Research Article

Forecasting hotel arrivals and occupancy using Monte Carlo simulation

Received (in revised form): 1st May 2009

Athanasius Zakhary[a], Amir F. Atiya[b], Hisham El-Shishiny[c] and Neamat E. Gayar[a]

[a]Department of Information Technology, Faculty of Computers and Information, Cairo University, Giza, Egypt; [b]Department of Computer Engineering, Cairo University, Berkeley, Giza, Egypt; and [c]Advanced Technology and Center for Advanced Studies at IBM Technology Development Center, Cairo, Egypt.

Athanasius Zakhary received his BS degree from the Department of Information Technology, the Faculty of Computers and Information, Cairo University, Egypt. Since then, he has been a researcher with the Data Mining and Computer Modeling Center of Excellence, Ministry of Information and Telecommunications (MCIT). In addition, he is affiliated with the Faculty of Computers and Information, Cairo University, where he is pursuing his Masters degree. His research interests are in the theory of forecasting, neural networks and machine learning.

Amir F. Atiya received his BS degree in from Cairo University, Egypt, and the MS and PhD degrees from Caltech, Pasadena, CA, all in electrical engineering. Dr Atiya is currently a professor at the Department of Computer Engineering, Cairo University. He recently held several visiting appointments, such as in Caltech and in Chonbuk National University, South Korea. His research interests are in the areas of neural networks, machine learning, theory of forecasting, pattern recognition, computational finance and Monte Carlo methods. He obtained several awards, such as the Kuwait Prize in 2005, and was an associate editor for the IEEE Transactions on Neural Networks from 1998 to 2008.

Hisham El-Shishiny is currently Senior Scientist and Manager of Advanced Technology and Center for Advanced Studies at IBM Technology Development Center in Cairo. He received his PhD and MSc in Computer Science with highest honors from Nancy University in France and his BSc in Electronics and Communications Engineering from Cairo University. He has also completed a 2-year post-graduate program in Operations Research and conducted post-graduate studies in Solid State Sciences at the American University in Cairo. His current research interests are Data Mining, Machine Learning, Image Processing and Visualization, HPC, Optimization, Knowledge Management and Natural Language Processing.

Neamat E. Gayar is an associate professor in the Department of Information Technology, Faculty of Computers and Information, Cairo University, Egypt. She is currently also appointed at the Nile University, School of Communication and Information Technology in Giza, Egypt on leave from Cairo University. Dr Neamat obtained her MSc and her PhD in Computer Science from the Faculty of Engineering, University of Alexandria, Egypt. Dr Neamat's research interests lie mainly in the fields of pattern recognition and machine intelligence, artificial neural networks, fuzzy systems and neuro-fuzzy systems, multiple classifiers, soft computing, data mining and intelligent data-analysis techniques. She served in the organizing and program committees of several conferences and is a member of the program committee of the International Journal of Graphics, Vision and Image Processing.

Correspondence: Amir Atiya, Cairo University, Giza, Egypt
E-mail: amir@alumni.caltech.edu

ABSTRACT Forecasting hotel arrivals and occupancy is an important component in hotel revenue management systems. In this article, we propose a new Monte Carlo simulation approach for the arrivals and occupancy forecasting problem. In this approach, we simulate the hotel reservations process forward in time, and these future Monte Carlo paths will yield forecast densities. A key step for the faithful emulation of the reservations process is the accurate estimation of its parameters. We propose an approach for the estimation of these parameters from the historical data. Then, the reservations process will be simulated forward with all its constituent processes such as reservation arrivals, cancellations, length of stay, no shows, group reservations, seasonality, trend and so on. We considered as a case study the problem of forecasting room demand for Plaza Hotel, Alexandria, Egypt. The proposed model gives superior results compared to existing approaches.

Journal of Revenue and Pricing Management (2011) **10,** 344–366. doi:10.1057/rpm.2009.42;
published online 13 November 2009

Keywords: hotel demand forecasting; hotel arrivals forecasting; occupancy forecasting; Monte Carlo; forecasting by simulation; demand uncontraining

INTRODUCTION

Revenue management (RM) is the science of managing the available amount of supply to maximize revenue by dynamically controlling the price/quantity offered (Bitran and Caldentey, 2003; Ingold *et al*, 2003; Talluri and Van Ryzin, 2005). RM systems have been widely adopted in the hotel industry. Because of the large number of existing hotels, any possible improvement in the technology will amount to potentially very large overall savings. A key component of hotel room RM system is the forecasting of the daily hotel arrivals and occupancy. Inaccurate forecasts will significantly impact the performance of the RM system, because the forecast is the main driver of the pricing/room allocation decisions (see Lee (1990) and Weatherford and Kimes (2003) for discussions of this issue). In fact, Chiang *et al* (2007) mentions that for the airline industry it is estimated that a 20 per cent improvement in forecasting error translates into a 1 per cent increase in revenue generated from the RM system (Lee (1990) estimates an even larger impact). For either the airline industry or the hotel industry this will probably impact the net income in a much larger way, because of the small margins existing in these industries.

