

Labor Scheduling, Part 1

Forecasting

Demand

by Gary M. Thompson

Labor scheduling involves balancing customer demands, employee work requests, and profitability—and it sure ain't easy.

Cost-effective labor scheduling is one of the most important tasks that front-line managers perform in service organizations. At any given moment, having too few employees—or enough employees but not those who have the necessary skills—can result in poor customer service; frustrated, overworked employees; and lost sales. On the other hand, having too many employees either (1) reduces operating margins, if extra hours are scheduled, or (2) results in low employee morale if employees work fewer hours than they desire because the available work is spread thinly among many employees.

Labor or workforce scheduling is the act of balancing customer demand, employee work requests, and profitability. Hospitality businesses

are typical in that their labor costs constitute a large portion of the costs under managers' control. Controlling costs through effective labor scheduling can be a challenge, particularly since no two employees have exactly the same skills or desire to work the same number of hours, and the manager must also heed government regulations, company policies, and contractual obligations.

Gary M. Thompson, Ph.D., an associate professor of operations management at the Cornell University School of Hotel Administration, has written a four-part series on labor scheduling, scheduled for publication in *Cornell Quarterly*, of which this is Part 1.

© 1998, Cornell University

In earlier publications I characterized workforce scheduling as comprising four tasks.¹ Briefly, those are: forecasting customer demand, translating those demand forecasts into employee requirements, scheduling employees, and fine-tuning the schedule in real time. The first task is to predict customer demand for your service. This initial step involves forecasting characteristics of the service transaction that change over time, such as customer-arrival rates. The second task is to calculate the number of employee hours required to satisfy the customer demand predicted in step one. In other words, step two requires setting the number and skill levels of employees needed to serve customers adequately during some time period. The third task is to develop the actual work schedule by taking into account employees' skills, desires, and requests, and then deciding who will do what work at what time. The final task involves changing the work arrangement as required by actual demand. This final step ensures effective service.

This article is the first in a four-part series to be published in *Cornell Quarterly* that will focus on those four steps of workforce scheduling that I just outlined. The current paper describes the steps one might use to forecast demand.

Forecasting Demand

My approach comprises eight steps to forecasting customer demand. Those are: (1) determine the nature of the work; (2) identify those factors that generate the work (i.e., the labor drivers); (3) determine whether the key labor drivers vary

over a short time period (i.e., are time-variant or time-invariant); (4) determine the time interval for tracking the time-variant labor drivers (e.g., 15 minutes, an hour, or an eight-hour work shift); (5) forecast the time-variant labor drivers;² (6) reduce the period-to-period variability of the time-variant labor drivers by smoothing; (7) check the accuracy of forecasts through measurement and tracking; and (8) define the time period during which the work can be actually performed (i.e., the work window). Those eight steps are described in detail below.

Step 1— Determine the Nature of the Work

Hotel managers' labor-scheduling task involves taking into account work that can be performed at almost any time (e.g., vacuuming public-area carpets) and work that must be performed on demand (e.g., filling a guest's room-service order). I coined the terms "controllable work" to describe those tasks that have substantial time flexibility, and "uncontrollable work" to describe those tasks that have little or no timing flexibility (for example, when customers and employees interact).³

Much of a hospitality firm's work is delivering service, which generally involves uncontrollable tasks that can only be done when customers are present. In a full-service restaurant, for example, uncontrollable work would include the interaction of the wait staff with customers and the preparation of the customer orders by the kitchen staff.

² How to forecast the work to be generated by the time-variant labor drivers will be covered in Part 2 of this series on labor scheduling, scheduled for publication in the December 1998 issue of *Cornell Hotel and Restaurant Administration Quarterly*.

³ Gary M. Thompson, "Improving the Utilization of Front-Line Service Delivery System Personnel," *Decision Sciences*, Vol. 23 (September-October 1992), pp. 1072-1098.

In contrast to uncontrollable work, controllable work affords some latitude (called a window) in when it can be performed—for instance, guest-room preparation. The timing of that task is flexible because it can be done anytime between when the room is vacated (window opens) until just prior to its being reoccupied (window closes). In a full-service restaurant, controllable work includes washing dishes, stocking shelves, and folding napkins.

