *Journal of Business Forecasting*. (The article can be downloaded at the Institute of Business Forecasting's website, www.ibf.org. There is a small fee for non-IBF members.) Lapide reports on survey data showing the forecasting function residing all over the organization, and suggests that "the right answer … is to put the forecasting function inside a department that will diligently execute an effective forecasting process." Lapide also provides a systematic "pros and cons" argument for locating the forecasting function in sales, marketing, operations, finance, strategic planning, or as a stand-alone forecasting department.

Overreacting to anecdotal information is a risk wherever forecasting resides. Human beings have generally poor ability to distinguish signal from noise. As such, we tend to put too much emphasis on the latest piece of information and are unable to distinguish the random ups and downs present in every demand pattern. As an owner of the forecast, you must guard against overreaction to random noise, and you must also guard against the systematic biases of those providing information to you. The best way to protect yourself is to gather data on each step and participant in the forecasting process and over time, track their performance. You might be able to identify chronic biases that can then be communicated back to the contributors (to help them improve). Or you can simply stop listening to their input.

LOW VOLUME/INTERMITTENT DEMAND

Q: We have a very long tail of active but very low volume SKUs (about 1,000 SKUs with annual volume of less than 1,000 units). How do you deal with a large population of low volume SKUs?

A: Have you considered pruning the low volume SKUs? What is the total revenue they deliver? What is the total cost of continuing to produce, inventory, and manage all of these low volume SKUs? Your best solution may be to simply get rid of them (and not have to worry about forecasting them!). I have seen (at a consumer packaged goods (CPG) and an apparel company) that 25% to 50% of the total SKU count may generate less than 1% of your total revenue. Unless there is a really good reason to keep them, you are probably better off getting rid of them. An annual pruning exercise should be undertaken by all companies to identify (and get rid of) nonperforming SKUs. This lets your company spend resources on products that are actually making money. (See Chapter 6 for more discussion of pruning.)

Q: What metric do you recommend when demand is intermittent (MAPE may be undefined in many cases)?

Q: How do I get the most accurate forecasts with intermittent demand?

A: Intermittent demand is a difficult situation not only for forecasting but for measuring forecasting performance. MAPE, and many other traditional metrics, are inappropriate because they yield infinite or undefined values when there are zero values for actuals.

If the goal is simply to achieve the best forecast accuracy, forecasting 0 every period will probably give you the lowest error. This was discussed in a presentation by Ruud Teunter at the 2008 International Symposium on Forecasting, and in the Forgetting the Goal section of Chapter 3. However, the business goal is probably not forecast accuracy, but properly managing the inventory and servicing customers. For guidance on this, see the June 2006 issue of *Foresight* for a special section on accuracy metrics for inventory control.

Q: We are a retailer trying to forecast store/item demand every week, and we aren't doing very well. How do we get better forecasts at this level of granularity?

A: Demand is often intermittent (lots of zeroes) and erratic at the most granular level of detail. You may be unable to get a good statistical model at that level, and even if you can find an appropriate model, you may not be able to expect highly accurate forecasts.

It is often better to focus your forecasting efforts at some intermediate level, and then apportion the forecast down to the most granular levels. Better still, you may be able to forecast and manage inventory, capacity, and so forth at that intermediate level, and use supply chain management techniques to handle the levels below.

Consider an example from retail. You may have thousands of items and hundreds (or thousands) of stores. One option is to try to forecast each of these store/item combinations each week, and you can do this with large-scale automated forecasting software, but may still not get highly accurate forecasts. Instead, consider the real business objective, which is to have the appropriate amount of inventory on store shelves to meet customer demand, without excess inventory and without lost sales. Often the cause of store-level stockouts is poor replenishment practices from the store's warehouses. If the warehouses aren't managing inventory properly, or aren't able to distribute it promptly, then it may not matter how well you can forecast as the store level. Shrink (loss, theft, or damage) of product at the store level can be another problem.