In this article, we consider the problem of forecasting daily hotel arrivals and hotel occupancy, with the forecast horizon being several months. In the theory of forecasting, there have been two competing philosophies. The first one is based on developing an empirical formula that relates the value to be forecasted with the recent history (for example, ARIMA-type or exponential smoothing models). The other approach focuses on developing a model from first principles that relates the value in question with the available variables/parameters and so on, and simulates that model forward to obtain the forecast. This approach has been prevalent in weather forecasting, where the partial differential equations (PDEs) relating the weather variables are simulated in a spatiotemporal way (see Sivillo *et al*, 1997). Because the majority of real-world systems are either intractable or very complex to model, most forecasting applications follow the first approach. In contrast, we follow here the second approach. In other words, the proposed model is based on simulating forward in time in a Monte Carlo fashion the actual hotel mechanisms. Rather than simulating PDEs such as in weather forecasting, we here simulate the hotel's reservations processes. What makes the modeling quite intricate is the existence of many often inter-related processes: reservation arrivals, cancellations, duration of stay, no shows, group reservations, seasonality, trend and so on. We propose methods that attempt to model all these processes as faithfully as possible. This is achieved by estimating the distributions of the different quantities from the actual data if feasible, and, if the data are insufficient, aggregating the data in reasonable ways so as they become sufficient for obtaining relatively accurate estimation. The advantage of such methodology is that it yields the density of forecasts. This is a very beneficial aspect from the point of view of RM. The RM problem can be formulated using a dynamic programming-type construction (Bitran and Mondschein, 1995; Liu *et al*, 2006). It is desirable to have the density of the forecasts, rather than just point forecasts. This is because of the probabilistic nature of dynamic programming-type RM formulations. For example, the 'value function' is typically computed using probability transitions. Other stochastic optimization methods as well would need at least an estimate of the variance of the forecasts.

In this work, we considered as a case study the problem of forecasting the arrivals and occupancy of the Plaza Hotel, Alexandria, Egypt.

There are several distinct advantages of the proposed approach:

- It estimates the density of the forecasts, and hence also confidence intervals.
- It allows for estimating other quantities of interest, for example the probability of reaching the hotel capacity limit or a certain fraction thereof.
- It allows for scenario analysis. For example, one can examine the effect of overbooking on future arrivals. As another example, one can explore the effect of a cancellation penalty beyond some date before arrival.
- The sensitivity of the arrival forecast and the occupancy forecast owing to changes of some control variables (for example, booking limits) can in many cases be estimated. This, of course, is very useful to the RM aspect of the problem.
- It allows for forecasting unconstrained demand. This means the total demand that would have occurred, had the hotel not been limited by its room capacity and had it accepted every single reservation. This is an important quantity from the point of view of RM.
- The presented approach is very flexible in that it can accommodate many possible inputs from the hotel manager, and a lot of possible judgmental information (for example, some expected rise in reservation arrivals due to some anticipated future event such as a major convention or a sports event).

The article is organized as follows: In the next section, we briefly review other work on hotel arrivals and occupancy forecasting. The next following section presents the problem description and definitions. The subsequent section describes the proposed approach, specifically the estimation of the system components. The next section details how we put these components together to obtain the forecast. The next following section discusses some miscellaneous aspects, such as the level of aggregation and unconstrained forecasting. The preceding section gives an overview of the considered case study (Plaza Hotel). The penultimate section presents the simulations results, and the final section is the conclusion of the work.

RELATED WORK

There have been few published articles on hotel arrivals forecasting. Most of the work derives from approaches developed for the airline reservations forecasting problem. The airline problem has many similarities with the hotel room problem, such as dealing with reservations, cancellations and so on. But, there are still non-trivial differences that have to be taken into account. For example, the length of stay (LOS) is a variable existing in the hotel room problem, but not in the airline problem. A good review of the forecasting approaches can be found in Lee (1990) for the airline problem and Weatherford and Kimes (2003) for the hotel room problem. Basically, the approaches can be grouped into two categories: time series models and reservations-based models. Time series models consider only the arrivals or the occupancy time series, and apply time series models (such as exponential smoothing, ARIMA and so on) on these. No use is made of the reservations data. Examples of applications of the time series approach include Sa (1987) and Lee (1990) both of which applied ARIMA models.

The reservations-based approach, on the other hand, makes use of the reservations data to forecast future arrivals. It typically utilizes the concept of 'pick-up'. This means that given K reservations for a future day T, we expect to 'pick-up' N more reservations from now until T. The forecast will then be $K+N$. There are two versions of the pick-up model (see Weatherford and Kimes, 2003). In the additive version, we add to the current number of reservations the average number of reservations typically picked up between the current date and the arrival date (of course, taking seasonality into account). The multiplicative pick-up model is similar except that we add a fraction of the current number of reservations, rather than an amount independent of that number. Examples of applications of the reservations-based approach include L'Heureux (1986), and also the extensions introduced by Sa (1987), Wickham (1995), Skwarek (1996), Weatherford (1997) and Bitran and Gilbert (1998) who extended the approach by adding a linear regression component. There has been work combining the reservations-based approach and the time series approach using concepts of forecast combination (see Ben-Akiva, 1987). The problem, however, with the reservations-based approaches is that they are designed to forecast only the arrivals, but not the occupancy. Forecasting occupancy is an essential task when developing a RM system.

There has also been analytical attempts to model net bookings in the presence of reservation arrivals and cancellations. The so-called stochastic model has been developed by Lee (1990) for the airline problem. In that approach, the reservation process is modeled as a 'birth-death process'; see Bailey (1964). This is a branch of the area of stochastic processes that analyzes the dynamics of births and deaths in a population. In our case, each reservation is considered as a 'birth', and each cancellation is considered as a 'death' of an existing reservation. It is not clear, however, how to extend this approach to the hotel room case, where there is a third dimension to the problem represented by the LOS.