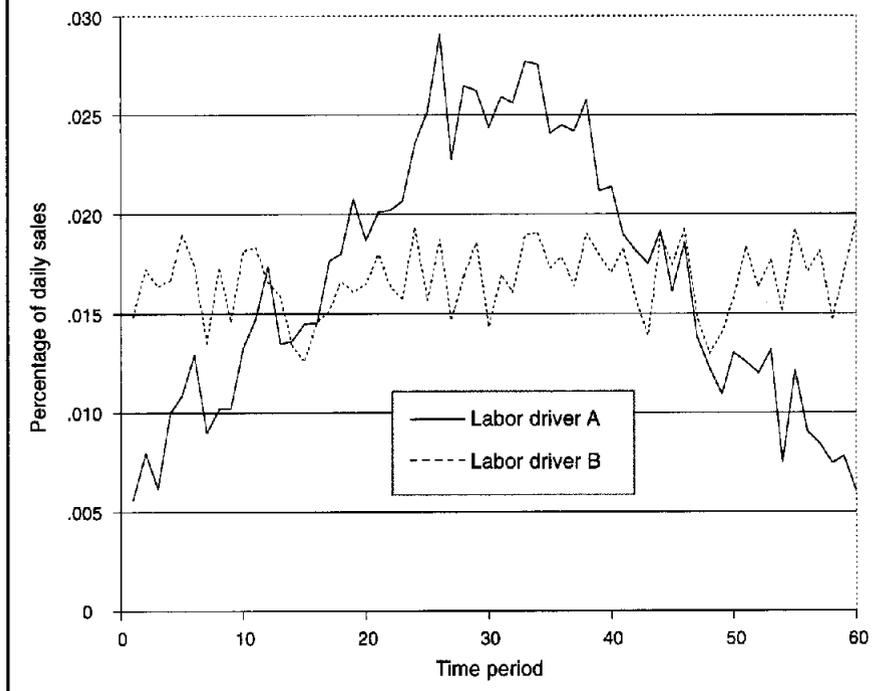
Controllable-work windows vary in length depending on the task. For example, preparing a stay-over guest's hotel room might offer just an eight-hour window, say, from 8:00 AM through 4:00 PM. In contrast, a room that turns over and is not immediately filled may have a much longer preparation window, perhaps stretching across several days, depending on the next guest's arrival.

If a job has any controllable characteristics, I treat the entire task as controllable. Some tasks that usually are controllable can become uncontrollable in some circumstances, particularly when full customer capacity is reached. Such uncontrollable work situations constitute a special case of controllable work that has a performance window of zero length. For example, consider what happens to housekeeping in a hotel having no defined check-in and check-out times as room occupancy approaches 100 percent. In that case, the housekeepers have essentially no latitude regarding when the guest rooms can be prepared. They must prepare each room immediately upon the departure of a guest so that it can be offered immediately to the next arriving guest.

Managers' scheduling task involves forecasting both uncontrollable and controllable work. For uncontrollable work, managers need

¹ See: Gary M. Thompson, "Labor Scheduling Using NPV Estimates of the Marginal Benefit of Additional Labor Capacity," *Journal of Operations Management*, Vol. 13 (1995), pp. 67-86; and Gary M. Thompson, "Assigning Telephone Operators to Shifts at New Brunswick Telephone Company," *Interfaces*, Vol. 13, No. 4 (July-August 1997), pp. 1-11.

Exhibit 1
Two labor drivers plotted by daily period



to predict the volume of work likely to be generated by the various labor drivers. Such predictions or forecasts should be expressed in terms of whatever labor-scheduling intervals are used by the organization (e.g., 15- or 30-minute work periods, also called planning intervals or planning periods) for some specified time into the future (e.g., typically a week or month, and commonly called the scheduling horizon). Managers also need to forecast the level of controllable work that must be performed during the same planning periods. Because the timing of that type of work is by definition flexible, however, they need not define specifically when it must be performed. Instead, managers must only identify the earliest and latest times when the work can be performed.

The next six sections describe the steps common to both controllable and uncontrollable work. The

last step, defining the allowable work window, applies only to controllable work.

Step 2— Identify the Independent Labor Drivers

The second step in forecasting demand is to identify those variables that will affect the number and skills of employees needed to perform or deliver the service. Consider housekeeping in a hotel, for example. What affects how long a housekeeper requires to clean and prepare a room? By answering that question, one has identified the labor drivers for room attendants.

Work-measurement studies may help a manager to determine how service characteristics affect the duration and nature of the service transaction. To conduct work-measurement studies, managers and employees (and oftentimes consultants) first brainstorm the characteristics of a job that could affect its

duration. Next, one measures and itemizes service interactions' actual duration and characteristics. Finally, the investigator determines the precise effect of each characteristic on the service transaction. Multiple regression can be helpful in this regard.

When identifying labor drivers, it will simplify things to choose measures that are independent of each other. In a quick-service restaurant, for example, both the number of items per order and the order value (i.e., check size) can be thought of as labor drivers. However, the order value will likely be related, or correlated, with the number of items per order. Thus, it would be better to select one of those drivers, but not both. Since one must identify and then forecast the effect of each labor driver, it is preferable to be parsimonious when selecting relevant drivers. One can use a correlation matrix to help identify relationships between labor drivers. (A correlation matrix lists the correlations between all pairs of potential drivers. Some spreadsheet software programs have a built-in function to generate correlation matrices.) Any labor drivers with correlations below 0.5 can be assumed to be independent.

Having identified the set of independent labor drivers, the next step is to determine whether the drivers are time-variant or time-invariant.

