9-2-2008

Forecasting Covers in Hotel Food and Beverage Outlets

Gary M. Thompson Ph.D.
Cornell University, gmt1@cornell.edu

Erica D. Killam

Forecasting Covers in Hotel Food and Beverage Outlets

Abstract

In this report we explain our finding that a lodging property can generally use information on its occupancy to improve the accuracy of cover forecasts in its food and beverage outlets. We examine twenty-seven forecasting methods. Six of the methods forecast covers using only an outlet's historical data, while the others include information on the property's occupancy. We conducted our study using four hotels that have a total of thirty-three combinations of food and beverage outlets and dayparts. The food and beverage outlets include room service, lounges, cafés, and main restaurants. Since we have extensive historical data from one of the properties, we split that into two samples, giving a total of forty-one outlet-daypart scenarios. In all of the cases we used an eight-week holdback data set to test the models. In thirty-four of the forty-one outlet–daypart scenarios, the best forecast originated with one or another of the models incorporating occupancy data. On average, forecast accuracy improved by over 11 percent when using occupancy data. In those thirty-four cases where using occupancy data improved the forecasts, the average improvement in accuracy was over 14 percent, while the accuracy improvement exceeded 25 percent in seven of the scenarios.

Keywords

hotel food and beverage outlets, food service forecasting

Disciplines

Business | Food and Beverage Management | Hospitality Administration and Management

Comments

The Center for Hospitality Research
Hospitality Leadership Through Learning

Forecasting Covers in Hotel Food and Beverage Outlets

Cornell Hospitality Report
Vol. 8, No. 16, September 2008

by Gary M. Thompson, Ph.D., and Erica D. Killam

Cornell University
School of Hotel Administration

The Robert A. and Jan M. Beck Center at Cornell University
Back cover photo by permission of *The Cornellian* and Jeff Wang.

Cornell Hospitality Report,
Volume 8, No. 16 (September 2008)
Single copy price US$50

Cornell Hospitality Report is produced for the benefit of the hospitality industry by The Center for Hospitality Research at Cornell University

David Sherwyn, *Academic Director*
Jennifer Macera, *Associate Director*
Glenn Withiam, *Director of Publications*

Center for Hospitality Research
Cornell University
School of Hotel Administration
537 Statler Hall
Ithaca, NY 14853

Phone: 607-255-9780
Fax: 607-254-2292
www.chr.cornell.edu

The Center for Hospitality Research

Hospitality Leadership Through Learning

Thank you to our generous Corporate Members

Senior Partners

American Airlines Admirals Club
General Growth Properties, Inc.
job.travel
Southern Wine and Spirits of New York
Taj Hotels Resorts Palaces
TIG Global LLC

Partners

AIG Global Real Estate Investment
Davis & Gilbert LLP
Deloitte & Touche USA LLP
Denihan Hospitality Group
Expedia, Inc.
Four Seasons Hotels and Resorts
Fox Rothschild LLP
FX Real Estate and Entertainment, Inc.
HVS
InterContinental Hotels Group
JohnsonDiversey
Jumeirah Group
LRP Publications
Marriott International, Inc.
Marsh's Hospitality Practice
Mobil Travel Guide
Nestlé
PricewaterhouseCoopers
Proskauer Rose LLP
Smith Travel Research
SynXis, a Sabre Holdings Company
Thayer Lodging Group
Travelport
WATG
WhiteSand Consulting

Friends

American Tescor, LLP • Argyle Executive Forum • Caribbean Hotel Restaurant Buyer's Guide • Cody Kramer Imports • Cruise Industry News • DK Shifflet & Associates • ehotelier.com • EyeforTravel • Fireman's Fund • 4Hoteliers.com • Gerencia de Hoteles & Restaurantes • Global Hospitality Resources • Hospitality Financial and Technological Professionals • hospitalityInside.com • hospitalitynet.org • Hospitality Technology • Hotel Asia Pacific • Hotel China • HotelExecutive.com • Hotel Interactive • Hotel Resource • International CHRIE • International Hotel and Restaurant Association • International Hotel Conference • International Society of Hospitality Consultants • iPerceptions • Lodging Hospitality • Lodging Magazine • Milestone Internet Marketing • MindFolio • Parasol • PKF Hospitality Research • RealShare Hotel Investment & Finance Summit • Resort+Recreation Magazine • The Resort Trades • RestaurantEdge.com • Shibata Publishing Co. • Synovate • The Lodging Conference • TravelCLICK • UniFocus • WageWatch, Inc. • WIWIH.COM

Cornell University
School of Hotel Administration

Forecasting Covers in Hotel Food and Beverage Outlets

by Gary M. Thompson and Erica D. Killam

EXECUTIVE SUMMARY

In this report we explain our finding that a lodging property can generally use information on its occupancy to improve the accuracy of cover forecasts in its food and beverage outlets. We examine twenty-seven forecasting methods. Six of the methods forecast covers using only an outlet's historical data, while the others include information on the property's occupancy. We conducted our study using four hotels that have a total of thirty-three combinations of food and beverage outlets and dayparts. The food and beverage outlets include room service, lounges, cafés, and main restaurants. Since we have extensive historical data from one of the properties, we split that into two samples, giving a total of forty-one outlet-daypart scenarios. In all of the cases we used an eight-week holdback data set to test the models. In thirty-four of the forty-one outlet–daypart scenarios, the best forecast originated with one or another of the models incorporating occupancy data. On average, forecast accuracy improved by over 11 percent when using occupancy data. In those thirty-four cases where using occupancy data improved the forecasts, the average improvement in accuracy was over 14 percent, while the accuracy improvement exceeded 25 percent in seven of the scenarios.