Most of the above references are developed mainly for the airline problem. There is little work applied to the hotel room problem. We review this work in what follows. Weatherford and Kimes (2003) compared among a number of time series models and reservations-based models for daily hotel room arrival forecasting. They found that the additive pick-up and a reservations-based approach that uses linear regression gave the best results. Zakhary *et al* (2008) focused on the reservations-based approaches, and compared among different variations. Yuksel (2007) applied several versions of exponential smoothing, as well as ARIMA and some Delphi methods, to forecasting monthly hotel arrivals. Ben Ghalia and Wang (2000) developed a forecasting system using aspects of fuzzy modeling that can accommodate judgmental facts. Pfeifer and Bodily (1990) considered a space-time ARMA approach for the hotel room arrivals problem. Andrew *et al* (1990) applied the Box-Jenkins approach and exponential smoothing to forecasting monthly hotel occupancy rates. Also, Chow *et al* (1998) used ARIMA for the hotel occupancy forecasting problem. Schwartz and Hiemstra (1997) applied a novel idea for daily occupancy forecasting. They compared the shapes of the booking curves for the previous days to that of the current day. Then, they based the forecast on the most similar booking curve. Rajopadhye *et al* (2001)'s model is probably the only model that has some aspect of simulation like our approach. Their approach, however, is different and is of a smaller scale than our approach. Their main approach is a time series forecasting model using Holt-Winter's exponential smoothing, but they also used a simulation approach to obtain short-term forecasts as well.

PROBLEM DESCRIPTION AND DEFINITIONS

Hotel arrivals are mainly driven by two processes of opposing effects: *reservations* and *cancellations*. A potential hotel guest makes a reservation, typically a few days or a few weeks before the intended arrival day (if space is available). Typically, the rate of reservation arrivals picks up significantly as the arrival day gets closer. A *denied* reservation request is a request, which is rejected by the hotel owing to lack of room availability for all or part of the intended stay period. Reservations can get canceled any time before arrival. The cancellation rate not only increases, the closer we get to the arrival day, but it is also influenced by the hotel's cancellation policy (which could include some penalties). The *total bookings* at any time τ before arrival day t is the total reservations net of cancellations made for the particular arrival day (thus, it equals total reservations minus total cancellations made up to this time τ). The *booking curve* is the graph of total bookings as a function of time until arrival (that is, $t-\tau$). The *arrivals* represent the net number of guests that check-in at a particular day t. *Occupancy* is the number of occupied rooms at a particular day t. It could as well be measured as a percentage of the hotel room capacity. These latter two time series are in this study the target variables to be forecasted.

Walk-in customers are customers that check-in without reservations. For example, they just show up at the hotel requesting a room for the current day. Also, some potential guests, who have reserved a room, do not show up on arrival day (potentially forfeiting their or part of their first night's payment). These are called *no-shows*.

Every room is reserved for a number of nights. This is called the *length of stay*. After the guest arrives, he could possibly checkout before the expected checkout date, leading to an *understay*. He could also checkout after the expected checkout date, leading to an *overstay*.

There has been recent interest in the hotel industry in applying RM systems. A large improvement in hotel revenue and/or profit could potentially be achieved when applying well-designed RM systems. Most RM systems work by segmenting the customers into categories, and dynamically allocating a number of rooms and specifying the price for each category according to the expected demand, in a way that maximizes hotel revenue (Vinod, 2004).

Another aspect of RM is the implementation of an overbooking strategy. This means that the hotel will allow bookings to exceed the available hotel capacity, in anticipation that several reservations will be canceled. This strategy is expected to increase the level of hotel occupancy, and hence also the revenue. However, there is the flip side that if more guests with valid reservations arrive than available rooms the hotel would lose some good will and possibly incur some extra costs related to rebooking the extra guests in neighboring hotels ('walking the guests').

The need for accurate arrivals and occupancy forecasting arises because of the following aspects:

- The optimal number of rooms in each room category in an RM system is mainly influenced by future room demand.
- The price of each category should also be fixed according to the future demand. This arises from the well-known supply/demand relationship. For this and for the previous item, these quantities are determined in the framework of some formulated optimization problem.
- The optimal overbooking strategy can be determined as well. In fact, the proposed model allows for obtaining forecasts in the presence of any specific overbooking strategy.

We note that even though our approach gives the flexibility to incorporate an overbooking strategy, at this point we will not consider it in our experimental simulations and we will focus only on the forecasting aspect.

ESTIMATION OF THE SYSTEM'S COMPONENTS

As mentioned, there are two major phases in our approach. In the first phase, we estimate all the parameters of the reservations process (only the in-sample data are used for this purpose). In the second phase, we simulate the reservations process forward in time to obtain the forecasts using the parameter estimates obtained in the first phase. In this section, we describe in detail the parameter estimation phase.

Seasonality

Seasonality is a major factor that considerably affects the level of room demand. Most hotels have busy periods, where demand pushes up to full occupancy, and low periods with plenty of vacant rooms. By mastering the periods of high and low demand, pricing and room allocation can achieve more efficient revenue optimization.

In the hotel business, the days are usually categorized into: high season and low season. Some hotels, however, have more seasonal levels. For example, Plaza Hotel, our case study, has a third seasonal level that they label 'very low season'. For concreteness sake, we will follow the case of Plaza Hotel in this description. Of course, the model can be easily customized to accommodate the seasons' convention of any other hotel. Thus, we classified seasonality into three categories:

- high season;
- low season; and
- very low season.

