CHAPTER 1

Fundamental Issues in Business Forecasting

Forecasting is a difficult, thankless, and sometimes futile endeavor. When accuracy is not quite where everyone wants it to be, we react by making significant new investment in technology, process, and people to solve the business problem.[1] While we live in an uncertain and largely unpredictable world, we prefer to operate under, as Makridakis and Taleb suggest, an "illusion of control."[2] We think a bigger computer, a fancier model, and a more elaborate process are all we need to get better forecasts, but the world doesn't work that way.

THE PROBLEM OF INDUCTION

Scottish philosopher David Hume formed an early statement of the problem of forecasting in his 1748 book, *An Enquiry Concerning Human Understanding*. Hume was concerned with induction, the reasoning from particular facts to general conclusions. Forecasting is an example of this.

Hume observed that when we eat a piece of bread, it nourishes us. So we extrapolate our finite particular experiences to a general belief that bread nourishes. When we do this, we are in fact creating a forecast that the next piece of bread we eat will nourish us.

But Hume was a philosopher, so it was his job to be a critic and to question everything that we blindly believe. He was compelled to ask the question, What justification is there for this belief that bread nourishes? Just because bread has exhibited this nourishment characteristic in the past, is that any proof that bread will continue to nourish in the future? In the business forecasting world, just because customer X has always ordered product Y in some particular pattern, is that any proof that this behavior pattern will continue into the future? Of course not! Hume came to the skeptical conclusion that there is no proof that the future will behave like the past. He resolved that, in fact, we have no justification for our forecasts. Hume is correct. It has been over 250 years, and there is still no refutation of Hume's argument, and no evidence that the business world is getting any easier to forecast.

We should not let Hume's ultimate skepticism completely ruin our day, however. Just because we have no *logical proof* that our forecasts will be any good doesn't mean we never ought to try. This historical perspective isn't meant to completely discourage us. But it does provide a valuable reminder that we aren't gods, we don't have special powers of omniscience, and that we have no right to expect to consistently predict the future very well.

THE REALITIES OF BUSINESS FORECASTING

The unfortunate reality is that investment in the forecasting function is no guarantee of better results. There are often fundamental issues that impact an organization's ability to forecast accurately, yet these are largely unrecognized or ignored. Until these issues are addressed, however, further investment in the forecasting function may be wasted. We begin by identifying several fundamental issues that must be dealt with in pursuit of our objective:

To generate forecasts as accurate and unbiased as anyone can reasonably expect them to be, and to do this as efficiently as possible.

So who is to blame for all the unrealistic expectations around ***forecast accuracy***? Unfortunately, these unrealistic expectations are perpetuated by many in the profession (including those selling forecasting-related software or services) who know better, or who at least should know better. The dream of forecast accuracy is always easier to sell than the harsh reality.

The harsh reality is that predicting the future is a very difficult thing! As statisticians and forecast analysts, the best we can ever do is discover the underlying structure or rule guiding the behavior that is being forecast, to find a model that accurately represents the pattern of behavior, and then pray the behavior pattern doesn't change in the future.

Assume we do discover the underlying structure of the behavior, we correctly model that structure in our forecasting software, and the structure does not change in the future. Should we then be able to achieve perfect forecasts? Unfortunately, the answer is no. In any complex business or social system (including things like the buying behavior of customers), there remains an element of randomness. Even though we know the underlying structure and model the behavior correctly, our forecast accuracy will still be limited by the amount of randomness, and no further improvement in accuracy will be possible. We can see how this works with a forecasting contest.

THE CONTEST

There are three processes to be forecast:

P10: The percentage of heads in the tossing of 10 fair coins

P100: The percentage of heads in the tossing of 100 fair coins

P1000: The percentage of heads in the tossing of 1000 fair coins.

Every day, the three processes will be executed: The coins will be tossed, and we have to predict the percentage of heads. What is our forecasted percentage of heads each day for each process? Can we forecast one process better than the others? What accuracy will we achieve? Are there any investments we can make (better software, bigger computer, more elaborate forecasting process, more skilled statistical analyst) to improve our accuracy?