ABOUT THE AUTHORS

Gary M. Thompson, Ph.D., is professor of operations management at the Cornell University School of Hotel Administration (gmt1@cornell.edu), where he teaches undergraduate and graduate courses in service operations management. His research, which focuses on wine cellars, restaurant operations, scheduling conferences, and on workforce staffing and scheduling, has appeared in a number of outlets. He has consulted for several prominent hospitality companies and is the founder and president of Thoughtimus, Inc., a small software development firm focusing on scheduling products.

Erica D. Killam is a 2008 graduate of the Cornell University School of Hotel Administration, where she focused her studies on hotel and restaurant food and beverage management. Currently working in operations for Aman Resorts, she plans to use her educationand experience in the management and design of small resort food and beverage outlets (erica.killam@gmail.com).

Forecasting Covers in Hotel Food and Beverage Outlets

by Gary M. Thompson, Ph.D., and
Erica D. Killam

Forecasts of customer demand are useful in many business settings, and that is certainly true for food and beverage demand in lodging properties. Good forecasts allow for better planning and superior execution. Our focus in this report is on forecasting customer demand in hotels' food and beverage outlets. Our guiding premise, which our examination supported, is that the accuracy of the forecasts will be improved by using information on the occupancy of the property, as compared to forecasting using only the outlet's historical data. To conduct our study, we employed twenty-seven forecasting methods and used data from forty-one combinations of food and beverage outlets and dayparts in four properties. These hotels are all in different cities and have different flags. Three of the properties are in the United States, while one is in Europe.

Exhibit 1

Characteristics of the lodging properties

Location	Abbreviation	Market Segment	Star Level	Number of Rooms	Data Date Range
Brussels	BRU	Urban, Business	★★★	500	2005-2007
New York	NYC	Urban, Business & Leisure	★★★★	900	2004-2007
Philadelphia	PHL	Urban, Business	★★★★★	400	2002-2003
Upstate New York	UPS	Urban, Business & Leisure	★★★	200	2003-2005 (UPS1), 2006-Mar 2008 (UPS2)

Notes: Room counts are rounded to the nearest 100 to keep the properties anonymous. The Upstate New York property's data were split into two samples because of the large amount of data.

The structure of the remainder of this report is as follows. We next review the relatively meager literature relevant to food and beverage forecasting in lodging. We then describe the forecasting contexts we examined and present the twenty-seven forecasting models we evaluated. Following that, we describe the evaluation process and present our findings and conclusions.

Literature Review

Surprisingly, we could find little literature on forecasting food and beverage covers in general, and even less on forecasting food and beverage covers in lodging operations. Earlier research that addressed forecasting of covers, in environments independent of lodging operations (or where lodging data were not considered), was performed by Miller, McCahon, and Miller and by Morgan and Chintagunta.[1] The only research we have been able to find on the topic of forecasting food and beverage covers in lodging operations was conducted by Hu, Chen, and McCain.[2] These authors evaluated methods of forecasting covers in a casino's buffet restaurant. One of their early regression models included the number of occupied rooms in the casino, but they dropped that measure because of redundancy with other variables. Thus, the single previous instance that examined the use of occupancy data did not find that it was useful to use such data.

Forecasting Contexts

For our study, we obtained data from four properties, in Brussels, New York, Philadelphia, and Ithaca, New York. For convenience, we'll refer to these properties respectively as BRU, NYC, PHL, and UPS (for upstate New York). As can be seen in Exhibit 1, which reports characteristics of the properties, the properties ranged from 200 to 900 rooms. We obtained the data from two of the properties (PHL and UPS) through personal connections, and for the others by contacting about two dozen people registered with the Center of Hospitality Research. The selection criteria for contacting people was that they had to have listed (what appeared to be) a lodging property as their employer, and they had to have downloaded at least two of the CHR's restaurant-related content items written by coauthor Gary Thompson.[3] Several of the people contacted declined to be involved in the study, for reasons including that sharing data violated company policy, the property was either not yet open or open for too short a period, or that the data were not available on a daily basis. Since the study was exploratory in nature, we are not concerned with issues of self-selection among the respondents.

Each of the properties provided at least two years of data, and the Ithaca hotel offered more than five years worth of daily data, from 2003 through the end of March 2008. To increase the size of the analysis pool, we decided to analyze this property's data separately for the periods of 2003-2005 (UPS1) and 2006-2008 (UPS2).

The properties provided data on their daily occupancy (measured as number of occupied rooms), and data on the daily cover counts in one or more of their food and bever-

[1] James J. Miller, Cynthia S. McCahon, and Judy L. Miller, "Food-service Forecasting Using Simple Mathematical Models," *Journal of Hospitality & Tourism Research*, Vol. 15, No. 1 (1991), pp. 43-58; James J. Miller, Cynthia S. McCahon, and Judy L. Miller, "Food-service Forecasting: Differences in Selection of Simple Mathematical Models Based on Short-term and Long-term Data Sets," *Journal of Hospitality & Tourism Research*, Vol. 16, No. 2 (1993), pp. 93-102; and Michael S. Morgan and Pradeep K. Chintagunta, "Forecasting Restaurant Sales Using Self-selectivity Models," *Journal of Retailing and Consumer Services*, Vol. 4, No. 2 (1997), pp. 117-128.