Step 3— Determine Whether the Labor Drivers Are Time-Variant or Time-Invariant

Examine each labor driver to determine whether its effects will vary over the course of the planning horizon. The effects of a time-variant labor driver will change over the duration of the planning horizon, while the effects of a time-invariant labor driver will remain constant. In the long run every driver is time variant (because, for

instance, technology improvements may introduce new efficiencies), but our focus is on just the scheduling horizon, typically somewhere between a week and three months. With a short work-schedule time horizon, many labor drivers will be essentially constant.

Identifying the time-variant labor drivers is important because of how one must forecast those labor drivers. Perhaps the easiest way to distinguish between time-variant and time-invariant labor drivers is to track the driver over time.⁴ This can be visualized by drawing a graph and charting the driver's effect on the "y" axis and time on the "x" axis. A time-variant driver's chart will show cyclical changes over time, while a time-invariant driver will remain relatively constant or exhibit random variation.

As an example, again consider housekeeping. If one plots the cleaning time per stay-over room, one should see a relatively constant relationship over a period of several weeks, suggesting a time-invariant labor driver (i.e., each stay-over room requires *X* minutes to clean). A consistent time to clean each room does not imply that the requirement for housekeepers also is constant, however. The key time-variant driver for housekeepers is the *number* of rooms to be processed across the scheduling horizon. Note, then, that different labor drivers for a particular job can be either time-variant (e.g., number of rooms) or time-invariant (e.g., effort per room).

Exhibit 1 shows a plot of two labor drivers by daily planning period. Both drivers exhibit random variation from period to period.

⁴ Ideally, managers will have a record of each labor-driver's effects so that those effects can be plotted as a direct function of time. If not, the drivers can be treated as constants at first (i.e., as time-invariant), although data should be collected to determine the drivers' actual effect as either time-variant or -invariant.

Exhibit 2

Labor-driver data lagged by one period

Mean check values:

\$13.28	\$15.41	\$14.62	\$14.93	\$17.03	\$15.56	\$12.10	\$15.51	
	\$13.28	\$15.41	\$14.62	\$14.93	\$17.03	\$15.56	\$12.10	\$15.51

Top row: Original data over eight hours, presented in 15-minute intervals.

Bottom row: Same data lagged by one period (as explained in the text).

The correlation of the shaded data is -0.201, which indicates that this particular driver is time invariant. (If the driver were time variant, one would expect to see a correlation of approximately 0.5 or higher.)

Driver B appears to have a stable mean over the day, while Driver A's mean appears to vary over time. Based on the picture recorded in Exhibit 1, one would likely decide that Driver A is time-variant while Driver B is time-invariant.

A correlation analysis can help distinguish between time-variant and time-invariant drivers. The correlation should be measured with the data lagged by one period (to compare each period with the period before it). For example, say that over a two-hour interval, broken into eight 15-minute periods, a manager observes the following mean check values in a restaurant: \$13.28, \$15.41, \$14.62, \$14.93, \$17.03, \$15.56, \$12.10, and \$15.51. She should then set up the data as shown in Exhibit 2 and measure the correlation of the shaded data.⁵ The correlation of these data is -0.201, which indicates that this particular driver is time invariant. If the driver was time variant, one would expect to see a correlation of 0.5 or higher.

Lagged-data correlation analyses of the labor drivers illustrated in Exhibit 1 gives a correlation of 0.927 for Driver A and 0.101 for Driver B. Those results suggest that Labor Driver A is time-variant

⁵ If, for example, the data are in an Excel spreadsheet, using columns A through I and rows 1 and 2, the correlation could be found using the formula «=CORREL(B1:H1,B2:H2)».

But be careful...

Tracking Labor Drivers Using Information Systems

When tracking labor drivers, one should focus on when the driver actually occurs. Information systems do not necessarily provide the correct information. For example, in a quick-service restaurant, the point-of-sale system (POS) records when service starts and finishes for each customer. As such, one could be tempted to use the start of service as the indicator of when labor was necessary. However, usually there is a lag between when customers arrive at the restaurant and when they are served. That lag, of course, is called waiting in line. The longer customers wait for service, the longer the lag. Plus, that waiting time will not be consistent across the operating day. Thus, the POS will provide misleading information. Using the POS data to define the service demand can, in effect, perpetuate poor service. To correct the problem, a manager would need to supplement the POS information with other data—for example, the number of customers in line or customers' queue time during each period of the day. Knowing that will help the manager estimate when customers actually arrive (versus when they are served).—G.M.T.

while Driver B is time-invariant. Again, the distinction between time-invariant and time-variant drivers is important because we want to be able to forecast the time-variant drivers, as described in the next section.

Step 4— Determine the Time Interval for Tracking Time-Variant Labor Drivers

Once the time-variant labor drivers have been identified, an appropriate time interval for tracking those labor drivers can be determined.