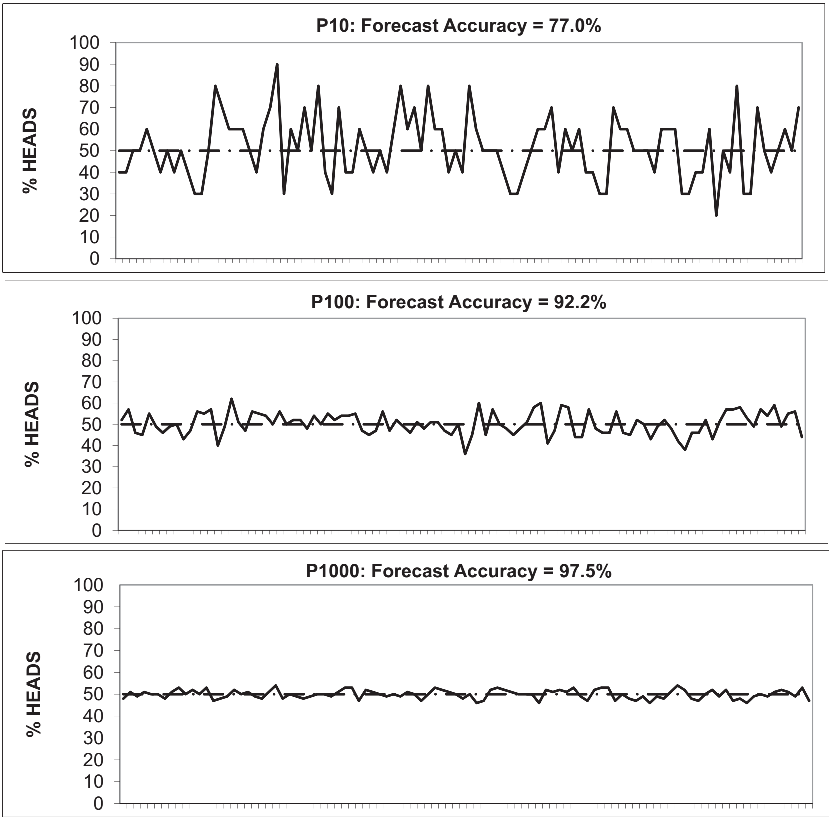

Exhibit 1.1 One Hundred Trials of P10, P100, and P1000[a]

[a]The accuracies reported in this exhibit are based on 100 simulations of each process. The mathematical expectations of accuracy calculated using Blythe (1980, 2.2) are 75.4% (P10), 92.0% (P100), and 97.5% (P1000). (Colin Blythe, "Expected Absolute Error of the Usual Estimator of the Binomial Parameter." The American Statistician, *August 1980, Vol. 34, No. 3, 155–157.)*

This isn't meant to be a trick question, and it doesn't take a doctorate in statistics to figure it out: The only rational forecast each day for each process is 50% heads. Exhibit 1.1 illustrates 100 daily trials of each of these processes. Since we are dealing with the independent tossing of fair coins, then, by definition, each process behaves according to the same underlying structure or rule—that over a large number of trials, each process will average about 50% heads. We fully understand the nature of each process, and we realize it makes no sense to forecast anything other than 50% heads each day for each process. However, as illustrated in the exhibit, the variation in the percentage of heads in each process is vastly different, as is the accuracy of our forecasts.

When there is a lot of randomness, or noise, in the behavior, we cannot expect to forecast it very accurately. Even when we know everything there is to know about the rules guiding the behavior, as we do here, the amount of randomness limits how accurate we can ever be. Also, in situations like these, any additional investment in the forecasting process would be a waste. There is nothing we could ever do to forecast P10 more accurately than P100, or P100 more accurately than P1000. The nature of each process, its underlying structure along with its random variability, determined the level of accuracy we were able to achieve.

Real life demand patterns are different from this in that the underlying mechanisms, knowable or unknowable, are not so simple as to be illustrated by independent tosses of a fair coin. Real life demand patterns may or may not have an underlying structure, we may or may not be able to discover and model that underlying structure, the underlying structure may or may not continue into the future, and there will be some degree of randomness.

What makes real life demand patterns so difficult to forecast is that the underlying mechanisms guiding their behavior may not be so apparent or may not even exist. Even if there is some structure to the historical pattern, it may not be obvious and can require good software or a skilled analyst to uncover it. But even then, even if we can discover and model the underlying behavior, there is no guarantee the behavior won't change over time. As forecasters, why do we even bother to try?

The coin tossing contest illustrates that there are limits to the forecast accuracy we can achieve. We can't assume that by applying more data, bigger computers, and more sophisticated software, or by exhorting our forecasters to work harder, we can always achieve the level of accuracy we desire. It is important to understand the limits of forecast accuracy, and to understand what level of accuracy is reasonable to expect for a given demand pattern. The danger is that if you do not know what accuracy is reasonable to expect, you can reward inferior performance, or you can waste resources pursuing unrealistic or impossible accuracy objectives. You can also miss opportunities for alternative (non-forecasting) solutions to your business problems.

WHAT IS DEMAND?

This book is about forecasting for products and services. It is not about forecasting the weather, or interest rates, or the outcome of sporting or political events. It is about forecasting the quantity of things people will buy or the quantity of services they will seek.

In business forecasting, we talk about demand every day. We don't think much about our use of the word because it seems pretty straightforward. Demand is commonly characterized as, "what the customers want, and when they want it," sometimes with the added proviso, "at a price they are willing to pay, along with any other products they want at that time." So far, everything seems to make sense.

When we refer to demand, we usually mean *unconstrained* or *true* demand because we take no consideration of our ability to fulfill it. (Note: I will treat demand, unconstrained demand, and true demand as synonyms.) We use *constrained* demand to describe how much of true demand can be fulfilled (after incorporating any limitations on our ability to provide the product or service demanded). Thus, constrained demand ≤ demand.

A good forecast of demand, far enough into the future, allows an organization to invest in all and only the facilities, equipment, materials, and staffing that it needs to most profitably fulfill that demand. The value of a good demand forecast is readily apparent, and we valiantly load demand history into our software and statistical models to start the forecasting process. The common characterization of demand becomes problematic, however, once we try to operationalize it (that is, when we start to describe the specific, systematic way to measure it). We need an operational definition to provide true demand history to our forecasting models and to measure the accuracy of our unconstrained demand forecast. We need to know what true demand really is, but soon realize that it may be unobservable.