[2] Clark Hu, Ming Chen, and Shiang-Lih Chen McCain, "Forecasting in Short-term Planning and Management for a Casino Buffet Restaurant," *Journal of Travel & Tourism Marketing*, Vol. 16, Nos. 2-3 (2004), pp. 79-98.

[3] For example, see: Gary M. Thompson and Heeju Sohn, "Accurately Estimating Time-based Restaurant Revenues Using Revenue per Available Seat-Hour," *Cornell Hospitality Report*, Vol. 8, No. 9 (2008); and Gary M. Thompson, "Restaurant Capacity Effectiveness: Leaving Money on the Tables," *Cornell Hospitality Report*, Vol. 7, No. 7 (2007), both at chr.cornell.edu.

We evaluated 27 forecasting approaches, using different mathematical rules for each one.

age outlets. Most properties provided room service covers, in addition to cover counts in one or more other outlets. Exhibit 2 reports the food and beverage outlets for which we received data from the four properties. We analyze each meal period or daypart separately. For most of the outlets, we had data for breakfast, lunch, and dinner, although a few had additional dayparts such as brunch, tea, late-night, and special events. The combination of the dayparts and outlets yielded a total of forty-one different environments.

Research Hypothesis

The hypothesis guiding our research is straightforward, as follows:

> Using room occupancy information will improve the accuracy of cover forecasts for a hotel's food and beverage outlets.

The rationale for this hypothesis is that common sense suggests that a hotel's food and beverage outlets will be busier when the hotel is busier (i.e., running a higher occupancy) and less busy when the property is running a lower occupancy. The extent of this relationship might vary across food and beverage outlets, however. Room service demand, for instance, will likely be much more closely tied to occupancy than will demand at a property's fine dining restaurant that draws many of its customers from the local area. Additionally, this relationship would vary based on a property's location; a resort will have a far different occupancy-to-cover relationship than a downtown hotel, for example.

In this report, we test our hypothesis by comparing cover forecasts developed without regard to occupancy (i.e., those developed solely using the outlet's historical data) to those developed using information on occupancy. While occupancy can be measured in different ways, we will use the number of rooms occupied as the key metric. However, for the single property for which we have the data, we will also examine the effectiveness of using the number of guests as the occupancy metric.

To judge forecast accuracy we use the mean square error (MSE). This measure averages, across all evaluation periods (in our case, dayparts by outlet), the squared difference between the actual and forecast cover counts. The advantage of using MSE as the performance criterion is that it magnifies big forecasts errors, because the measure squares each error value. This is important, in that managers have many more options at their disposal for dealing with small forecast errors than they have for dealing with large forecast errors. That is, smaller errors are much less of a problem to managers than are large errors.

Even if we find that our hypothesis is not supported, it will not necessarily mean that covers are not affected by occupancy. For example, occupancy might well affect food and beverage demand, but, because the food and beverage outlet's historical data on cover demand would have captured the effects of the property's historical occupancy, then the forecasts implicitly would be based on those historical occupancy relationships. In this case, we would expect the forecasts incorporating occupancy data to perform as well as, but not better than, forecasts based solely on past food and beverage demand.

Forecasting Methods

We evaluated a total of 27 forecasting methods, as summarized in Exhibit 3 and explained in more detail overleaf. We abbreviate each of the methods with a multi-factor code. The first portion of the code indicates whether the primary forecasting component was based on the commonly used techniques of exponential smoothing (ExpSm) or regression (Regr).[4] The middle portion of the code indicates whether the exponential smoothing parameter alpha is fixed or adaptive; or it indicates the main independent variables of the regression. The final portion of the code indicates whether the forecasts are left as is (NoAdj), or whether they are adjusted upward or download based on the previous day's occupancy (DayBf), the current day's occupancy (DayOf), or an average occupancy for the previous and current day (Avg).

With regard to those adjustments, the rationale for evaluating the three occupancy adjustments is as follows. Breakfast is likely to be affected more by the previous day's occupancy, while dinner may well be most affected by the

[4] Both types of models were used, for example, in the earlier study of casino buffet cover forecasts by Hu, Chen, and McCain, *op.cit*.

Exhibit 2

Food and beverage outlets of the lodging properties

Property	Food and Beverage Outlets Providing Data
BRU	Room Service (B,L,D), Main Restaurant (B,L,D)
NYC	Main Restaurant (B,L,D)
PHL	Room Service (B,L,D,N,S), Café (R,L,D), Lounge (R,L,T,D), Main Restaurant (B,L,D)
UPS	Room Service (B,L,D), Main Restaurant (B,L,D), Lounge (F,V)

Key: B = Breakfast, F = Food, R = Brunch, L = Lunch, T = Tea, D = Dinner, N = Late night, S = Special event, V = Beverage