The best way I know to illustrate the process of selecting an appropriate time interval is to use an example. Exhibit 3 shows four sets of data for a single labor driver, plotted at four different time intervals, ranging from 15 minutes to 450 minutes. Looking at Exhibit 3, the longest data-collection interval of 450 minutes (7.5 hours) clearly provides a misleading view of the labor driver. In fact, at 450-minute intervals, the flatness of the plotted line suggests that the labor driver is time-invariant. Both the 150-minute and 45-minute intervals in this example also are too long, since by comparing those data to the 15-minute-interval data it's clear that there are periods where the 150-minute and 45-minute intervals over- and under-estimate the level of the driver. In this case, then, using those intervals to forecast demand would not give the best results.

In an unpublished study I found that 15-minute intervals work well for tracking time-variant labor drivers, though periods as short as 5 minutes can be effective when labor drivers are undergoing rapid change (for example, when considering the period between 11:00 AM and 1:00 PM in a quick-service restaurant). Moreover, 15-minute intervals are convenient. They are commonly used anyway when developing work schedules, since employees' rest

breaks often are 15 minutes long. On the other hand, tracking time-variant labor drivers using different time intervals is acceptable.

• • •

Once the four steps outlined above have been completed, they need be updated only periodically. Repeating those steps every six months to a year should be adequate. (One does not need to re-evaluate the chosen time intervals every week, for example.) In contrast, the following four steps should be performed every time a new schedule is developed.

Step 5— Forecast the Work Generated by the Time-Variant Labor Drivers

Step 5 involves forecasting the work to be generated by the time-variant labor drivers. To do this, one needs to forecast the level of each time-variant labor driver for every time interval (e.g., every 15 minutes) for the entire work schedule (e.g., for a week or month). There are two ways to make such forecasts: the labor driver can be forecast in each period independently, or the labor driver can be forecast using an aggregation-disaggregation approach. Again, allow me to explain by example.

Assume that a manager wishes to develop a workforce schedule for the coming week, broken into 15-minute intervals, and that the facility operates 12 hours per day. Also assume that the key labor driver is the number of customers served and that the workload per customer is time-invariant. The manager, then, must develop forecasts of demand for 336 planning periods (7 days × 12 hours per day × 4 periods per hour = 336). A manager using the independent approach would develop the forecasts independently for each of the 336 periods. A forecast for period X, then, would be based on the customer demand observed in the same period during

some previous time period, say, six months. The independent approach thus requires considerable calculation to generate a large number of forecasts. It also assumes that each period is independent of every other period, which is not true in most hospitality operations. If a hospitality firm is busier than average between 10:00 and 10:15 on a Monday morning, say, it is probably also busier than average between 10:15 and 10:30. Thus, a better approach to forecasting demand is frequently the aggregation-disaggregation approach.

The aggregation-disaggregation approach takes advantage of consistency in the labor driver. It first combines demand data across all the planning periods (in this instance, 15-minute intervals). For example, the manager can combine all the historical Monday demand data and measure how each 15-minute planning period compares to Monday's total business. He can then forecast demand for Monday as a whole (one forecast for each day of the week) and then separate the total day's demand into the demand for each individual period, using historical data.

One can apply the aggregation-disaggregation approach to longer planning periods as well. For example, the manager can combine the daily demands to obtain a measure that can be used to forecast the weekly demand (one forecast). He would then break down the weekly forecast into the demand for specific days, which in turn would be broken down into demand forecasts for the individual 15-minute planning periods.

For aggregation-disaggregation to work there must be a degree of consistency in the data. For example, Monday must consistently be the Xth busiest day of the week. Consider the data table in Exhibit 4, for example. Those sample data were

Exhibit 3
A labor driver plotted according to different data-collection intervals