The classification of the different days of the year into these three seasonal regimes is obtained by consulting with the hotel managers. A good strategy when forecasting a time series is to deseasonalize the time series (see Franses, 1998), in order to have the forecasting model focus on the medium-term or long-term variations or trends. Towards this end, the seasonal average curve has to be estimated. We have chosen a multiplicative seasonality model. The seasonal average is estimated as follows:

$$s_{\text{avg}}(t) = \frac{1}{N_H} \sum_{t' \in S_H} \frac{s(t')}{\text{Avg}(s(\tau))} \quad \text{for} \quad t \in S_H \tag{1}$$

$$s_{\text{avg}}(t) = \frac{1}{N_L} \sum_{t' \in S_L} \frac{s(t')}{\text{Avg}(s(\tau))} \quad \text{for} \quad t \in S_L \tag{2}$$

$$s_{\text{avg}}(t) = \frac{1}{N_{VL}} \sum_{t' \in S_{VL}} \frac{s(t')}{\text{Avg}(s(\tau))} \quad \text{for} \quad t \in S_{VL} \tag{3}$$

where S_H, S_L and S_{VL} are the sets of, respectively, the high season days, the low season days and the very low season days. As mentioned, these are determined by the hotel managers. The size of these sets is, respectively, N_H, N_L and N_{VL}. The term $s(t)$ that is averaged here is the total reservations that arrived for arrival day t, taking out the cancellations that occurred. A precise definition of this variable, as well as the rationale for excluding the cancellations will be given in the next subsection. Concerning $\text{Avg}(s(\tau))$, it is a normalization factor representing the average of $s(t)$ over the year in which it exists.

One can see that equations (1), (2) and (3) lead to a tri-level piecewise constant function. However, in some cases this might be too broad of a classification to track the detailed seasonal variations. So, a possible modification is to partition large contiguous periods of the above three periods into subperiods by detecting clusters of relatively high or relatively low periods using some moving average mechanism. For example, if from 6 October 2008 to 4 December 2008 is designated a low season period, we partition it further to have a seasonal average that varies somewhat in this period. We have implemented this approach rather than a tri-level season average in our case study (see the details in Seasonal analysis section). In any case, whether any modeler should have three fixed levels, or should partition further to obtain a finer-grain seasonal average depends on the problem at hand. He should analyze the data and estimate the potential merits of each approach. In any case, as we will see, the broad classification into three major seasonal regimes will also be utilized in the categorization of the reservation curves and the LOS distributions.

Subsequent to computing the seasonal averages, we deseasonalize the series, as follows:

$$s_{\mathrm{des1}}(t) = \frac{s(t)}{s_{\mathrm{avg}}(t)} \qquad (4)$$

where the subscript 'des1' means a deseasonalized series, and the number 1 attached to it signifies that it is still an intermediate step as one more step will be considered in the next paragraphs.

This analysis so far considers only the seasonality regimes in the different periods of the year. There is another source of seasonality, namely day of the week seasonality (or in short weekly seasonality). While hotels that cater to business guests will have high weekday guest traffic, the converse is true for resort hotels. Their busy periods will be during the weekend. As such, the day of the week arrival numbers will follow a distinct pattern. When estimating the weekly seasonality, we consider the deseasonalized time series, obtained in the previous seasonality analysis step, that is $s_{\mathrm{des1}}(t)$. We then apply the following weekly deseasonalization algorithm:

1. For every time series point (of the series $s_{\mathrm{des1}}(t)$) compute the average value for the week containing it. Denote this by $AvgWk$.
2. Compute the relative or normalized time series value:

$$s'_i(t) = \frac{s_{i,\mathrm{des1}}(t)}{AvgWk} \qquad (5)$$

where $s_{i,\mathrm{des1}}(t)$ is the time series value at time t (that is, $s_{\mathrm{des1}}(t)$) that happens to be of day of the week i, and $s'_i(t)$ is the normalized time series value assuming it is of day of the week i. This normalization step is designed to take away the effects of any trend or level shift that would affect the relative values for the different days of the week.

3. The seasonal average s_i^W (where i designates the day of the week) is given by

$$s_i^W = \mathrm{Median}_t(s'_i(t)) \qquad (6)$$

A distinctive feature of this deseasonalization algorithm is the use of the median instead of the average. Weekly seasonality involves in most cases sharp pulses. If using the average, the peak of the seasonal average becomes blunter and shorter, leading to detrimental results. The median, on the other hand, leads to preserving the typical shape of the peaks observed during the week. One can also have a distinct weekly average for each of the three seasonal regimes (the high season, the low season and very low season). In our implementation we used this latter approach.

The deseasonalization is then obtained as

$$s_{\mathrm{des}}(t) = \frac{s_{i,\,\mathrm{des1}}(t)}{s_i^W} \qquad (7)$$

where $s_{\mathrm{des}}(t)$ represents the final deseasonalized series. The forecasting model will be applied on this time series. Of course, after the forecasting step all these normalizing factors will be multiplied back to restore the seasonal effects of the forecasted portion of the time series.

As mentioned, we obtain the classification of the days of the year to the different seasonal regimes from the hotel managers. It is generally possible to estimate the high, medium and low seasonal periods exclusively from the data using some statistical technique. However, we believe that the information provided by hotel managers should take precedence. They are more knowledgeable about future events, such as a convention or a sports event. We therefore believe that the best approach is to obtain the seasonal average using a combination of information provided by the managers (to obtain the major high and low season periods and relevant future city events) and statistical approaches (to obtain the precise seasonal level within each season and to obtain the weekly seasonal average).

Reservations

Reservations represent the amount of bookings that arrive with time for a particular arrival day. It is a central variable in our whole simulation approach. But, at the same time it is also the most challenging component to model. The reason is its dependence on two time indexes: the reservation day (that is, the day the room is booked) and the arrival day (the intended day for the guest to check-in). A possible way to model reservations arrivals is to use a Poisson process. Because of the discrete nature of the way the data are recorded (that is, by days) this approach can lead to difficulties and rounding inaccuracies. So we opted for a different approach whereby we model the reservations arrivals as Bernoulli trials.