Exhibit 3

Forecasting methods evaluated

Method	Base Method	Scope	Occupancy Adjustment	Identifying Code
1	Exponential Smoothing (fixed alpha)	Day	N/A	ExpSm(Fixed)-NoAdj
2			Day Before	ExpSm(Fixed)-DayBf
3			Day Of	ExpSm(Fixed)-DayOf
4			Avg	ExpSm(Fixed)-Avg
5	Exponential Smoothing (adaptive alpha)	Day	N/A	ExpSm(Adapt)-NoAdj
6			Day Before	ExpSm(Adapt)-DB
7			Day Of	ExpSm(Adapt)-DO
8			Avg	ExpSm(Adapt)-AV
9	Regression (month)	Day	N/A	RegrByDay(Mth)-NoAdj
10			Day Before	RegrByDay(Mn)-DayBf
11			Day Of	RegrByDay(Mn)-DayOf
12			Avg	RegrByDay(Mn)-Avg
13	Regression (week)	Day	N/A	RegrByDay(Wk)-NoAdj
14			Day Before	RegrByDay(Wk)-DayBf
15			Day Of	RegrByDay(Wk)-DayOf
16			Avg	RegrByDay(Wk)-Avg
17	Regression (occupancy only)	Day	Day Before	RegrByOcc-DayBf
18			Day Of	RegrByOcc-DayOf
19			Avg	RegrByOcc-Avg
20	Regression (month, weekday)	Week	N/A	RegrByWk(Mn+Dy)-NoAdj
21			Day Before	RegrByWk(Mn+Dy)-DayBf
22			Day Of	RegrByWk(Mn+Dy)-DayOf
23			Avg	RegrByWk(Mn+Dy)-Avg
24	Regression (week, weekday)	Week	N/A	RegrByWk(Wk+Dy)-NoAdj
25			Day Before	RegrByWk(Wk+Dy)-DayBf
26			Day Of	RegrByWk(Wk+Dy)-DayOf
27			Avg	RegrByWk(Wk+Dy)-Avg

Forecasting Model Details

Exponential smoothing is the basis for Models 1-8. Exponential smoothing, which derives its name from the fact that the weights on the historical data decline exponentially, is:

$$A_t = \alpha D_t + (1- \alpha) A_{t-1} \quad (1)$$

$$F_{t+1} = [A_t] \quad (2)$$

$$E_t = F_t - D_t \quad (3)$$

where:

A_t = average as of time period t;

D_t = actual demand in time period t;

E_t = forecast error in period t (sometimes measured as $D_t - F_t$);

α = smoothing parameter ($0 \leq \alpha \leq 1$); and

[x] = integer closest in value to x.

Lower values of the parameter alpha—that is, values closer to zero—put relatively less weight on the most recent data and relatively more weight on older data. Conversely, higher values of the parameter alpha—values closer to one—put relatively more weight on the most recent data and relatively less weight on older data. At the extremes, an alpha value of zero never updates the average from its initial value (in other words, no weight is placed on the newest information), while an alpha value of one is equivalent to the traditional "naïve" method of forecasting that uses the current observation for the next forecast (in other words, all of the weight is placed on the most recent data). The forecast for next period (F_t+1), given by equation (2), simply rounds the current average (A_t) to the nearest integer, since the covers will always be whole numbers. Finally, the error in the period is the difference between the forecast and actual values, as shown in equation (3). We then square the error, and average it across all relevant periods to yield the performance criterion MSE (mean squared error).

Models 1-4 and 5-8, while all based on the exponential smoothing formulas above—that is, equations (1). (2), and (3), differ in their treatment of alpha. Models 1-4 used a fixed alpha. In this case, the alpha value that yields the best forecast over the model initialization period (eight weeks less than the entire data available) is selected. Models 5-8, in contrast, use an adaptive alpha. With an adaptive alpha, all of the data available prior to a particular day is considered. Alpha is reoptimized with every new observation, thus yielding a model that adapts to the current conditions.

Methods 2-4 and 6-8 adjust the exponential smoothing forecast upward or downward based on occupancy. Here is an example of how that adjustment is determined and made. First, the best value of alpha found in Method 1 (for Sunday, say) yields a set of error values. We then develop a regression equation that uses these exponential smoothing forecast errors (on Sundays) as the dependent variable and the room occupancy as the independent variable. (We force the regression intercept to be zero.) In essence, the results of this regression will tell us if we should adjust our exponential smoothing forecast upward or downward based on the hotel occupancy. As noted earlier, Methods 2 and 6 consider previous day occupancy, Methods 3 and 7 consider day-of occupancy, and Methods 4 and 8 use an average of the current and previous day's occupancy.

Models 9-12 are based on linear regression, where covers are the dependent variable and the independent variables are the months of the year (and occupancy, for Models 10-12):

$$Covers_t = [Intercept + \beta_{mnth2} * Mnth2 + \beta_{mnth3} * Mnth3 + \ldots + \beta_{mnth12} * Mnth12 + \beta_{occ} * Occ] \quad (4)$$

where:

β_{mnthX} = regression coefficient for month x;

MnthX = binary coding variable, given a value of 1 in month X and a value of zero otherwise;

β_{occ} = regression coefficient for occupancy (only Models 10-12); and

Occ = Room-nights—day before (Model 10), day-of (Model 11), average of day before and day-of (Model 12).

Similarly, Models 13-16 are based on linear regression with covers as the dependent variable and the weeks of the year (and occupancy, for Models 14-16) being the independent variables:

$$Covers_t = [Intercept + \beta_{wk2} * Wk2 + \beta_{wk3} * Wk3 + \ldots + \beta_{wk53} * Wk53 + \beta_{occ} * Occ] \quad (5)$$

where:

β_{wkX} = regression coefficient for week x;

WkX = binary coding variable, given a value of 1 in week X and a value of zero otherwise;

β_{occ} = regression coefficient for occupancy (only Models 14-16); and

Occ = Room-nights—day before (Model 14), day-of (Model 15), average of day before and day-of (Model 16).