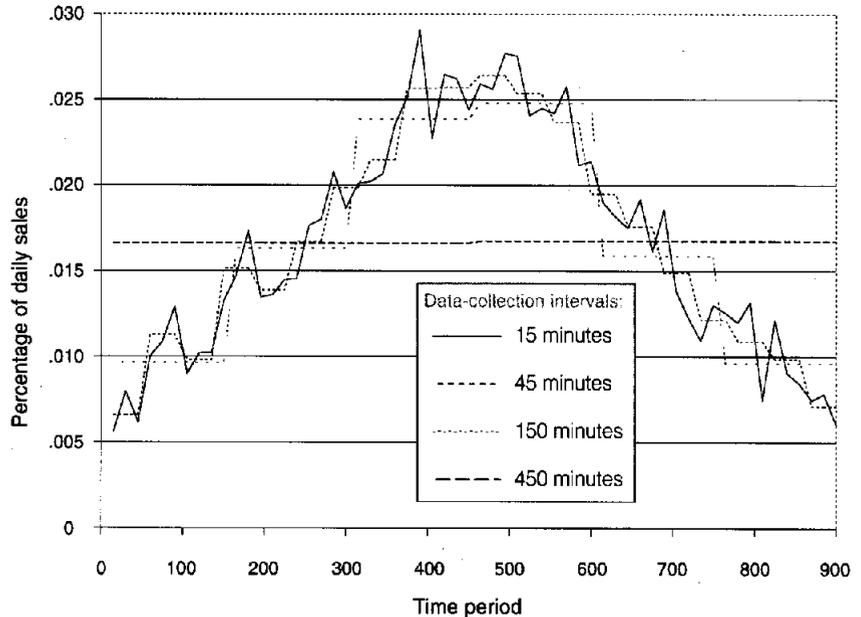
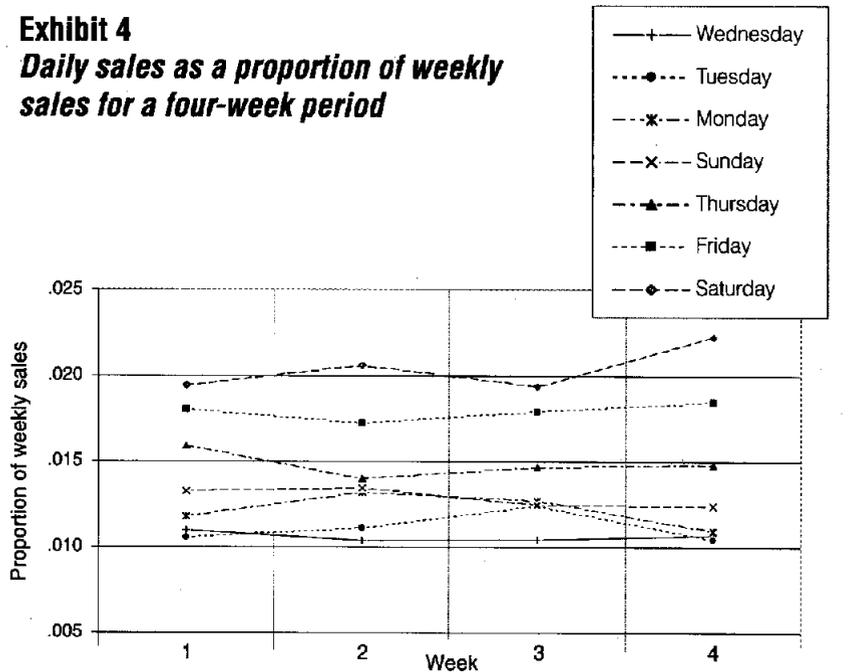


Exhibit 4
Daily sales as a proportion of weekly sales for a four-week period



Week	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1	0.1326	0.1177	0.1060	0.1101	0.1588	0.1804	0.1944
2	0.1346	0.1321	0.1114	0.1038	0.1403	0.1721	0.2057
3	0.1249	0.1273	0.1240	0.1044	0.1469	0.1790	0.1935
4	0.1243	0.1095	0.1043	0.1068	0.1481	0.1847	0.2223

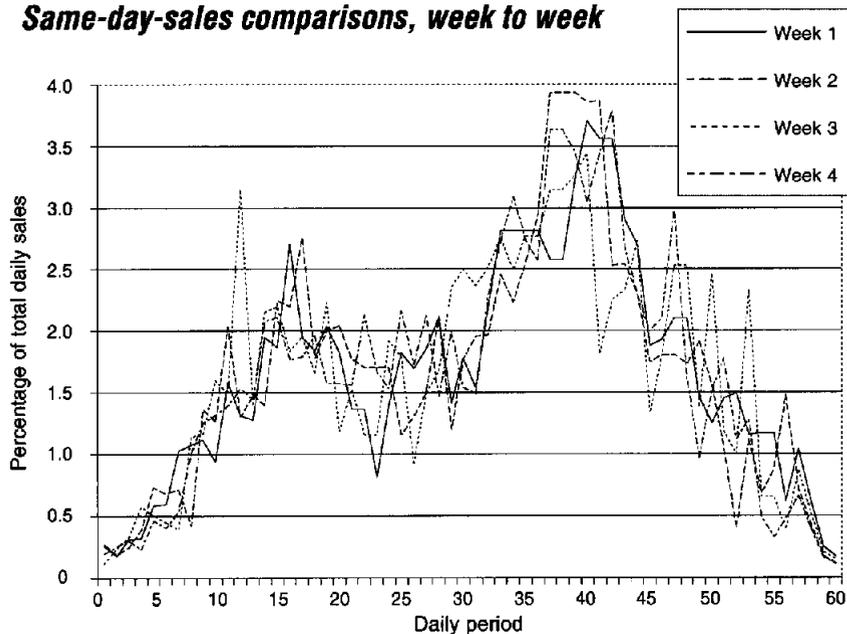
Exhibit 5

Correlations between weeks for the data in Exhibit 4

	Week 1	Week 2	Week 3	Week 4
Week 1	1			
Week 2	0.9454	1		
Week 3	0.9563	0.9664	1	
Week 4	0.9760	0.9701	0.9679	1

Exhibit 6

Same-day-sales comparisons, week to week



Sales are shown for every 15-minute period as a proportion of total daily sales, for a particular day of the week (e.g., Mondays), for four consecutive weeks.