Let $B(i, t)$ be the expected number of reservations for arrival day t that are booked exactly i days before arrival. (In that sense $B(0, t)$ represents the expected number of walk-in guests.) We call $B(i, t)$ the *reservation curve*. Because of the random nature of the reservations process, more or less reservations than $B(i, t)$ will actually occur. Among the guests that could potentially come some reservations would materialize and some would not. As such, we assume that reservations obey a binomial distribution with probability p. Thus,

$$B(i,\ t) = Np \tag{8}$$

where N is the size of 'potential' population from which possible reservations can come for the particular reservation and arrival dates, and p is the probability that the reservation will materialize. The previous equation is for the purpose of preserving the fact that the mean of this binomial experiment should equal $B(i, t)$ (as by definition $B(i, t)$ is the expected number of reservations).

The main issue here is to estimate $B(i, t)$. Once available, then one can simply generate future reservations data by generating Bernoulli trials. The problem, however, is that for each arrival time t we have only one realization of the reservation curve $B(-,\ t)$, and it will, of course, be grossly inaccurate to base the estimate on that single realization. However, beyond up and down shifts with the seasonality variations, the shape of the reservation curve does not change much. We have verified that by visually screening the data. However, we found that this uniformity is up to a certain limit. That is, the actual shape would change somewhat between the extreme seasonal conditions.

To separate the effect of 'shape' from 'level' (of the reservation curve), we assume

$$B(i,\ t) = s(t)B'(i) \tag{9}$$

where $B'(i)$ is the *normalized reservation curve* (it sums to 1, that is $\sum_{i=0}^{\infty} B'(i) = 1$), and it represents the absolute shape of the reservation curve, ignoring the effect of its magnitude. The variable $s(t)$ represents the level or magnitude of the reservation curve. It more or less represents how the seasonal effects adjust the level of the reservation curve by shifting it up or down in a multiplicative manner.

Based on the above realization that the shape varies somewhat between different seasonality regimes, we assume that there are three distinct shapes of the normalized reservation curve $B'(i)$: one pertaining to the high season, $B'_H(i)$, one pertaining to the low season $B'_L(i)$, and another representing the very low season $B'_{VL}(i)$. To estimate these quantities, we average over the days occurring in the respective seasonal regimes:

$$\hat{B}'_H(i) = \frac{1}{N_H} \sum_{t \in S_H} \frac{R(i,\ t)}{\sum_{j=1}^{\infty} R(j,\ t)} \tag{10}$$

where S_H is the set of high season days, as defined in Seasonality subsection (let its size be N_H), and $R(i,\ t)$ denotes the actual number of reservations for arrival day t that are booked i

days before arrival. In contrast to $B(i, t)$, which is the *expectation* and is an *unknown* quantity, $R(i,\ t)$ is the actual number of reservations and it is an observed quantity. Note that we likewise normalized $R(i,\ t)$ in the above summation (in equation (10)) so that we focus on the shape rather than level. We have similar equations for the other two regimes:

$$\hat{B}'_L(i) = \frac{1}{N_L} \sum_{t \in S_L} \frac{R(i,\ t)}{\sum\limits_{j=1}^{\infty} R(j,\ t)} \qquad (11)$$

$$\hat{B}'_{VL}(i) = \frac{1}{N_{VL}} \sum_{t \in S_{VL}} \frac{R(i,\ t)}{\sum\limits_{j=1}^{\infty} R(j,\ t)} \qquad (12)$$

where S_L and S_{VL} are the sets of, respectively, low season days, and very low season days. The sizes of these sets are, respectively, N_L and N_{VL}.

Concerning the level multiplier $s(t)$, it is estimated as follows:

$$\hat{s}(t) = \sum_{i=0}^{\infty} R(i,\ t) \qquad (13)$$

The reason for the previous equation is that by summing both sides of equation (9) over the index i, we get

$$\sum_{i=0}^{\infty} B(i,\ t) = s(t) \qquad (14)$$

(because $\sum_i B'(i) = 1$). Since $R(i,\ t)$ is a realization from a distribution whose mean is $B(i, t)$, then equation (13) can be considered an estimate of $s(t)$. Note that it is fair to assume that the reservations process has little serial correlation (with i), thus making the estimate in equation (13) reasonably accurate. Of course, in all the previous summations the upper limit will practically be some bound I (rather than ∞) beyond which no reservations usually come. Figure 1 shows the normalized reservation curve $B'(i)$ as a function of the booking horizon i (time before arrival) for the three seasonal regimes, as estimated from the in-sample period for our Plaza Hotel case study. We can see that in the low season and especially in the very low season regimes more reservations come immediately before arrival day, which is an expected observation. Please note that the random fluctuations in the curve (including the blips at times 8 and 24) are because of a statistical estimation error, as we are dealing with a limited amount of data.

The estimated $\hat{s}(t)$, which represents the sum of all reservations for arrival day t (see equation (13)), represents a measure of room demand. This is the variable for which we apply the deseasonalization step, as detailed in the last subsection. After the deseasonalization step, we apply a forecasting method to project this variable forward in time. The reason for using this variable instead of the net bookings is that it purely handles reservations only. On the other hand, the net bookings variable takes away the cancellations, and it will therefore be hard to disentangle the two processes (that is, the reservations and cancellations processes) by observing only the arrival forecast.