The nonchalant use of demand will not work. We know orders, we know shipments, and we know sales. We know calls handled at call centers, transactions processed at retail stores, and hours billed by consultants. We can track inventory, out of stocks, fill rates, back-orders, and cancellations. We have all this data available to us, but none of it is the same as true demand. Consider the situation at a manufacturer:

If customers place orders to express their demand, and if the manufacturer services its customers perfectly by filling all orders in full and on time, then we have our operational definition. In this case, demand = orders = shipments. If both order and shipment data are readily available in the company's system, then we have the historical demand data, which we can use to feed our statistical forecasting models.

Unfortunately, few organizations service their customers perfectly. As such, orders are not a perfect reflection of true demand. This is because when the order fill rate is less than 100%, orders are subject to all kinds of gamesmanship. Here are three examples:

1. An unfilled order may be rolled ahead (carried over) to a future time bucket.
2. If shortages are anticipated, customers may inflate their orders to capture a larger share of an allocation.
3. If shortages are anticipated, customers may withhold orders or direct their demand to alternative products or suppliers.

In the first example, demand (the rolled ahead order) appears in a time bucket later than when it was really wanted by the customer. Rolling unfilled orders causes demand to be overstated—the orders appear in the original time bucket and again in future buckets until the demand is filled or the order is cancelled.

In the second example, the savvy customer (or sales rep) has advanced knowledge that a product is scarce and will be allocated. If the allocation is based on some criterion, such as fill all orders at 50%, the customer simply orders twice its true demand and then hopes to receive what it really wanted in the first place.

The third example not only contaminates the use of orders to reflect true demand, but it can also cause significant financial harm to your business. In a period of chronic supply shortages (due to either supply problems or much higher than anticipated demand), customers may simply go elsewhere. Customers may truly want your product (so there is real demand), but it won't be reflected in your historical data because no orders were placed. While orders are often perceived as equal to or greater than true demand, this third example shows that what is ordered may also be less than true demand.

As with orders, the use of shipments to represent demand has a number of potential problems. Shipments are often perceived as equal to or less than true demand. Thus, shipments and orders are thought to represent true demand's lower and upper bounds. But, as we see in example three, orders can be lower than the true demand. Furthermore, by example one, shipments can actually be greater than true demand in a particular time bucket. (This would occur when an unfilled order is rolled ahead into a future time bucket and then filled. In this situation the shipment occurs later than the true demand and inflates demand in the time bucket in which it is finally shipped.)

It should be noted that any operational definition of demand involving shipments should use gross shipments, not shipments net of returns. A return expresses an overstatement of demand in the period in which the returned product was originally shipped. We would need to attribute that return to some past time period and subtract it from demand in that period. But which past period? If a customer orders the product every week, we don't know in which week(s) they ordered too much. Alternatively, if we simply subtract the return from the period in which it is received, this does not accurately reflect true demand in that period. It is correct to adjust for returns for financial calculations (such as, net sales = gross sales − returns). However, such adjustments are not necessary or appropriate when we try to calculate true demand.[3]

To summarize, a suitable operational definition of demand may be unique to each organization and may be difficult to construct given the available data.[4] For a manufacturer, what a customer orders may not be the same as true demand (for the various reasons described above), nor is true demand what the manufacturer actually ships. At a retailer, what is actually sold off the shelves may not be the same as true demand, either. For example, customers may not be able to find what they want in the store (due to out-of-stocks or poor merchandise presentation), so there is true demand but no recorded sale. Determining true demand for a service can be equally vexing. I may wish to stay at a cheap hotel but have to upgrade when my preferred choice is sold out. Or I may call the cable company to complain about my television reception only to hang up in frustration trying to wade through their voice menu system.

As a practical matter, while we can't know exactly what true demand really is under most circumstances, we can often get close enough to make the concept useful in forecasting. It is true that shipments ≠ demand ≠ orders, yet if a manufacturer does a good job at filling orders (say 98%+), then shipments, orders, and true demand are virtually the same. Likewise, if a retailer's shelves are fully stocked (or nearly so), then ***point-of-sale (POS) data*** (that is, cash register receipts) may be an adequate representation of true demand (within a few percentage points).

Contrast imperfections in our true demand history with our real-life forecast errors, which are often 25%, 50%, or even much more. The fact that the demand history on which we build our unconstrained forecasts is not perfect (but may be off by a few percentage points from true demand) is inconsequential compared to the magnitude of the forecast error. The takeaway is this: Making heroic efforts to capture a perfect history of true demand is unlikely to result in significantly improved forecasts and is probably not worth the effort.

CONSTRAINED FORECAST

True unconstrained demand may be unobservable, but we can often construct some proxy that is close enough to be useful in driving our business. We begin the planning process, therefore, by compiling (our approximation of) true demand. The demand history is fed into statistical software and the unconstrained forecast is generated.

We prefer to start the planning process with an unconstrained forecast. This provides the supply chain with an unfettered prediction of what customers are going to want in the future and allows the organization to take actions to meet this demand. For example, if future demand is predicted to exceed the current supply capacity, the organization can hire workers, build new facilities, or outsource. (Alternatively, the organization could increase prices, drop customers, or eliminate sales channels to help reduce demand to levels it can fulfill.)

In order to measure the accuracy of our unconstrained forecast, we need to know the true demand that actually occurred, and herein lies a problem. Because of the unobservability of true demand, we cannot with certainty measure the performance of our unconstrained forecast. At best, we can measure its accuracy versus our proxy for true demand.