To code the weeks, we used the "= weeknum(date)" function in Excel. This function works in a way that may not be obvious. For example, the second day of the year could actually fall in the second week of the year according to this function, if the first day of the year was Saturday. The best way to understand the function is to realize that the first day of the year will always fall in the first week of the year. After that, the week number increases at every Sunday. Thus, this week coding will differ from one where the first seven days of the year are always considered to fall in the first week, days 8-14 to fall in the second week, and so on. Additionally, this means that Excel calculates a year to have 53 weeks rather than the normal 52.

Linear regression also forms the basis of Models 17-19. These models have covers and only occupancy as the dependent variables:

$$Covers_t = [Intercept + \beta_{occ} * Occ] \quad (6)$$

where:

β_{occ} = regression coefficient for occupancy; and

Occ = Room-nights—day before (Model 17), day-of (Model 18), average of day before and day-of (Model 19).

While regression also forms the basis of Models 20-27, these models are week-based. Models 20-23 are given by:

$$Covers_t = [Intercept + \beta_{mnth2} * Mnth2 + \beta_{mnth3} * Mnth3 + \ldots + \beta_{mnth12} * Mnth12 + \beta_{dy2} * Day2 + \beta_{dy3} * Day3 + \ldots + \beta_{dy7} * Day7 + \beta_{occ} * Occ] \quad (7)$$

where:

β_{dyX} = regression coefficient for weekday x;

DayX = binary coding variable, given a value of 1 in weekday X and a value of zero otherwise;

β_{occ} = regression coefficient for occupancy (only Models 21-23); and

Occ = Room-nights—day before (Model 21), day-of (Model 22), average of day before and day-of (Model 23).

Finally, Models 24-27 are similar to Models 20-23, except that Models 24-27 use binary coding variables for weeks instead of months:

$$Covers_t = [Intercept + \beta_{wk2} * Wk2 + \beta_{wk3} * Wk3 + \ldots + \beta_{wk53} * Wk53 + \beta_{dy2} * Day2 + \beta_{dy3} * Day3 + \ldots + \beta_{dy7} * Day7 + \beta_{occ} * Occ] \quad (8)$$

where:

β_{occ} = regression coefficient for occupancy (only Models 25-27); and

Occ = Room-nights—day before (Model 25), day-of (Model 27), average of day before and day-of (Model 27).

The greatest number of accurate forecasts of F&B covers included room occupancy data.

current day's occupancy. Lunch might be affected by both days—that is, those people who stayed in the hotel the previous night and those who are staying overnight for the current day. By evaluating all three occupancy values, we can determine which measure is the best predictor of cover demand.

Methods 1 through 19 use separate-day forecasting. This means that these forecasting models use only data from the same day of the week when forecasting a particular day's covers. So, for example, these models would forecast one Monday's cover count using only data from previous Mondays (unless, of course, the previous day's occupancy was included in the model, in which case occupancy values from Sundays would be used, too). Models 20 through 27 use weekly forecasting, where the models include all of the data for each week, and the models themselves contain some means of forecasting the individual days of the week.

Exponential smoothing approaches. Methods 1 through 4 were based on exponential smoothing. Each method has the same fixed value for the smoothing parameter, alpha, which represents the weight on the most recent data point. Alpha was determined from the value of the parameter that gave the most accurate cover forecast for the initialization period (all but the last eight weeks of the available data, which were, as we said, held back). Methods 1 through 4 then use that fixed value of alpha to obtain the forecasts for the (eight-week) holdback period. Since Methods 1 through 4 take a separate-day approach to forecasting, the best value of the parameter alpha may vary across days. Method 1 does not use any information on occupancy; in contrast, Methods 2, 3, and 4 adjust the exponential smoothing forecast based on occupancy data. Method 2 uses occupancy from the day before (in this example, Saturdays), Method 3 uses the same day's occupancy (or day-of, Sundays in this example), and Method 4 uses the average occupancy from the previous day and current day (an average of Saturday and Sunday occupancy).

As do Methods 1 through 4, Methods 5 through 8 use exponential smoothing, but Methods 5 through 8 use adaptive values of the smoothing parameter alpha, rather than fixed parameter values. This means, then, that as new data become available, a new best value of alpha is identified for data up to that point. That value of alpha is then used in forecasting the next period's cover count. Method 5 uses the unadjusted forecast, but Methods 6, 7, and 8 adjust the forecasts upward or downward based on occupancy in the same way as the corresponding Methods 2, 3, and 4.

Regression-based forecasts. Methods 9 through 12 develop regression-based cover forecasts, using the month as the independent variable (where January is the reference month and the months of February through December are each coded using a separate binary variable) and covers as the dependent variable. Method 10 accounts for the previous day's occupancy, Method 11 for the current day's occupancy, and Method 12 includes an average occupancy for the previous and current day.

Regression-based forecasts that use the week of the year as binary-coded independent variables and covers as the dependent variable constitute Methods 13 through 16. As with the previously described methods, Method 13 includes no occupancy information, while Methods 14 through 16 include either the previous day's occupancy, the current day's occupancy, or an average occupancy for the previous and current day.

In contrast, Methods 17, 18, and 19 take a different approach to forecasting the covers, developing regression models with covers as the dependent variable and only occupancy as the independent variable. As with other approaches, Method 17 uses the previous day's occupancy, Method 18 applies the current day's occupancy, and Method 19 includes an average occupancy for the previous and current day.

