DEMAND PLANNING AND FORECASTING WITH POS DATA: A CASE STUDY

By Fred Andres

The retailer is the final frontier of supply chain planning. So, it is important for manufacturers to have a serious look at what is happening at the retail outlet level because that is where there is an interface with a real customer. If your customer does not get the products it wants, in the package it wants, and in the stores where they are needed, then your entire supply chain has failed. The most common shortcoming of retail execution is high levels of stock outs. Numerous studies by organizations like the Grocery Manufacturers Association and the National Association of Convenience Stores over the past 10 years have shown that the average manufacturer faces an out of stock rate of 6% to 8%, a number that can reach as high as 12% during price promotions.

The data that is needed for supply chain planning at the retail level is the actual sales to consumers at each retail outlet, which is called Point of Sale (POS) data. Twenty years ago, collecting all the POS data a manufacturer needed was extremely expensive, because it meant getting a massive amount of data from hundreds and thousands of independent businesses. Many of these businesses still had old-fashioned cash registers and did not store their POS data in a database. However, there have been three significant changes over the past 20 years that have dramatically reduced the cost of collecting this data:

1. There has been a massive consolidation of retail stores into large chains. This change has been led by Wal-Mart, which now accounts for nearly 10% of all U.S. domestic retail sales. All the large chains consolidate the POS data from all their stores in a database at regional locations or at corporate headquarters. So it is now possible to get POS data for thousands of stores from one location. These large chains usually have an IT staff that can work with the manufacturers' IT staff to set up automated data exchange.
2. Data storage has become much less expensive and database engines have become powerful enough to move terabytes of data instead of megabytes of data.
3. Even most small retail outlets, like Mom and Pop Grocery Stores, now use scanners to record sales; with the result, POS data are stored in the database.

FRED ANDRES

Dr. Andres is president of PhredTek, Inc., a company with most of its clients in the beverage industry. At PhredTek, he provides logistics consulting services, market research, and software development. He spent 28 years in the Logistics and IT Divisions of Anheuser-Busch where he was responsible for building and operating several large-scale beverage sales forecasting systems. Prior to Anheuser-Busch, he worked in the Michigan Department of Commerce where he advised the Governor on energy policy. He also worked on large-scale economic models at the Lawrence Berkeley Laboratory and switching systems engineering at Bell Telephone Laboratories. He holds a Master's degree in Operations Research from Cornell, and a Ph.D. in Operations Research from the University of California at Berkeley.

Retail chains do not account for 100% of retail sales yet because there are many independent retail outlets. But as the use of POS data by manufacturers grows, it has become profitable for independent IT businesses such as Nielsen and IRI to build data marts with the POS data gathered from independent retailers and then sell to manufacturers at a price.

THE NATURE OF POS DATA

The basic element of POS data is an individual sales transaction. It contains:

- Universal product code or UPC
- Price
- The number of units per transaction (over 97% of transactions that I have seen involve a single unit; sometimes there is a transaction of multiple units)

  POS data is converted into revenue by multiplying the quantity of SKUs by their price.

The accuracy of each transaction is generally very high because the UPC is usually scanned from product itself and the prices are rigorously maintained by the retailers because their profit depends on accurate prices.

The most basic supply chain decision that a retailer or manufacturer would make using POS data is the quantity to deliver to a store on a specific day. Therefore, we add the data across all transactions for a day or for each hour. When doing so, it is important to remember that retailers can charge different prices to different customers, so not all transactions will have the same price. I have found that the best daily price to use in forecasting models is the weighted average price of all the day's transactions. This can be computed simply by adding the revenue of all the transactions and dividing it by the sum of units sold.

There are still some channels of retail sales that do not have good POS data because some of the products they sell are not easily classified by a product number such as UPC. An example of such a channel is a restaurant. Even when they sell an item that has been assigned a UPC number—like a bottle of beer—they generally serve it in another container (like a glass) that does not have the UPC printed on it so it cannot be scanned. From what I have seen of the POS data that restaurants have, it will be many years before they are capable of providing POS data sufficient for Supply Chain Planning.