While the planning process begins with an unconstrained forecast, supply constraints are identified and communicated to the organization through a sales and operations planning (S&OP) or similar process.[5] The constrained forecast accounts for these supply limitations and indicates the expected shipments, or expected sales, or expected services to be provided. It represents the organization's best guess at what is really going to happen. It is appropriate to report performance of the constrained forecast rather than the unconstrained forecast. The constrained forecast can be evaluated against what really does happen. Unlike the nebulous measurement of true demand, we should be able to unambiguously measure what really does ship or sell, or the amount of services we provide.

As a final note, financial projections should always be made from the constrained forecast. (It makes no sense to project revenues for any unconstrained demand you know in advance you can't fulfill.) Any gap between the unconstrained and the constrained forecasts is also useful information for managing customer service. For example, when a shortage is anticipated, customers can be contacted and their demand redirected to a future date (when their demand can be fulfilled) or to alternative products. It is a failure of management to continue the solicitation of orders that are known in advance to be unfillable.

DEMAND VOLATILITY

Forecasting performance is traditionally evaluated by forecast error (or accuracy), and forecast ***bias*** (whether forecasts are chronically too high or too low). There are dozens of error and accuracy metrics available, but these don't tell the whole story. There are other less recognized metrics, such as the ***coefficient of variation (CV)***—the measure of a demand pattern's volatility—that have an important relationship to forecasting performance

The CV is computed over some time frame, such as the last year of weekly or monthly data, using this formula:

$$CV = \text{standard deviation}/\text{mean}$$

The CV indicates whether a demand pattern is smooth and stable, staying close to its average value, or is highly erratic. CV is a more useful measure of volatility than standard deviation alone, because it is scaled by the mean, and expressed as an easy to interpret percentage. Thus, while two demand patterns may each have a standard deviation of 100, their CV will differ based on their means. (If one has a mean of 1000 then its CV = 10%, while if the other has a mean of 100 then its CV = 100%.)

You observe high CV values when there is a lot of growth or decline in demand, when there is a lot of seasonality, when the demand is intermittent, or when there is a lot of randomness in the pattern. (In the coin-tossing contest above, CV was roughly 32%, 10%, and 3% for the processes P10, P100, and P1000, respectively.) CV is often in the 25% to 100% range for real-life demand at intermediate levels of aggregation (such as by item or by product group at a region). At more granular levels, such as by item at a store for a retailer, CV will often exceed 100%. (Note that CV can explode to huge values when the mean is close to zero.) Exhibit 1.2 provides a sample calculation of the CV for a demand pattern over a 13-week time frame.

What makes the study of demand pattern volatility so interesting and useful is that there tends to be a strong inverse relationship between the volatility of a demand pattern, and our ability to forecast it accurately. Exhibit 1.3 is a scatterplot of approximately 5,000 ***stock keeping units (SKUs)*** at a consumer products company (approximately 500 items sold through 10 distribution centers). The vertical axis shows

Week	Demand
1	75
2	93
3	105
4	110
5	89
6	95
7	120
8	101
9	88
10	79
11	105
12	134
13	153
Total	1347
Mean	103.6
Std Dev	21.0
CV	**20.3%**

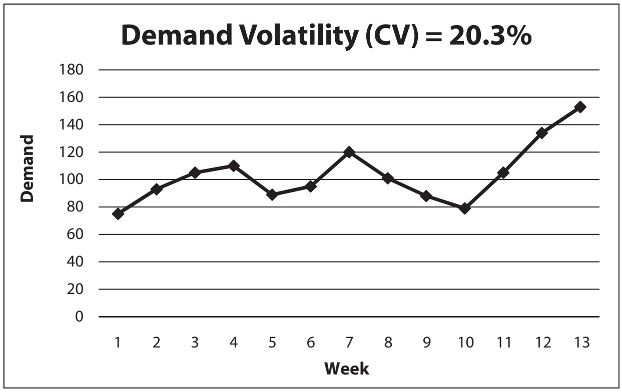

Exhibit 1.2 Coefficient of Variation of a Demand Pattern

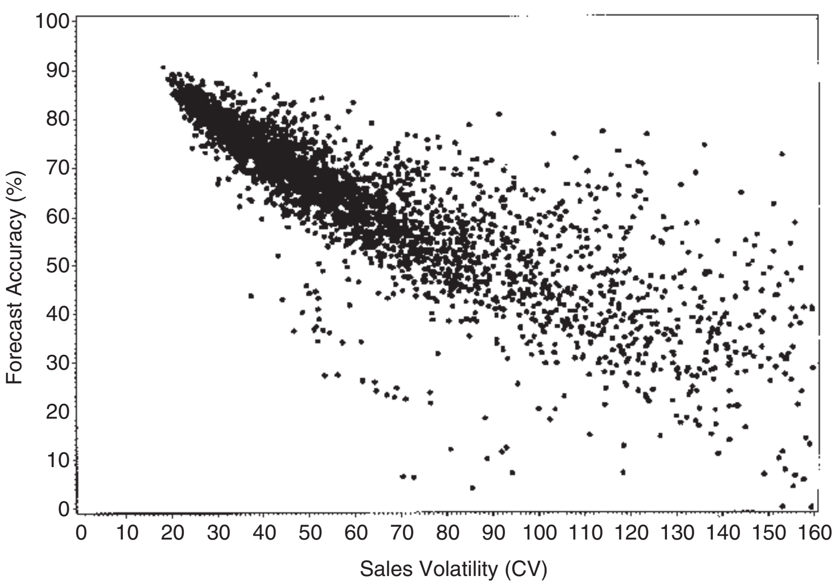

Exhibit 1.3 Forecast Accuracy versus Sales Volatility

forecast accuracy for each SKU, from 0% to 100%. Sales volatility, as expressed by the CV of ***actual*** sales for each SKU, is along the horizontal axis. The inverse relationship between volatility and forecast accuracy is apparent. (See Chapter 8, Practical First Steps, for detailed instructions on how to construct this scatterplot with your own data.)