Regression forecasting methods also are implemented in Methods 20 through 27. These methods, though, develop week-long forecasts, rather than separate forecasts by day. To accomplish this, these methods all incorporate binary coding variables for the days of the week (where Sunday is the reference day). In addition to the daily binary-coded independent variables, Methods 20 through 23 use binary-coded independent variables for the month of year, while Methods 24 through 27 use binary-coded independent variables for the week of the year. Methods 20 and 24 do not use occupancy data, while Methods 21 through 23 and 25 through 27 include as additional independent variables, either the previous day's occupancy, the current day's occupancy, or an average occupancy for the previous and current day.

In summary, Methods 1, 5, 9, 13, 20, and 24 do not use occupancy data when developing the forecasts. Methods 2, 6, 10, 14, 17, 21, and 25 use occupancy from the previous day; Methods 3, 7, 11, 15, 18, 22, and 26 use occupancy from the current day; and Methods 4, 8, 12, 16, 19, 23, and 27 use an average of the occupancy of the previous day and the current day.

Process

Some of the data provided by the properties required cleaning, as described in the sidebar at right. For all four properties, we used all but eight weeks of data for model initialization. (Again, the last eight weeks of data were a holdback sample used for evaluating the models.) As noted earlier, we used mean squared error to evaluate the performance of the forecasting models. To facilitate the evaluation, we built an Excel spreadsheet model that automated the process of developing and evaluating the forecasts. We validated the automated spreadsheet by comparing its results to those we found using specially developed, single-purpose spreadsheet models. The eight weeks of holdback data that we used to evaluate the performance of the forecasting models translates into fifty-six performance data points for each food and beverage outlet for each daypart. We averaged the squared error values across the fifty-six days in the holdback data set to obtain the MSE for each forecasting method.

Findings

Our study findings are presented in the accompanying tables. Exhibit 4 lists the number of times each forecasting method yielded the best forecast for a particular combination of hotel, food and beverage outlet, and daypart, based on the forty-one scenarios we analyzed. The best performing forecasting method was the regression model using covers as the dependent variable and occupancy as the independent

Exhibit 4

Number of best forecasts, by model

Model	Number of Best Forecasts	Uses Occupancy
ExpSm(Fixed)-NoAdj	5	No
RegrByOcc-Avg	5	Yes
RegrByWk(Mn+Dy)-Avg	4	Yes
RegrByOcc-DayOf	4	Yes
RegrByWk(Wk+Dy)-DayBf	4	Yes
RegrByOcc-DayBf	3	Yes
RegrByDay(Wk)-DayBf	3	Yes
RegrByWk(Mn+Dy)-DayOf	2	Yes
RegrByWk(Wk+Dy)-DayOf	2	Yes
ExpSm(Adapt)-DayBf	1	Yes
ExpSm(Adapt)-NoAdj	1	No
ExpSm(Fixed)-Avg	1	Yes
ExpSm(Fixed)-DayOf	1	Yes
RegByDay(Mn)-Avg	1	Yes
RegByDay(Mn)-DayBf	1	Yes
RegByDay(Mn)-DayOf	1	Yes
RegrByWk(Mn+Dy)-DayBf	1	Yes
RegrByWk(Wk+Dy)-NoAdj	1	No

Data Cleansing

For each of the properties we performed a variety of steps to "clean" the data. Fifteen data points were removed from the original BRU data sent to us from the property. Nine of the data points were negative numbers, mostly occurring in the room service cover counts, with the additional six points being cover counts over 1,000. We assume that the cover counts that exceeded 1,000 were the result of data entry errors, but no explanation for the negative numbers could be found. These gaps in the data sets were filled in by repeating the cover count from the same meal period or daypart and the same day of week from the previous week. For PHL, we removed the day after Thanksgiving and Christmas Day, since these were outlier days when the property's café was closed. We also removed New Year's day. Somewhat surprisingly, we needed to remove no data points from the NYC or UPS data.

Exhibit 5

Result by property, F&B outlet, and daypart

Property/ F&B Outlet/DayPart	Best MSE	Best MSE without Occupancy	Percentage Improvement
BRU/RS/B	173.3	173.3	0.0%
BRU/RS/L	52.5	47.8	8.9%
BRU/RS/D	214.6	146.8	31.6%
BRU/MR/B	3691.0	2085.2	43.5%
BRU/MR/L	221.9	168.2	24.2%
BRU/MR/D	281.0	281.0	0.0%
NYC/MR/B	1947.1	1945.2	0.1%
NYC/MR/L	605.6	582.4	3.8%
NYC/MR/D	1132.4	749.8	33.8%
PHL/RS/B	554.3	468.7	15.4%
PHL/RS/L	215.4	202.6	5.9%
PHL/RS/D	478.6	336.8	29.6%
PHL/RS/N	91.4	87.4	4.4%
PHL/RS/S	10.8	10.8	0.0%
PHL/MR/B	867.5	804.1	7.3%
PHL/MR/R	757.3	751.1	0.8%
PHL/MR/L	712.4	706.5	0.8%
PHL/MR/D	1004.6	993.9	1.1%
PHL/CF/R	339.8	331.6	2.4%
PHL/CF/L	70.4	70.4	0.0%
PHL/CF/D	151.6	151.6	0.0%
PHL/LN/R	1101.3	1075.4	2.3%
PHL/LN/L	470.8	467.9	0.6%
PHL/LN/D	145.4	145.4	0.0%
PHL/LN/T	287.7	277.5	3.6%
UPS1/RS/B	54.8	42.5	22.6%
UPS1/RS/L	22.5	21.2	5.6%
UPS1/RS/D	98.4	75.9	22.9%
UPS1/MR/B	1086.8	925.5	14.8%
UPS1/MR/L	1288.1	1152.8	10.5%
UPS1/MR/D	880.7	646.9	26.5%
UPS1/LN/F	312.1	181.2	41.9%
UPS1/LN/V	471.2	404.3	14.2%
UPS2/RS/B	33.3	26.7	19.9%
UPS2/RS/L	3.3	3.1	5.4%
UPS2/RS/D	54.5	46.1	15.4%
UPS2/MR/B	420.4	283.3	32.6%
UPS2/MR/L	7488.9	7488.9	0.0%
UPS2/MR/D	1483.7	1451.9	2.1%
UPS2/LN/F	121.1	100.0	17.4%
UPS2/LN/V	170.4	148.8	12.7%