Exhibit 7

Correlations between weeks for the data in Exhibit 6

	Week 1	Week 2	Week 3	Week 4
Week 1	1			
Week 2	0.8708	1		
Week 3	0.7954	0.7889	1	
Week 4	0.9039	0.8679	0.8262	1

The high correlation values, ranging from 0.79 to 0.90, indicate the applicability of an aggregation-disaggregation approach to forecasting within-day sales.

collected over a four-week period, and each day's total sales is expressed as a proportion of the total sales for the week. From the plot of data in Exhibit 4, one sees that those data clearly exhibit the consistency necessary for an aggregation-disaggregation approach to forecasting. That is, Saturday is consistently the busiest day of the week, followed by Friday, Thursday, Sunday, Monday, Tuesday, and Wednesday. To further verify the consistency of the data, one can run a correlation on it. Exhibit 5 lists the correlation of the daily proportions across weeks. Again, the high correlation values (all greater than 0.94) indicate that daily sales are very consistent as a proportion of weekly sales.

Managers can also look for consistency of within-day customer demand over time. For example, Exhibit 6 shows total sales by 15-minute intervals as a proportion of the total daily sales for that particular day (e.g., Mondays) for each of four weeks. In this case, the data are consistent. Moreover, one can again run a correlation analysis on the data to confirm the data's consistency (Exhibit 7).

When such data show similar same-day demand patterns week after week, it's possible to use an aggregation-disaggregation forecasting approach. In this case, the high correlation values, from 0.79 to 0.90, indicate the applicability of forecasting using the aggregation-disaggregation approach to within-day sales.

Remember, the aggregation-disaggregation approach is preferable to the independent-forecast approach when the labor driver exhibits some consistency over time, which is typically the case in hospitality organizations. Charts and correlation analyses, as just described, are helpful in evaluating whether the data are in fact, consistent.

**Step 6—
Reduce Random Variation by
Smoothing**

Some of the variation in customer demand over time is predictable, although clearly some of that change is random and therefore unpredictable. The goal of Step 6 is to deal with both predictable and random change in customer demand over short periods of time. This is best done by example.

Exhibit 8 illustrates a forecast of sales for a particular weekday (e.g., Mondays), by 15-minute intervals, using four weeks of historical data (in this case, based on the data in Exhibit 6). If that picture (Exhibit 8) represents a good forecast, a manager should be able to explain why she expects demand to materialize this way. In other words, based on the service she delivers to customers, in general she should know why customer demand peaks and lags at the times and volumes they do. It is unlikely, however, that she could say exactly what causes the “teeth” in the demand in the circled area of the chart. That is, it’s reasonable that she may be unable to explain why the forecast of demand in period 25 is higher than the forecast for period 26, which in turn is lower than the forecast for period 27.

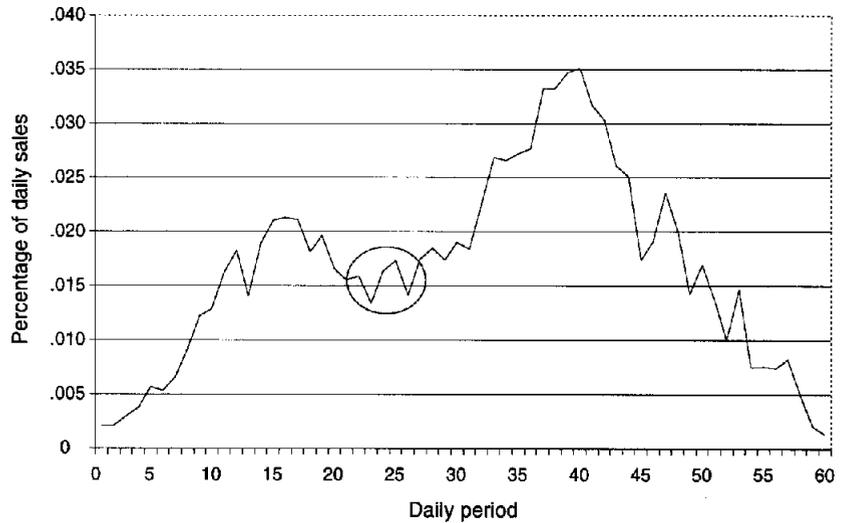
The most likely reason for the “teeth” is that the limited amount of historical data (just four Mondays’ worth) results in random variation around the true level of demand at that time of day for that particular day of the week. To eliminate some of that random variation one can smooth the forecast by averaging the original forecast for the period with the original forecasts for the two adjacent periods.⁶ Exhibit 9

⁶ Mathematically, here’s how it’s done. Let OF_i be the original forecast of demand in period i . A smoothed forecast for period i , SF_i , would then be found by:

$$SF_i = (OF_{i-1} + OF_i + OF_{i+1}) \div 3,$$

for $1 < i < \text{number of periods}$.