THE POWER OF POS DATA

The real advantage of POS data over shipment data is that it is free of inventory decisions, because consumers rarely make explicit decisions to build or deplete their home inventory of a particular product. Inventory decisions can hide the underlying sales trend. For example, if sales to consumers are flat, but the retailer chooses to cut inventory over the history being used to build a forecast model, then sales will appear to be declining. As shown in Table 1, sales to consumer are flat at 30 per month. However shipments are declining by one per month and any reasonable forecast of shipments is going to continue the decline. By March of 2008, the forecast will be half the actual sales to consumer. This is very common with new products. Retailers initially build inventories to levels that are well above expected sales to make sure they do not run out of stock in the first week of sales. Then if actual sales are not depleting the inventory fast enough, they will order less and less until the inventory is more reasonable. Once they get the inventory down to the desired level, they will increase their orders to keep it at that level. However, if the supplier is basing its forecast on shipments, and producing to meet that forecast, it may not have sufficient quantities of that product available to cover the order.

TABLE 1
EFFECT OF SHIPMENT BASED FORECAST ON ENDING INVENTORY

Month	Beginning Inventory	Shipments	Sales to Consumer	Ending Inventory
1/2007	100	29	30	99
2/2007	99	28	30	97
3/2007	97	27	30	94
4/2007	94	26	30	90
5/2007	90	25	30	85
6/2007	85	24	30	79
7/2007	79	23	30	72
8/2007	72	22	30	64
9/2007	64	21	30	55
10/2007	55	20	30	45
11/2007	45	19	30	34
12/2007	34	18	30	22

Also, POS data responds immediately to changes in causal factors, such as price or weather. If, for example, there is a factor that causes a retailer's sales to increase this week, it will order more of that product next week. This property allows us to determine the parameters of these causal factors with a high level of statistical significance.

FORECASTING USING POS DATA

POS data is rich in information for building forecast models. We will demonstrate how we can build a good forecasting model with POS data. Figure 1 gives daily sales of a single brand and package of a common consumer product at one supermarket for 919 days.

We can make a few observations when looking at this chart. Sales appear to be seasonal, rising higher in summer than winter. If you look very carefully, you can see a strong repeating pattern by day of week. However, since there is so much data and the daily and weekly seasonal variations are so strong that they hide other potential of sales variation, it is difficult to draw any further conclusions just by observing the chart. So we need to take two additional steps:

1. We need to identify potential causal variables using our knowledge of the business.
2. We need powerful estimation software to find the best forecast model including the causal factors, time series effects, and correction for outliers.

First, we will identify possible causal variables. They are:

- **Price**. This is a product that is frequently promoted with discounts up to 20%. Sales go up significantly during promotion times, which usually last one or two weeks.
- **Temperature**. Consumption of this product increases in warm weather. It has been observed that when the daily high temperature is over 65 degrees

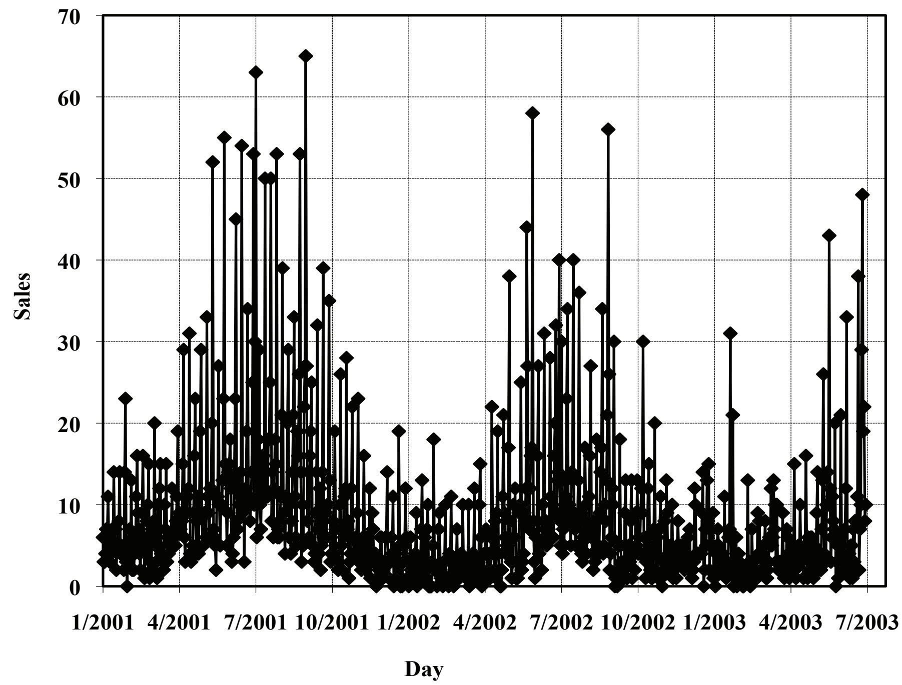
Fahrenheit, sales increase in proportion to the high temperature minus 65. However, when the temperature is below 65, sales do not change with temperature. This is very similar to cooling degree days used by the utilities to explain air conditioning load.