There are several lessons to draw from this kind of volatility analysis. First, we must be clear that just because one product is less volatile than another, it doesn't guarantee we will be able to forecast it more accurately, but that is the general tendency. Second, our expectations for forecast accuracy must take into consideration the volatility of the behavior we are trying to forecast. Setting the same accuracy objective for every product being forecast is inappropriate and unfair, as this would fail to account for the underlying forecastability of the demand pattern.[6] Finally, volatility analysis suggests a novel way to improve forecasting performance—by smoothing demand! (Find more discussion of this approach in Chapter 6, Alternative Approaches to the Problems of Business Forecasting.)

INHERENT VOLATILITY AND ARTIFICIAL VOLATILITY

The notion of demand volatility can be decomposed into two elements called inherent volatility and artificial volatility. This decomposition is most meaningful at organizations that do not sell directly to the consumer of their products, but instead to distributors (or retail stores) who then sell to the final consumer:

Manufacturer → Distributor → Consumer

Most products and services have natural variation in consumption (more sales of sunscreen and lawn mowers in the summer; more gloves and snow boots in the winter; more service center calls during the day and evening than overnight). Although it is possible to impact natural consumption patterns through our sales and marketing activities, we often accept inherent volatility as a given.

Inherent volatility ≡ Variation in consumption

Inherent volatility is measured by the coefficient of variation of sales to the consumer (for example, POS data at a retail store). Contrast

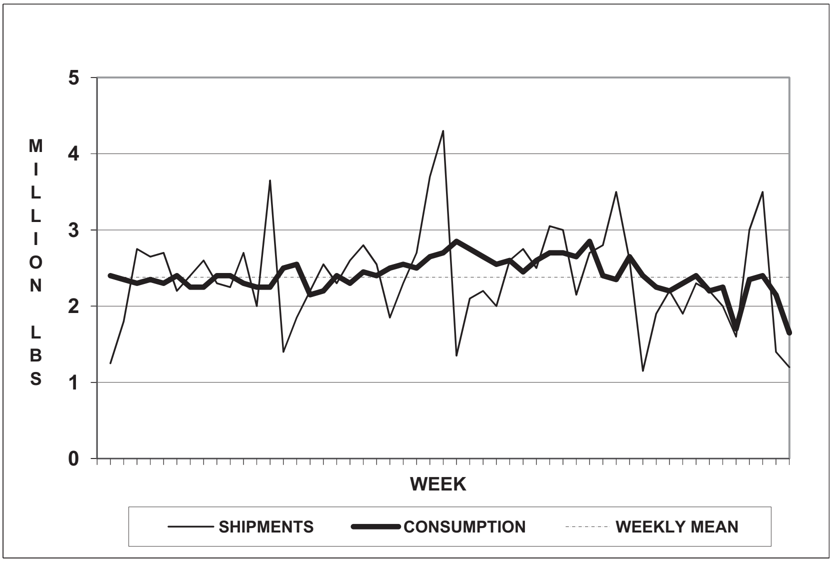

Exhibit 1.4 Shipments versus Consumption

this with shipment volatility, which is measured by the coefficient of variation of shipments from the manufacturer to the distributor or retail store. Shipment volatility is often much higher than inherent volatility, as illustrated by data at a consumer products company in Exhibit 1.4.

As seen by the thick line, weekly consumption (consumer purchases at retail) is very stable—a little higher in summer, a little lower in winter. In fact, the mean weekly sales would provide a very good forecast. The thin line, however, shows shipments from the manufacturer to the retailer, and this kind of pattern is very typical. There is a big push at the end of each quarter to ship enough goods to meet short-term financial objectives. This might involve a lot of promotions and price-cutting and other concessions to retailers, who have become skilled at playing the quarter-end waiting game. Each new quarter then starts anew with very little demand from the retailers, and very low shipments, simply repeating the vicious cycle. Despite the wide swings in shipments, consumption is little changed.

Artificial volatility is the difference between the volatility of shipments (to the distributor or retailer), and the inherent volatility of sales to consumers (by the distributor or retailer).

Week	Shipments	POS
1	70	105
2	80	90
3	105	100
4	130	95
5	90	110
6	115	105
7	120	85
8	70	95
9	110	100
10	80	115
11	75	105
12	130	95
13	125	100
MEAN	100.0	100.0
STD DEV	22.4	7.8
CV	22.4%	7.8%

Shipment Volatility	22.4%
Inherent Volatility	7.8%
Artificial Volatility	14.6%

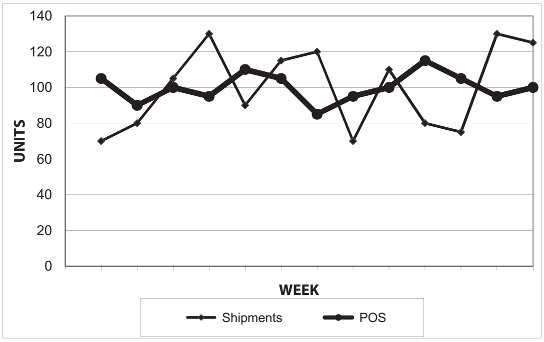

Exhibit 1.5 Calculating Inherent and Artificial Volatility

$$\text{Artificial volatility} \equiv \text{Variation due to organizational policies and practices}$$

Exhibit 1.5 illustrates the calculations.