**Exhibit 8
Forecast of demand, by 15-minute periods**



The circled area shows the demand forecast’s unreliable “teeth.”

**Exhibit 9
Comparing unsmoothed and smoothed demand forecasts**

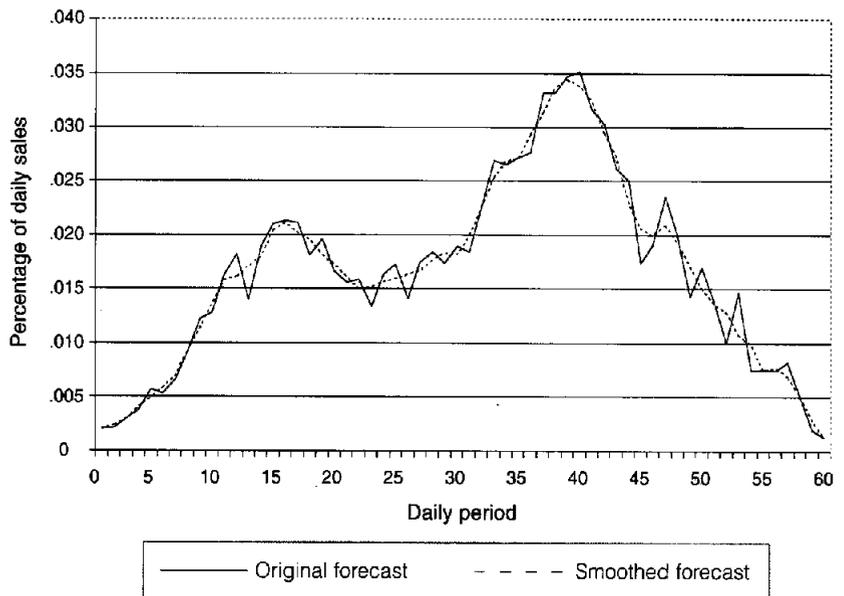


Exhibit 10
Tracking forecast error

	Forecast	Actual	Error	Percentage Error	Absolute Value Percentage Error
Week 1	10	12	2	-16.67	16.67
Week 2	11	8	3	37.50	37.50
Week 3	9	10	1	-10.00	10.00
Week 4	10	11	1	-9.09	9.09
Average		10.25			18.31

shows the results of applying the smoothing technique to the data used in Exhibit 8. Note that the smoothed forecast retains the general shape of the original forecast while eliminating the randomness (the teeth) of that forecast.

Smoothing is intended to make up for erratic data. The danger in smoothing forecasts comes when the spikes and valleys of demand are caused by real phenomena, rather than by lack of data. For example, a hotel's central reservations office may experience a large but short-lived increase in call volume following a television advertisement. If smoothing is applied to the demand trend for that day, the short-duration peak caused by the TV ad will disappear. That, in turn, will result in an understaffed CRS and poor service when the ad campaign runs again.

The determination of whether smoothing is appropriate comes down to understanding the various drivers of the service. In the call-center example, managers should know why demand peaks occur. In general, if a manager cannot offer a valid reason why brief-duration demand peaks exist, either they do not know their service as well as they should or they should control for random variation in their forecasts using smoothing.

Step 7—
Measure and Track Forecast Accuracy

Forecasts are rarely perfect. Step 7 measures and tracks forecast accuracy to ensure that the forecasting method is appropriate. There are two common yardsticks for measuring forecast accuracy: mean absolute

percentage error (MAPE) and coefficient of variation of the forecast error (COV). Both MAPE and COV measure relative error (i.e., actual demand appears in the denominator of both measures).