As mentioned, for the purpose of simulating the reservations process, we consider a binomial distribution, with equation (8) guaranteeing equality of the mean of the distribution to the expected reservations number. There are however, two variables involved, N and p, and this could therefore allow us to fit an additional quantity (other than the mean) for more faithful representation. We took the additional quantity to be the variance. Hence, we set the two quantities N and p so that the following two equations are satisfied:

$$\hat{B}(i,\ t) = Np \qquad (15)$$

$$\frac{1}{TI} \sum_{t=1}^{T} \sum_{i=1}^{I} \left(\hat{B}(i,\ t) - R(i,\ t)\right)^2 = Np(1-p) \qquad (16)$$

where $\hat{B}(i, t) \equiv \hat{s}(t)\hat{B}'(i)$ is the estimate obtained using the procedure described above (equations

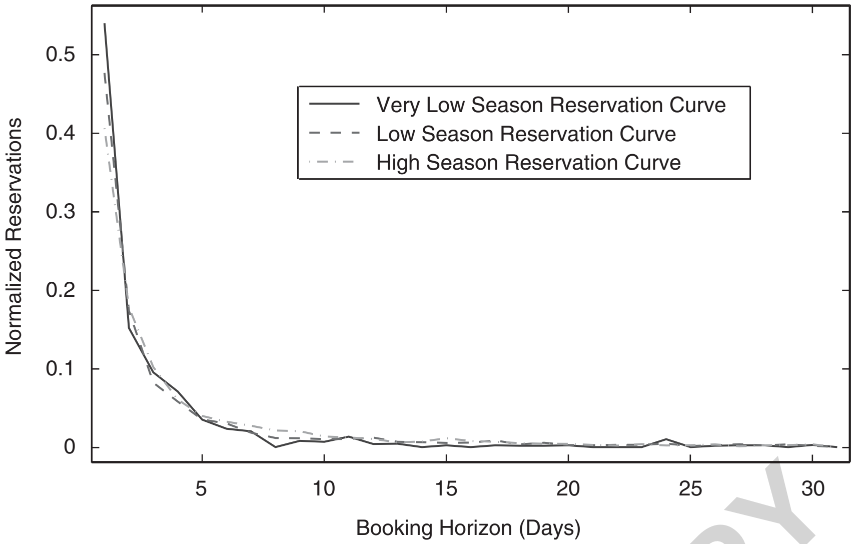

Figure 1: Plaza hotel's normalized reservation curve $B'(i)$ as a function of the booking horizon (time before arrival) for the three seasonal regimes.

(10)–(13)). Note that for each i and t, we have distinct N and p. The second equation specifies that the variance of the binomial process equals the empirical variance observed from the data. Notice that this empirical variance is computed using all arrival times t and all number of days i before arrival (or else, if we assume a variable variance, data will not be sufficient). So, in summary, we will have an estimate of $\hat{B}(i, t)$, and from that we generate the reservations. So, we estimate the appropriate N and p from (15) and (16). We then generate N Bernoulli trials with probability p to obtain the number of reservations for that particular i (number of days till arrival) and t (arrival day).

Cancellations

Reservations can be canceled any day before arrival day. The rates of cancellations vary according to the time until arrival. Typically, they increase as we get close to arrival day. However, if there are some penalties for cancellations that occur beyond a certain day, cancellations will decrease dramatically. We assume that the cancellation rate (say $c(i)$) is a function of the number of days i until arrival day. It is defined as the mean fraction of net bookings that get canceled. For example, consider that we are focusing on arrival day t, and that we are at ith days before that arrival day. Assume that at the close of the previous day there are $H(i+1, t)$ bookings (or reservations at hand) for that arrival day t. If $c(i) =$ 5 per cent, then the expected number of reservations to be canceled at the current day (day i before arrival) is 0.05 $H(i+1, t)$. Also, as a result, $c(0)$ represents the mean fraction of no-shows plus last day's cancellations.

Of course, $c(i)$ gives only the mean value. The actual number that ends up being canceled is a random variable. We model that random variable as binomial. Specifically, we assume $H(i+1, t)$ Bernoulli trials, each one with probability $p \equiv c(i)$ of being canceled. The cancellation mean curve $c(i)$ is estimated from the reservations data of the in-sample period. Note that we have assumed a similar cancellation mean curve $c(i)$ for all seasonal regimes. The reason is that the estimate of $c(i)$ turned out to be a little noisy. Disaggregating it among the three main seasonal regimes will aggravate the estimation error. Figure 2 shows the cancellation mean curve $c(i)$ for the case study of Plaza Hotel, as estimated from the in-sample period.

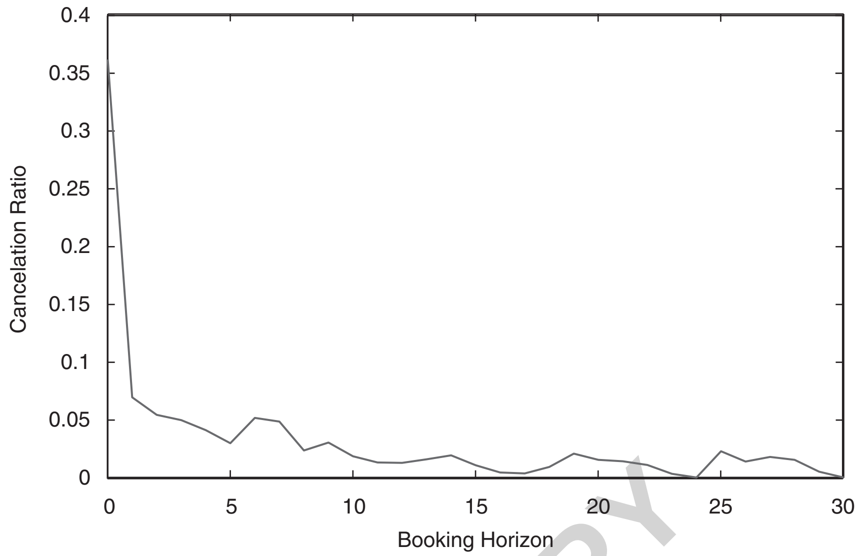

Figure 2: Cancellation rate $c(i)$ as a function of the booking horizon (time before arrival) for Plaza hotel.