Artificial volatility occurs when the manufacturer ships its products in patterns that are more erratic than the pattern in which consumers buy. While this kind of behavior is common, adding variation to a process rarely adds value or reduces costs.

Some types of businesses recognize the evils of volatility and proactively do things to spread demand more evenly. Industries with fixed capacity and little flexibility in changing the available supply of their product or service are most active in this approach. An example is electric utility companies charging higher rates during peak hours to shift demand to off-peak hours (and thereby avoiding the need to build costly new power plants). Similarly, airlines and hotels use revenue management techniques to constantly adjust prices (and maximize revenue), driving demand to flights or nights when there is excess availability, and away from those expected to be oversold.

EVILS OF VOLATILITY

Demand volatility can impact our ability to forecast accurately, often increasing costs and harming our customer service. Many organizational practices (such as promotional activities, sales contests, and the quarter end push) only serve to increase volatility. Typical practices are designed to produce record sales weeks rather than promote smooth, consistent, and more profitable growth. Consider this example of organizationally induced artificial volatility and the harm it caused at a consumer packaged goods company:

There was a long-running product that had very consistent (and predictable) sales over the past three years, and was very easy to manage. (Although not the actual data, Exhibits 1.6 and 1.7 provide an accurate representation of these events, illustrated with a mean of 100 units and a standard deviation of 10.) To increase the volume of this rather ordinary product, it was included in a major sales contest with significant incentives (for example, a trip to Hawaii) for the biggest selling sales representatives.

The sales contest was held and appeared to be wildly successful. In the final contest week, shipments were five standard deviations above the long-running

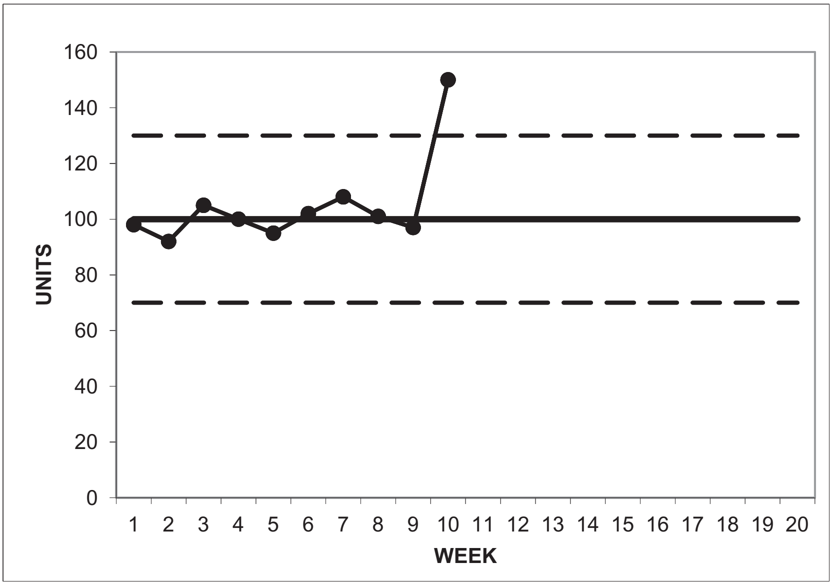

Exhibit 1.6 Sales Contest Shipments

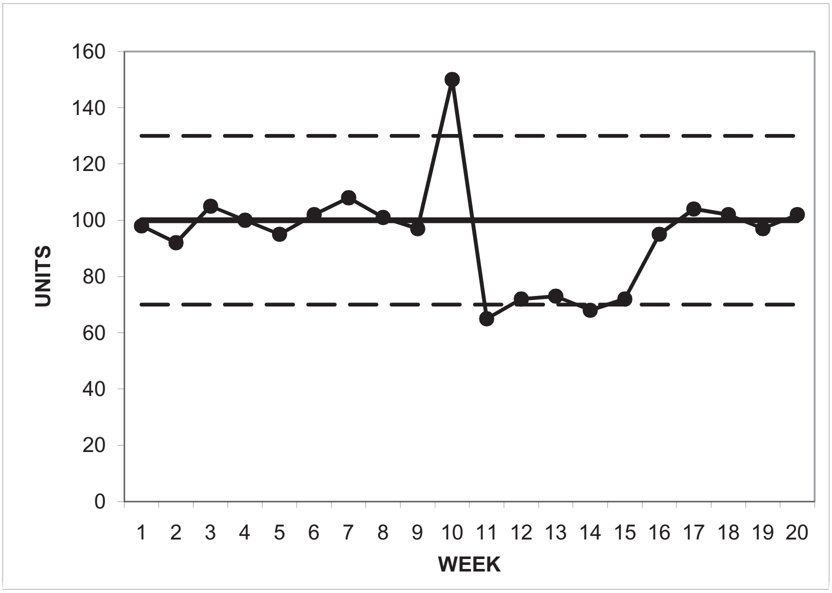

Exhibit 1.7 Post Contest Shipments

weekly average! Banners were hung proclaiming the grand event, the top sales reps were packed off on their award trips, and all involved were caught in the frenzy of self-congratulations.