- **Holidays**. In general, sales increase on selected holidays. They are represented by dummy variables that equal 1 on the day of the holiday and equal 0 on all other days. Sometimes sales can increase up to six days before a holiday and possibly decrease up to six days after a holiday. Important holidays include: New Years Day, Memorial Day, July 4th, Labor Day, and Christmas.
- **Super Bowl**. There is generally a significant increase in consumption of this product while people are watching the Super Bowl. Its impact on sales is also measured by dummy variables—which equal one for Super Bowl Sunday and zero for all other days.

I chose Autobox to build the forecasting model because it automatically finds the best model including the causal variables, time series effects, and adjustments for outliers. Further, it automatically determines the important leads and lags of causal variables in conjunction with holidays. One needs only to provide it with the data. The program goes through numerous iterations until it finds one for which all parameters are statistically significant. Furthermore, it estimates the entire model at once including all the causal, time series, and outlier effects. The Autobox also revealed that sales are mostly higher than average during summer.

Outliers (unusual values) can hide the effect of certain factors, and thus they have to be taken care of before doing any analysis. Outliers include pulses or one-time incident of unusual value, level changes (either up or down), seasonality, and trends (a trend that is either up or down starts at some point in time and continue into the future). To do this type of analysis, one needs the highly granular data that is provided only by daily POS data.

FIGURE 1
DAILY SALES OF A SINGLE BRAND AND PACKAGE OF A CONSUMER PRODUCT AT ONE SUPERMARKET



USING FORECASTS FOR STORE REPLENISHMENT

Good forecasts alone are not enough. We also need a robust program for creating orders for store replenishment for optimal results. Here are the steps to be used for generating replenishment orders:

Step 1: Forecast the daily sales of all of your products and packages until at least the second expected delivery date. If today is Monday, and you have deliveries on Tuesday and Friday, you need a daily forecast for at least today through Friday. This demonstrates the need for a daily forecast. The second delivery on Friday generally faces higher sales levels on Saturday through Tuesday than the Tuesday delivery. Also, delivery schedules may change requiring a different partition of the week. For very high volume items, daily deliveries may be appropriate requiring daily forecasts.

Step 2: Compute a safety stock using the following formula:

Safety Stock = Protection Factor × Standard Deviation of Error × Square Root of Number of Days until Second Delivery

Let us assume:

i. The probability of stock out is .05. If the error is normally distributed, then based on the normal probability function, you would need a safety stock of 1.61 standard deviation of forecast error. The 1.61 would be your protection factor.

ii. The standard deviation of error is 4.02.

iii. Today is Monday, the next delivery day is tomorrow, and the subsequent delivery day is Friday. You want to be 95% confident that if the Friday delivery arrives late in the day, you will still not run out of product from the Tuesday delivery. If today is Monday, when you are placing the order, you have five days of forecast error to cover before the Friday delivery. The standard deviation of the five days is the square root of 5 or 2.24 times the daily standard deviation.

In that case, safety stock will come to 14.50 (1.61 × 4.02 × 2.24).