Once $c(i)$ is estimated, it is used when we simulate the reservations and cancellations processes forward in time using the aforementioned binomial generation process.

Length of stay

The LOS (call it T_L) for hotel guests varies according to the type of the hotel's clientele. Business travelers tend to stay for one or two nights, whereas vacationers could stay up to a week. The LOS plays a major role in our simulation system. In fact, the LOS can actually impact the hotel occupancy of the near future, as well as lead to denials of booking attempts (even for future days). When tracking an arriving reservation until it materializes and the guest arrives, the LOS has to be specified. Towards this end, we consider a distribution of the LOS, and estimate it from the in-sample portion of the data. Then, in the simulations phase we generate an actual stay scenario for every reservation using this distribution.

It is conceivable that T_L would-be influenced by some factors. In our work with hotel data, we observed that typically the time from booking to arrival does not impact the LOS. That leaves one potential influence factor: the seasonal cycle. To test this possible dependence, we have estimated $p(T_L | S_H)$, $p(T_L | S_L)$ and $p(T_L | S_{VL})$, that is the distribution of the LOS in each of the three seasonal regimes. Figure 3 shows these curves for the Plaza Hotel case study. One can see that these distributions differ somewhat. We therefore decided to use in our model these season-specific LOS distributions. Day of the week versus weekend arrival should also influence the LOS distribution, but for simplicity we did not take it into account in the current version.

One could also in principle model understays and overstays. One way to do that is to estimate a distribution of the number of days the guest stays more (or less) than he reserved. We did not model this aspect in our simulation as it is a bit involved, and its impact on accuracy will probably be limited.

Group reservations

A large amount of tourism travel nowadays is through pre-arranged tour packages. This means that the tourist has a planned itinerary, with stays at specific hotels for specific dates. This way the tour operator can achieve block

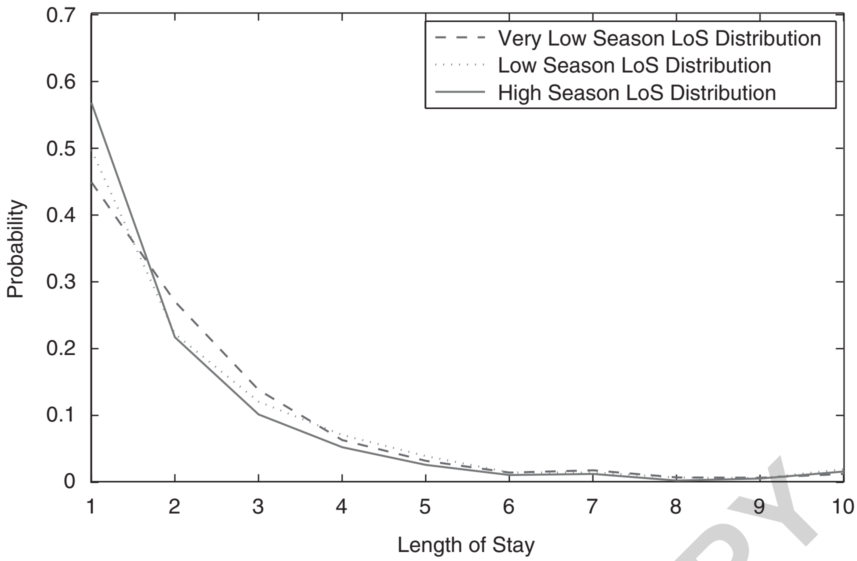

Figure 3: Length of stay distributions for the different seasonal regimes for Plaza hotel.

reservations in the hotels and hence obtain a lower cost that can be passed on to the traveler. Kimes (1999) performed an insightful analysis and developed a forecasting model for group reservations in hotels. Group reservations have their specific dynamics, which we consider in this simulator. For example, we allow whole block cancellations. We define a block or a group as a group of reservations that are reserved at the same day, for the same arrival and departure dates, and by the same travel operator. In our system, we model the group size g by some distribution $p(g)$. This distribution is estimated from the historical data by checking the sizes of all the reservation blocks. Actually, we lump all group and non-group data into one set and estimate the distribution from this set. In such a case, $g=1$ represents an individual (or non-group) reservation, and $p(1)$ represents the probability or fraction of non-group reservations.

Once estimated, in the simulation stage the group reservations are generated as follows. We generate a reservation according to the estimated reservation curve (as described in the Reservations subsection). This reservation could be a group (of some specific size) or a non-group reservation. Then, we generate a number g using the distribution $p(g)$. This generated g will then represent the group size. It could equal 1 (actually with a high probability), which simply means it was just an individual (or non-group) reservation.

Trend estimation

As mentioned, the variable $s_{\text{des}}(t)$ represents some measure of deseasonalized room demand (where we exclude the cancellations' effect). It is also possible to use the $s_{\text{des}}(t)$ time series to gauge trends in overall room demand. For this reason, we apply a forecasting model for predicting this variable (that is, $s_{\text{des}}(t)$) in the considered forecast horizon. Our ultimate goal is to simulate the room reservations process in some forecast horizon. The variables used should therefore reflect future values, rather than present values. It is anticipated that the general room demand will exhibit some medium-term trend, owing to changing hotel conditions, and external effects that affect tourism demand and business conditions in the area. As such, this forecasting step should estimate this trend. However, the overall reservation arrivals (as exemplified by $s_{\text{des}}(t)$) are not the only variable in our model. There are other parameters in the model, mainly distributions and mean values (such as the normalized reservation curves, the cancellation curve, the LOS and so on), and one may also hope to use values of these parameters that reflect the future rather than the present. However, we argue that they are not anticipated to change in the medium term (beyond changes because of the seasonality effect discussed before). These quantities reflect customer behavior issues that typically change only in the long term. We have verified this claim, by estimating these distributions and averages in two contiguous 6-month periods for our case study. We found that the estimates are close. So, in summary, other than the overall demand $s_{des}(t)$, the present estimates of the other hotel parameters are perfectly suitable for use in the forecast period, and no forecast step is needed to project forward these parameters.