However, upon further review:

Over the five weeks immediately after the sales contest, shipments averaged three standard deviations below the long-running weekly average. (This was not due to lack of inventory. There was plenty.) The net result was fewer units sold than what would have been expected by the long run weekly average for this product. Also, significantly less revenue was generated because of price incentives to encourage the record week.

To make matters worse, over one million units of excess inventory (worth over $1 million) were distressed—sold at a significantly reduced price or destroyed—because of its limited shelf life (these were processed food products). Grocery stores were in no hurry to replenish their stocks, having loaded up on the product at low prices during the contest. Tens of thousands of dollars were spent on celebrations and sales rep award trips that, in hindsight, encouraged absolutely the wrong kind of behavior!

Rather than creating costly incentives to spike demand, it may make more sense to design incentives that smooth demand. Smooth and stable growth can be managed profitably. Growth via boom and bust cycles is not necessarily best for an individual business, or for the economy as a whole.

The good news is that reducing the variability of a demand pattern is an almost sure way to improve the accuracy of our forecasts. Even better news is that reducing demand variability can give you better forecasts for free, without any new investment in forecasting process, systems, or people. Smoother demand can be achieved by reengineering or eliminating those policies and practices that encourage customers to order in spikes or erratic patterns. Management has control over the often misguided policies and practices that serve to increase volatility, and management can change them. One of the surest and cheapest ways to get better forecasts is to simply make the demand forecastable.

EVALUATING FORECAST PERFORMANCE

Forecasting performance should be evaluated by more than just accuracy or error. Traditional and commonly used forecasting performance metrics, such as the ubiquitous ***mean absolute percent error (MAPE)***, show the ultimate result of the forecasting process. MAPE tells you the magnitude of your forecast errors but gives no indication of how efficient the organization was in achieving that level of performance. Also, MAPE, by itself, does not tell you whether other methods would have been equally or more accurate, or whether you could have achieved similar results with less effort.

Due to company politics, personal agendas, lack of training and tools, wishful thinking, or sheer incompetence, many (if not most) management efforts fail to improve the forecast, or even make it worse! The application of traditional process performance metrics, such as MAPE, does not address this issue. By failing to consider the value added by each step and participant in the forecasting process—their contribution to improving the forecast—the traditional approach to performance measurement misses a source of significant process improvement.

An important, yet sometimes overlooked metric, is the bias in the forecast. Bias indicates whether the forecast is chronically too high or too low, and is often a sign of political influences that are contaminating the forecasting process. When forecasts are unbiased, an organization can operate effectively with less forecast accuracy because the positive and negative errors are balanced over time, tending to cancel each other out. When forecasts are biased too high, however, there may be unfavorable consequences in terms of inventory (carrying costs, and added risk of spoilage, loss, or obsolescence) or other expenditures (adding unneeded capacity or staffing). When forecasts are biased too low, there can be unfavorable impacts on customer service (failure to fill orders, extended call center wait times, and so on) and missed opportunity for revenue.

The most basic exercise is to compare the results of your forecasting process to the results you would have achieved using an alternative method. The fundamental standard for comparison is with a ***naïve forecast***—something simple and easy to compute with minimal effort. Common examples of naïve models include the random walk (using the last known actual as your future forecast), a seasonal random walk (such as using the corresponding period from last year as your forecast for this year), or a moving average. The difference between your process results and the results of a naïve method is the value added by your efforts.

Forecast value added (FVA) analysis is a method for evaluating each step and participant in the forecasting process (and will be thoroughly discussed in Chapter 4, Forecast Value Added Analysis). The objective is to identify and eliminate non-value adding activities from the process. Eliminating unnecessary steps and participants will make the forecasting process more efficient, delivering its results with fewer resources. Eliminating those activities that are actually making the forecast worse has the added benefit of improving forecast accuracy.

Looking solely at MAPE (or other traditional metric), by itself, can create the wrong impression. "If only we hired more analysts, if only we engaged more participants in the process, if only we had a bigger computer and more sophisticated software, then our forecasting problem would be solved," is a serious misconception. In reality, results are often improved by doing less, by doing less harm.

EMBARKING ON IMPROVEMENT

Our forecasts never seem to be as accurate as we would like them to be, or need them to be. As a result, there is a strong temptation to throw money at the problem in hopes of making it go away. There are plenty of consultants and software vendors willing to take that money in exchange for promises of improved forecasting performance, but these promises are often unfulfilled. Many organizations, perhaps even your own, have thrown thousands or even millions of dollars at the forecasting problem, only to end up with the same old lousy forecasts.

This chapter introduced fundamental, yet frequently overlooked, issues in the practice of business forecasting. Addressing these issues can lead to immediate improvement in forecasting performance, often without any new investment. Before embarking on improvement efforts, a few simple tests can help determine both opportunities for improvement and the organization's readiness to make the necessary changes:

- *Data/systems infrastructure.* Forecasting requires systems and data. There is no magic formula that solves all the forecasting problems, but well-designed software can help. Before implementing a software solution and expecting it to work, certain basic data elements are required. Can your information technology department provide a master file of items (with their attributes), a master file of customers (with their attributes), and a clean file of historical orders, shipments, forecasts, inventory, production, and POS (for consumer products companies)? If so, you have the minimum necessary data infrastructure in place.
- *Defining demand.* What are you trying to forecast (orders, shipments, unconstrained demand)? What historical data are you feeding into your forecasting models, and what actuals are you using to evaluate forecasting performance? Sometimes what we can measure (for example, orders or shipments) is not the same as what we are forecasting (for example, unconstrained demand). Make sure your organizational language is understood, operationally defined, and used appropriately.