Step 3: Calculate the order quantity, which will be safety stock plus the sum of the daily forecasts Monday to Friday minus the current inventory. You can find the inventory by counting it every time an order is placed. But this is very time consuming and I've found it to be highly error prone. A better approach may be to do a thorough inventory of all items in the store, say, once in a calendar quarter. Then compute current inventory by adding all the deliveries to the store since then and subtracting the sales to consumer (POS sales). This is called a perpetual inventory. There are also other schemes for improving the perpetual inventory between physical counts.

IS THIS PRACTICAL?

The number of SKUs in any practical application with this kind of methodology is going to be in the thousands, tens of thousands, or even hundreds of thousands. Consider that you are a manufacturer with an average of 10 items in a store and you sell your products at 10,000 stores. This means that you have to deal with 100,000 SKUs. Furthermore, the process requires close collaboration among as many as three independent business entities—the manufacturer, the distributor, and the retailer. I know that this kind of system is practical because I built one for a major manufacturer of retail goods in 2002. The manufacturer had a strict three-tier distribution system so all three business entities were involved. However, since the retailers were major chains, we were able to arrange collaboration with thousands of stores by working with their corporate headquarters. The core forecasting engine was Autobox so we had models of the above quality for every SKU. We implemented the system on a blade server so the capacity for adding new SKUs was virtually unlimited. We simply added blades as we added SKUs. Processing time did not change because the blades operated in parallel.

We were able to convince the chains that the system would benefit them by reducing their stock outs. As a result, each of the major chains agreed to automatically send the daily POS data from each of their stores by 8:00 AM the next day. We automatically stored this information in a database along with other causal factors such as temperature and holidays.

Although building such models is quite complex, you can get forecasts virtually instantaneously once your model is built. To improve the quality of forecasts, we generated new forecasting models for each SKU about once a month. This was done overnight when the real time load on the system was low. The models for each SKU were stored and used for forecasting each day until new models were built. The distributor's sales representatives would physically go to a store to place the order. They had a hand-held computer, which would establish a wireless connection over the Internet to the forecasting system, located at the manufacturer's headquarters. First, the hand-held computer would send the delivery schedule for that store to the system. Then, the system would retrieve the forecasts for the manufacturer's items in that store, compute the safety stock based on the delivery schedule, add to it the forecast, subtract a perpetual inventory, and send the resulting order back to the hand-held computer. The entire communications session lasted less than 2 minutes, sometimes even less than 20 seconds. At its peak, the system served 1,200 stores owned by five major chains, serviced by 30 distributors with a total of 40,000 SKUs. That means that we had to generate 40,000 models.

Solving the technical issues proved to be easy with Autobox, blade servers, and the Internet, but it was far more difficult to maintain the necessary collaboration between the three business entities. By 2005, many distributors decided that the incredible detailed information provided by the system exposed faults in their business practices that they would rather not let the manufacturer know. There was also a fear by some distributors that the system was the first step of a plan by the manufacturer to take them over. As a result, many distributors said that they were no longer going to use the system even though it worked perfectly. It soon reached a point where only two distributors were using the system; thus, it became impractical for the manufacturer to continue operating it.

RETAIL DISTRIBUTION CENTER REPLENISHMENT

The store replenishment process described earlier works for products that are delivered directly to individual stores by a manufacturer or distributor or for retail stores that are not part of a chain. Most products, however, are delivered to chain stores through the retailers' regional distribution centers. Consider the following extension to the above process:

i. Create orders for individual stores using the above process a day in advance of the next planned delivery.
ii. Consolidate all orders of one of the retailer's distribution centers at the nearest manufacturer's distributor or warehouse.
iii. Stage each store's order on a separate pallet that evening. Load all of the pallets for the one retailer's distribution centers (DCs) on a truck. Many of DCs of a large retailer like Wal-Mart will have sufficient demand for many manufacturers to support a full truck load daily.
iv. Schedule the departure of a truck so it arrives at the retailer's DC just before daily deliveries are made to individual stores. The pallets can be unloaded one at a time and directly transferred to a waiting truck going to an individual store. This way product will never touch the floor of a retailer's DC. This is called cross docking.