A reasonable approach in this situation is to focus your forecasting efforts at the item/warehouse level. You have a much better chance of having good forecasts at this intermediate level (aggregating all the stores that pull from the warehouse). As long as you maintain the appropriate level of inventory at the warehouse, and are able to promptly fill orders from the stores, then shelf inventory in the stores should be managed properly. You may be much more effective in meeting your overall inventory and service objectives by using good inventory management practices, rather than expecting to solve the store/item issues just by better forecasting at that level. (You could have perfect forecasts at the store/item level, but if your warehouse can't fulfill replenishment orders in a timely fashion, the perfect store/item forecasts are of little value.)

NEW PRODUCT FORECASTING

Q: How to do forecast when you don't have enough data points?

A: When you have little or no historical data, judgmental methods can be used. One common example is forecasting by analogy, which is when you forecast demand for a new product based on historical demand for a similar product from the past. There are many other judgmental methods and new product forecasting approaches discussed in the forecasting literature (see Chapter 5). The approach of new product forecasting by structured analogy, also discussed in Chapter 5, combines judgment with statistical analysis of past new product introductions, and allows you to visually assess the likely uncertainty in the new product forecast.

It is important to manage expectations for accuracy when forecasting new products. It is unlikely you will have a high level of accuracy, and so the organization must assess the risks of forecasting significantly too high, or significantly too low. These risks must be accounted for in your operational plans (procurement, manufacturing, distribution, etc.) and your financial plans. A big mistake is to assume that your new product forecasts are going to be highly accurate, and to make significant financial commitments based on this assumption without proper assessment of risk. The structured analogy approach is an application of time series exploration techniques to provide visualization of past new product introductions and the range of outcomes. This can help provide some indication of the possible range of errors in the new product forecast.

FORECASTING HIERARCHY

Q: In our company we like to see our forecasts rolled up into lots of different levels for reporting: sales, marketing, finance, operations and so on. As a result, we have a huge and complex hierarchy with over 10 levels, which bogs down system performance. Is there a right number of levels to use for hierarchal forecasting?

A: The right number of levels is the fewest you will need for forecasting. This will vary by the nature of your business and the structure of your organization. There is a common mistake of confusing levels needed for reporting with what you need for forecasting. Reporting should be separate from generation of the forecasts.

Having too many levels in your hierarchy will degrade software performance. Extraneous levels will also bloat the storage requirements for all the historical data and future forecasts. If you do not need to generate a statistical forecast or make a manual override at a level, it should not be in your forecasting hierarchy. Forecasts can always be exported to a reporting system, and reaggregated or apportioned to the levels required for reporting only.

Many software packages require you to specify hierarchies during system implementation. Later on, if you decide or add or remove levels, this is a major effort, since you have to reimplement the software with the new hierarchy structure. One of the great benefits of SAS® Forecast Server is that forecasting hierarchies are completely flexible and are specified on-demand by each user. If you find that the hierarchy you are using needs to be changed, simply create a new "project" specifying the new hierarchy. This is done by the forecaster and takes minutes. Reimplementation of software with fixed hierarchies can take weeks or months, and cost tens or hundreds of thousands of dollars.

Q: You talked about minimizing the complexity of the forecasting hierarchy and not putting in attributes. Can you elaborate on that?

A: Be sure to distinguish the generation and manual adjustment of forecasts, from the reporting and use of forecasts for management and downstream systems. If you aren't routinely generating a statistical forecast at a level, or making manual adjustments at that level, then it probably does not belong in the forecasting hierarchy.

Attributes are things like color or package size. You might be interested in how many units of red products you are forecasting, so that you can merge with your bill-of-materials and determine raw material requirements (e.g., how much red dye you need to procure). But this is reporting, not forecasting. You do this by exporting the forecasts from the forecasting system, into a reporting system, to create reports based on attributes. You probably don't have your forecast analysts assigned by red products or green products, so these attributes don't belong in the forecasting hierarchy. Software performance will always be better with fewer levels and nodes in the forecasting hierarchy.