- *Demand volatility.* Utilizing historical shipment and POS data, determine the inherent volatility of consumption and the artificial volatility caused by your organizational practices. (For example, measure the CV of POS and the CV of shipments over the past 52 weeks.) Significant artificial volatility indicates opportunities to smooth shipments, and thereby get better forecasts. This may call for reengineering practices that encourage volatility in shipments. Significant inherent volatility (in consumer demand) may prompt new sales, marketing, and financial practices to help coax that demand into more stable (and predictable) patterns.
- *Forecast value added.* For a quick analysis, compare the accuracy you achieved over the past year to the accuracy you would have achieved by just using a moving average. Many organizations find that the moving average would have done better! Such a finding can actually be good news, as it indicates your forecasting performance can be improved immediately by scrapping the current forecasting process and just using a moving average.

The fundamental problem of business forecasting was articulated by David Hume over 250 years ago, and that problem remains today. While we are much better at building models (that purport) to explain past observations, we still have no assurance that future behavior will follow the past, or that our forecasts will be any good. We must temper our expectations for forecast accuracy, recognize the uncertainty inherent in all prediction about the future, and address business forecasting in practical terms, not with wishful thinking.

But all is not lost. The practical approach to business forecasting is to recognize that highly accurate forecasts may not be possible. We, therefore, turn our attention away from generating the perfect forecast, to generating forecasts that are as good as anyone can reasonably expect them to be, and doing this efficiently. Though we cannot control the ultimate accuracy of our forecasts, we can control the processes and systems we use and the resources we invest in the forecasting effort. We should not squander resources in pursuit of unrealistic accuracy objectives that will never be achieved. Instead, we must prepare our organizations to deal with the uncertainty, not assume we can overcome it simply with forecasting.

While fundamental issues are being addressed, there are other common mistakes and bad practices that can stifle forecasting improvement efforts. The next two chapters explore several worst practices in business forecasting—practices that may be occurring at your organization—that need to be recognized and not repeated. FVA analysis comes into play throughout these chapters, as a tool for identifying activities that are failing to improve the forecast.

NOTES

1. Portions of Chapter 1 originally appeared in Michael Gilliland, "Fundamental Issues in Business Forecasting," *Journal of Business Forecasting*, Vol. 22, No. 2, Summer 2003, 7–13.
2. S. Makridakis and N. Taleb, "Living in a World of Low Levels of Predictability." *International Journal of Forecasting* 25 (2009), 840–844.
3. For more discussion of this topic, see Mark Chockalingam, "What Is True Demand?" *Demand Planning Newsletter*, April 2009.
4. More complicated (but not necessarily better) operational definitions of true demand can be constructed by some hybrid of orders and shipments. Examples include:

    1. Demand = (shipments + orders)/2
    2. Demand = shipments + incremental shortages
    3. Demand = shipments + latest shortages

    The first formula simply defines demand as halfway between orders and shipments. It assumes half of the shortages represent legitimate demand. If order is 120 and shipment is 100, then demand = 110.

    The second formula avoids over-counting repeat shortage rollovers by only adding increases in shortages to shipments. Thus, if the shortage in time period t is 20, and the shortage in period $t + 1$ is again 20, then demand = shipment for period $t + 1$ (the shortage amount, 20, did not increase from the prior time period). If the shortage in period $t + 2$ is 25, the demand in period $t + 2$ is shipment + 5 (because there was an incremental 5 units of shortages from 20 to 25).

    The third formula also avoids overcounting repeat shortages by including in demand only those shortages still showing at the end of the time bucket. In this case, the demand for a month will include all shipments of that month + unfilled orders of the last week only. If, for example, shortages in a four week month were 10, 20, 40, and 30, the total demand for the month would be shipments + 30 (adding the last week's shortages). The following table illustrates various demand definitions over a one-month period:

	Week				Month Total
	1	2	3	4	
Orders	50	50	60	60	220
Shipments	50	40	55	40	185
Shortages	0	10	5	20	35
Incremental shortage		10		15	25
Latest shortage				20	20

Demand = (shipments + orders)/2 = (185 + 220)/2 = 202.5
Demand = shipments + incremental shortages = (185 + 25) = 210
Demand = shipments + latest shortages = (185 + 20) = 205

Thanks to Jacqueline Lawrence, Derrick Yuen, and Joe Anderson (former colleagues at Sara Lee Intimate Apparel) for their suggestions on alternative definitions of demand.

5. For the standard guide to the sales and operations planning process, see Thomas Wallace and Robert Stahl, *Sales and Operations Planning: The How-To Handbook* (3rd. ed.). T.F. Wallace & Co., 2008. I would also recommend Wallace and Stahl's *Sales Forecasting: A New Approach*, T.F. Wallace & Co., 2002, for its practical guidance to real-life business forecasting.
6. See the Spring 2009 issue of *Foresight: The International Journal of Applied Forecasting*, for several articles on assessing forecastability. (Len Tashman, Ed., 2009, "Special Feature: Assessing Forecastability." *Foresight* 13, 23–45.)