CF = Café, LN = Lounge, MR = Main Restaurant, RS = Room Service.
B = Breakfast, D = Dinner, F = Food, L = Lunch, N = Late night, R = Brunch, S = Special event, T = Tea, V = Beverage.
BRU=Brussels; NYC=New York City; PHL = Philadelphia; UPS = Upstate New York.

variable. In its three forms (Methods 17, 18, and 19), this forecasting method yielded the best forecast in twelve of forty-one scenarios. Overall, the forecasting methods that incorporated occupancy data yielded the best forecast in thirty-four of the forty-one scenarios—a substantial majority. In contrast, the forecasting methods that omitted occupancy yielded the best forecasts in only seven of the forty-one scenarios. Of the thirty-four scenarios in which using occupancy data improved the forecasts, thirteen used day-before occupancy, ten used same-day occupancy, and eleven used an average of day-before and day-of occupancy.

Exhibit 5 presents the forecast errors (MSEs) for each of the forty-one scenarios that we tested. Errors are presented for the best method that does not use occupancy data and the best forecasting method overall. Our assumption in presenting these results is that a manager would not choose to use a method that incorporates occupancy unless it was superior to the best of the methods that are not occupancy based. As we saw with Exhibit 4, using occupancy data improves the forecasts in about four out of five scenarios. If occupancy data are considered, the average improvement in MSE across all forty-one scenarios is 11.8 percent. On the thirty-four scenarios where occupancy data improve the forecasts, MSE is reduced by 14.3 percent, on average. There were seven scenarios where the accuracy improvement exceeded 25 percent.

The histogram in Exhibit 6 shows the proportion of scenarios yielding different percentage improvements in forecast accuracy. In just over half of the scenarios, the improvement in forecast accuracy exceeded 5 percent, while accuracy improved by more than 20 percent in just under one-fourth of the scenarios.

An interesting analysis is to examine the number of times that the best forecast was generated using a particular occupancy measure. Exhibit 7 reports the results of such an analysis. Across the four properties and ten food and beverage outlets, the best forecasts for breakfast typically included, not surprisingly, the previous day's occupancy (sometimes averaged with the current day's occupancy). Less intuitively, the most accurate lunch cover forecasts also typically used the previous day's occupancy. As anticipated, the most accurate dinner cover forecasts generally used occupancy from the current day (sometimes averaged with the previous day's occupancy).

The Brussels property provided guest counts in addition to the number of occupied rooms. We used those data to examine whether it would be better to use rooms occupied or guest counts when developing forecasts. As shown in Exhibit 8, using rooms occupied was better in four of the six outlet and daypart scenarios. Its average excess error was 1.8 percent, meaning that using rooms occupied would result in an MSE that was 1.8 percent higher, on average, than the best

MSE found using either rooms occupied or guest counts. This performance is better than that obtained using guest counts, which has an average excess error of 3.1 percent. A key point, however, is that the quality of the forecasts can be improved if both rooms occupied and guest count data are available, and the best approach is selected on a case-by-case basis.

Conclusions

Our findings show strong support for our hypothesis that using occupancy data can improve the accuracy of food and beverage cover forecasts. Indeed, forecast accuracy improved by over 11 percent for all the scenarios we examined when we included occupancy data, by over 14 percent in cases where improvements occurred, and by more than 25 percent in more than one in every six scenarios. Nevertheless, we caution that you should not expect that occupancy data will always improve forecasts. For example, three of the cases where occupancy data did not improve the forecasts were in the main restaurant of the Philadelphia property, suggesting that this restaurant is little affected by the hotel's occupancy. These data show that this outlet is operating independently of the property. Indeed, this restaurant is well regarded in the local community and its draws much of its business from the local market.

Contrasting findings. In light of our finding that using occupancy data improved forecasts, let's return to Hu, Chen, and McCain's finding that occupancy had no such effect on forecasting casino buffet covers.[5] We can see several possible reasons for this. First, it appears that these researchers tested only same-day occupancy figures, while we also considered occupancy from the previous day and an average of the same-day and previous-day occupancies. Second, Hu *et al.* used only the week-based regression models and not the day-based models that were so effective in our study. Third, Hu and colleagues evaluated far fewer forecasting models—eight compared to our twenty-seven—and only one of their

[5] Hu, Chen, and McCain, *op.cit*.