Such a process requires very close collaboration between the manufacturer, distributor, and retailer. All have to collaborate on schedules, data, and forecasts. However, it has significant

(Continued on page 38)

months in the same way. Notice that each ending inventory figure translates back to eight weeks of projected inventory based on the consumption of the POS forecast.

What happens if, after all this work, you miss your demand forecast? Other than the obvious impact on your forecast error, the chances are you will be asked to explain the variance and how the forecast was derived. Tying your forecast with the inventory your customer has agreed to maintain will make it easier to convince your management team that your forecast was developed with sound assumptions.

Finally, a word about third party POS data consolidation software. There are tools available in the market that can help in bringing disparate data together—normalizing dates (problem mentioned earlier), aggregating data at different levels, and enabling retailer tools to interface with your ERP or forecasting system. However, there are a few problems, particularly the cost of implementing such tools and the fact that you are limited to a stock reporting system that may not exactly match your needs. If a tool or system does not meet your requirements, you may wind up spending extra time and money to re-configure it.

CONCLUSION

Although some of the processes and activities that have been described here may not be the same at your company, they can be adapted with some modifications. One thing is clear: POS is a critical element in the development of any demand forecast. Not only does it provide a basis for the future and rationale for your forecast numbers but it also helps to bring order to a chaos in the demand planning process. ■

(RShapiro@jardencs.com)

TABLE 1
CALCULATION OF POS FORECASTS

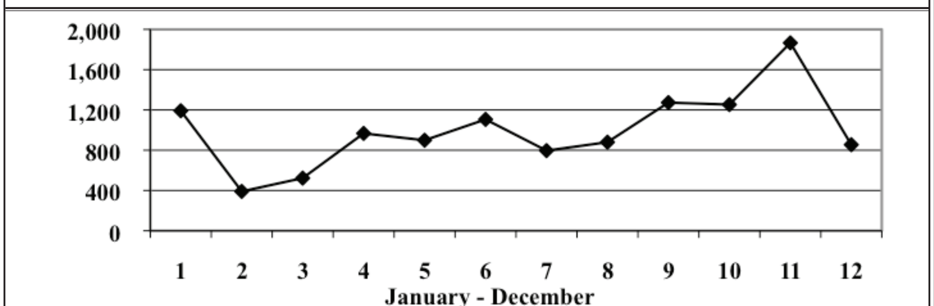
Period	Seasonal Index	POS Forecast
January 2009	119.20	1,192
February 2009	39.10	391
March 2009	52.30	523
April 2009	96.70	967
May 2009	89.90	899
June 2009	110.60	1,106
July 2009	79.60	796
August 2009	88.00	880
September 2009	127.30	1,273
October 2009	125.30	1,253
November 2009	186.60	1,866
December 2009	85.50	855

TABLE 2
HYPOTHETICAL EXAMPLE OF DETERMINING ENDING INVENTORY

Status	January	February	March	April	May
Demand Forecast	4,000	5,000	6,000	8,000	8,000
POS Forecast	2,000	4,000	4,000	6,000	6,000
Ending Inventory	8,000	10,000	12,000	14,000	16,000
Weeks of Supply	8	8	8	8	8

Notes: 1. Beginning inventory = 6,000.
2. Goal = Maintain 8 weeks of inventory.
3. Assumes 4 weeks in a month

FIGURE 1
SEASONAL INDICES



DEMAND PLANNING AND FORECASTING WITH POS...

(Continued from page 32)

cost advantages for all. Manufacturers and distributors are now delivering exclusively in full truckloads using full-size semi-tractor trailers. That can easily be a 50% savings over delivering in partial loads or smaller, less cost-efficient trucks. It eliminates all of the manufacturer's products from the retailer's DC. Instead, they go directly from truck to truck without ever hitting the floor. Holding and warehouse operating cost reductions can be significant. If enough manufacturers implement such a process, retailers could even decrease the size of their DC. Finally close attention to store level details when determining deliveries to the warehouse will reduce stock outs, which would mean increase in sales. This is good news for everyone. ■

(phredandres@earthlink.net)