Q: We typically have four to six levels in our hierarchy. What's the best way to "roll it up"? Top down? Middle out?

A: Usually middle out, from one of the intermediate levels works best. But you will have to determine this from your own data.

Demand at the most granular level of the hierarchy is often difficult to model effectively, because demand is too sparse or erratic. You end up with a bunch of flat line models at the lowest level, representing average demand, failing to account for seasonality that truly exists but is undetectable amidst all the random noise.

At the highest level of the hierarchy, you miss out on the subtleties of behavior in lower levels. A retailer might sell swimsuits and snow shovels, both highly seasonal items with very different seasons—and this detail would be lost in a model of overall behavior.

By focusing your attention at some intermediate level, you will probably have enough data to build decent models capturing underlying trend, seasonality, and so on. Forecasts can then be apportioned down to the most granular levels using some method of percentage splits, and summed up to higher levels.

Q: Is there a fixed or practical limit to the number of levels that can be defined within a hierarchy?

A: While there isn't a fixed limit on number of levels within a hierarchy (unless your software has such a limit), it is good practice to keep the number of levels as small as necessary for *forecasting*. Users of forecasting software (from whatever vendor) usually want to be able to see forecasts at a variety of levels, based on various product or location attributes. It is important to distinguish such *reporting* needs from what is required for good forecasting, and only use levels necessary for forecasting in the forecasting project. For example, in SAS® Forecast Server, reporting can be done with a stored process, or outside the forecasting project, and can be based on attributes not used in forecasting.

If forecasters think they need more than 10 levels in their forecasting hierarchy, they are probably mistaken. Are they really going to care about creating models at each level, and is someone going to be paying attention to the statistical forecast at each of those levels, and doing overrides at each level?

Another consideration is the level of granularity in the hierarchy. Ultimately, you have to come up with a forecast by product and location (in whatever time bucket is appropriate). But sometimes it makes sense to not go all the way to the most granular level within the hierarchy. In apparel, for example, the amount of data starts to explode once you get down to style/color/size. There may be dozens of size combinations available (e.g., waist and inseam on pants), and several colors. Of course, data can become very erratic at the most granular level, and you may not be able to detect true seasonality from the randomness, and end up with flat line models. It is often better to forecast at a higher level, such as style, or style/color, and then apportion the forecast down to SKU level (style/color/size) by size curves based on historical demand.

In short, less is more. Don't make the hierarchy any deeper or more complicated than necessary—you may get better forecasts (and better performance) by keeping the hierarchy smaller and simpler.

SOFTWARE SELECTION

Q: Our company spent over $1 million buying and implementing a forecasting software package. But a year later we still aren't forecasting or planning any better than before. What are we doing wrong?

A: Before buying new forecasting software you should run a thorough test on your own data to determine how well it will be able to forecast. Don't just look at how well the software can create models to fit your history—that is easy! What is hard is generating good forecasts, and even good software may not always be able to do this. Also, always compare the statistical forecasts from the software to a naïve model, to see how much better (if any) are the statistical models. As discussed throughout this book, naïve models are often difficult to beat, particularly if your demand patterns have a lot of randomness. Some demand is simply unforecastable to the degree of accuracy desired.

If you determine that the software is sound, and can generate forecasts that are better than before, the next step is to look at the nature of your forecasting process. Are you truly providing an unbiased best guess at what is really going to happen? Or are process participants contaminating what should be an objective and scientific process with their own personal agendas. Perhaps you have an evangelical forecasting process in which all numbers must be approved by executive management? This sort of thing is just an invitation to increase forecast error. You can spend a million dollars (or even much more) on good software, have it generate better forecasts, but still see no benefits if management won't use the numbers. You may need to make fundamental changes to your process, such as eliminate the management approval step, to be able to achieve better results. FVA analysis can help expose these sorts of political biases and process inefficiencies.