Exhibit 6

Result by property, F&B outlet, and daypart

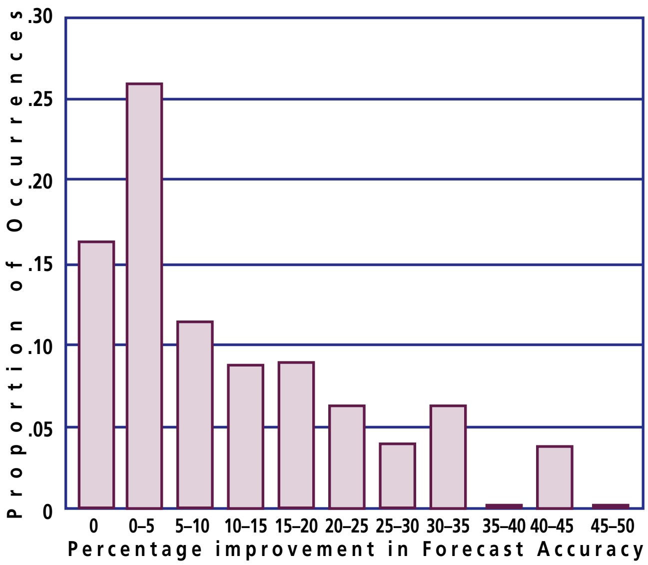

Exhibit 7

Number of instances where each occupancy value yielded the best forecast, by meal period

Meal Period	Occupancy Data Used: None	Day Before	Day Of	Average
Breakfast	1	4	0	4
Brunch	0	3	0	0
Lunch	2	5	0	4
Tea	0	1	0	0
Dinner	3	0	5	3
Late Night	0	0	1	0
Special Events	1	0	0	0
Food Covers	0	0	2	0
Beverage Covers	0	0	2	0

Exhibit 8

Best forecasts for Brussels hotel, comparing rooms occupied with number of guests

	Room Service			Main Restaurant			Average Excess
	Breakfast	Lunch	Dinner	Breakfast	Lunch	Dinner	Error
Rooms Occupied	173.3	47.8	146.8	2085.2	168.2	281.0	1.8%
Number of Guests	166.8	49.3	148.9	2065.5	192.5	266.0	3.1%

The results encourage further exploration of the value of using room occupancy or guest counts to forecast hotel F&B sales.

models incorporated occupancy (the one that they eventually dropped). By contrast, half of our models considered occupancy data. Fourth, it may simply be that the particular casino or particular buffet that Hu *et al*. studied happened to be one of those food and beverage outlets where occupancy data do not help improve the forecasts (as is the case of the main restaurant in the Philadelphia property in our study). In our study, using occupancy data failed to improve the forecasts in seven of the forty-one scenarios.

Given that only a small proportion of the hotels we approached to participate in this study were able to meet our data requests, we believe it is likely that few properties are performing the types of analysis on which we reported in this report. Given our findings, the implication is that a large number of hotel-based food and beverage managers have the potential to improve the quality of their forecasts. To make this happen, we recommend that properties begin immediately to capture data every day, for each daypart or meal, and store this information in a form that facilitates the types of analysis we performed. With accurate data, the ability to forecast covers then becomes a matter of determining the most accurate forecast model for each of a property's outlets—keeping in mind that each outlet and each daypart may require a different model. Once the initial forecast model has been developed, it can be automated to allow individual outlet managers to plug in their data and easily generate a forecast, without worrying about the calculations. Periodically, the model creator would need to update the model by adding current occupancy and cover data to maintain the model's accuracy. Because of the complexity of this type of analysis and the large amount of data needed to create an accurate model, the problem for most operations would be having a manager or director with the knowledge and time to create a model in the first place.

Five areas for future investigations are suggested by our findings. First, extending the analysis to a broader set of properties would be worthwhile, especially given the comparison of our findings to those of Hu and colleagues. Given that our existing set of properties is diverse (based on markets and geography), we would expect to see that occupancy will continue to be of value in improving forecast accuracy. Second, one could evaluate a larger set of forecasting methods. Even though we evaluated twenty-seven forecasting methods, we expect that the overall forecast accuracy might improve with additional methods. That said, it will still be beneficial to consider occupancy data. Third, if we were able to obtain data for more properties, it would be interesting to repeat the analysis we performed on the Brussels hotel on other properties, that is, using both rooms occupied and guest counts. With a broader set of properties, it may emerge that one of these measures of occupancy dominates the other in terms of forecast accuracy. Fourth, it might be interesting to repeat the analysis using different metrics for forecast error, such as the commonly used mean absolute percentage error (MAPE). While the numbers will change, we see no reason to expect that occupancy would not improve forecast accuracy, regardless of the metric used. Finally, the investigations we performed should be considered as a way to improve the accuracy of day-of (or next-day) cover forecasts (i.e., forecasts made for the current or following day) since we are using occupancy data from the current day or the day before to improve the forecasts. While these forecasts can certainly be useful for fine-tuning one's plans, a richer, though much more complex analysis would attempt to forecast covers farther into the future. To do this, occupancy forecasts for the appropriate future days would have to be used instead of actual occupancy figures (most hotels do create such occupancy forecasts). An investigation like this would need to consider such things as the property's booking curve, for example, and would require significantly more historical data to conduct. Properties with the necessary data should consider contacting us, since we are interesting in pursing this line of inquiry. ■

Cornell Hospitality Reports

Index

www.chr.cornell.edu

2008 Reports

2008 Hospitality Tools

2008 Industry Perspectives

2007 Reports

Cornell Hospitality Reports Index (continued)
www.chr.cornell.edu

2007 Hospitality Tools

2006 Reports

www.chr.cornell.edu