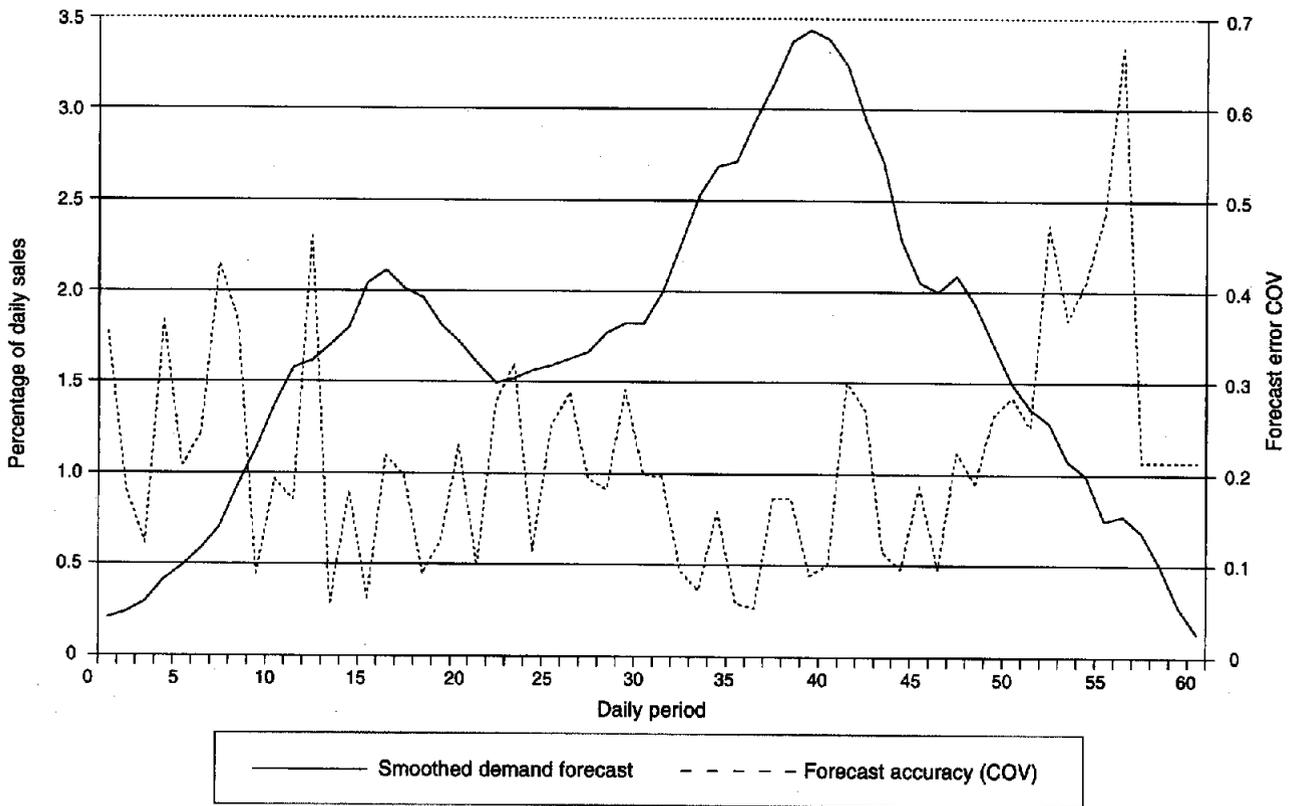
MAPE is found by taking the mean of the absolute value of the error divided by the actual demand, multiplied by 100 percent. The data shown in Exhibit 10 has a MAPE of 18.3 percent. COV is found by taking the standard deviation of the error (2.22 for the data in Exhibit 10) and dividing by the average demand (10.25 for the data below), which for this example yields a COV of 0.216. In the spreadsheet software Excel, for example, the formula is «=STDEV(data)».

In general, the forecast errors should be tracked using the time intervals used for tracking the labor driver. Using the earlier example of a 12-hours-a-day, 7-days-a-week operation, one would measure the forecast error in each of the 336 planning periods. If one plots the forecast errors, such as in Exhibit 11, one often observes high relative variability at low-demand times. In other words, forecasts commonly are relatively less accurate at low-demand times compared to a high-demand time.

Step 8—
Define the Allowable Window for the Work

As discussed in Step 1, only controllable work has timing flexibility. Thus, defining a work window applies only to controllable work. A manager must take into account the fact that the window opens at the earliest the job can begin (e.g., a guest vacates a room) and closes at the latest time the job can finish (e.g., moments before the next guest occupies that room). Work windows can be related to the operating hours of the facility or to the con-

Exhibit 11
Forecast accuracy



trollable work's drivers (e.g., having shelves stocked in a concession prior to the peak sales periods).

Controllable work requires more information than does uncontrollable work. Instead of just forecasting the labor drivers that define work volume, one also needs to forecast the drivers that define the timing of that work. Again consider the housekeeping example. The number of guests arriving that day defines the room-preparation workload. The earliest that the rooms can be prepared is defined by guest departures, while the latest that rooms can be prepared is defined by guest arrivals. So, one needs to forecast both guest arrivals and departures to determine both the volume and timing of the required work.

Drivers that determine controllable work can be either time-variant or time-invariant. For example, the drivers that define the preparation of turned-over rooms are time-variant, depending on the arrival and departure times of guests. In contrast, the driver that defines the allowable time to prepare a stay-over room is time-invariant, since by management policy, those rooms will always be cleaned within certain specified time limits (say, 9:00 AM to 4:00 PM). To forecast the timing window for controllable work, one should follow a procedure much like that outlined in Steps 2 through 7.

The next article in the series picks up with how to calculate the requisite level of staffing to meet the forecast. **CQ**