We use Holt's exponential smoothing model for forecasting the room demand variable $s_{des}(t)$. The Holt's exponential smoothing model is based on estimating smoothed versions of the level and the trend of the time series Hyndman *et al* (2008). Then, the level plus the trend is extrapolated forward to obtain the forecast. The governing equations for updating the trend and the level are given by Gardner (2006):

$$l_t = \alpha s_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{17}$$

$$b_t = \gamma(s_t - l_{t-1}) + (1-\gamma) b_{t-1} \tag{18}$$

where $s_t \equiv s_{des}(t)$ is the variable to be forecasted (room demand variable), l_t is the estimated level and b_t is the estimated trend or slope of the time series. The forecast is given by a linear extrapolation in time:

$$\hat{s}_{t+m} = l_t + m b_t \tag{19}$$

There are five parameters that have to be set, before applying the forecasting step. These are α and γ, the smoothing constants for, respectively, the level variable and the trend variable, l_0 and b_0, the starting values for, respectively, the level and trend, and σ, the standard deviation of the error term. We used the approach by Andrawis and Atiya (2009) that is based on the maximum likelihood concept.

Once the forecast of $s_{des}(t)$ is obtained for the required horizon, the seasonal effects will be restored back, to obtain $\hat{s}(t)$. Then, the detailed reservation curves $B(i, t)$ for the future can be constructed (as they are the product of the normalized reservation curves $B'(i)$ and the forecasted $\hat{s}(t)$ variable, see equation (9)). Then we can generate future reservations according to the forecasted reservation curves $B(i, t)$ with the help of Bernoulli trials as detailed in Reservations subsection section using N and p variables appropriately estimated from (15) and (16).

THE OVERALL MONTE CARLO SIMULATION SYSTEM

Once we have estimated the parameters such as the seasonal average, the reservation curve and so on, as detailed in the last section, we can now apply the forecast step. In this step, we simulate forward the processes of reservations arrivals, cancellations and so on, exactly as they happen in the model that we have developed. We use all the parameter values obtained in the estimation step. Because of the randomness aspect of the reservations process one realization of this simulation is naturally not sufficient. We need to generate many paths in a Monte Carlo fashion, and then take the mean of these paths at any future instant of time t as the forecast. This applies to whatever quantity we would like to forecast, such as reservations arrivals or occupancy.

When at a particular day a forecast is needed for some horizon, we make use of the information that we have about the reservations already at hand. This will be the starting point upon which reservations will keep building. For example, we are at time t and we would like to forecast arrivals at time $t+5$. Assume that the hotel has already 20 reservations for that future date. Then any new reservations that will be simulated forward will add to these 20 reservations. Conversely, any future cancellations simulated will be subtracted out from these 20 reservations. These starting reservations could significantly influence the forecasted variables, with this effect slowly decaying as the lead time increases. Below are the forecasting algorithm's details:

Algorithm Demand Forecasting:

1. Let t be the current time and $t+1: t+T$ be the horizon to be forecasted. From the hotel records, we know that we have $R(i, t')$ reservations for arrival day $t' = t+1, \ldots, t+T$, $i \geqslant t'-t$.
2. We know the variable $s(t')$ for all previous times $t' \leqslant t$ (it is defined by equation (13)). Forecast $s(t')$ for the considered horizon, as discussed in Trend estimation section. Let the forecasts be $\hat{s}(t')$, $t' = t+1, \ldots, t+T$.
3. For $\tau = t+1$ to $t+T$ perform the following:

   (a) *Cancellations:* Generate cancellations from a binomial distribution, as follows. For every arrival day t', $t' = \tau, \ldots, t+T$ the bookings (or reservations at hand) are $H(t'-\tau+1, t')$. The average fraction of cancellations is $c(t'-\tau)$ where the c function has been already estimated from the historical data in the parameter estimation phase. We generate a Bernoulli trial for each reservation with probability $p = c(t'-\tau)$ that a cancellation will actually occur for that reservation. Remove the canceled reservations and correspondingly update the new booking matrix $H(t'-\tau, t')$.
   (b) *Reservations:* Generate new reservations for every arrival time t', $t' = \tau, \ldots, t+T$. The reservation curve for some time t' will be one of the normalized reservations templates: $B'_H(i)$, $B'_L(i)$ or $B'_{VL}(i)$ (depending on which seasonal regime t' falls in) multiplied by the forecasted level $\hat{s}(t')$. Generate the reservations using a binomial distribution with number of trials N and probability p determined according to (15) and (16).
   (c) *Group Reservations:* For every generated reservation determine the group size by generating a number (per reservation) according to the group size distribution. Note that if this number turns out to equal 1 then this means it is an individual reservation (which is usually the higher probability case).
   (d) *Length of Stay:* For every reservation generate an LOS according to the estimated season-specific LOS distribution.
   (e) For every reservation generated in steps (b)–(d) determine if it will be accepted or denied. A reservation is accepted if during the intended duration of stay there is room availability, else it is denied. If the hotel has an overbooking policy, then a reservation is accepted if it is within the bounds of this overbooking policy, else it is denied. Note that steps (b)–(d) for generating a reservation, with all its features including the decision of whether to accept or deny the reservation, have to be performed in a sequential manner, one reservation at a time.

4. Repeat step (3) for K times to get K Monte Carlo paths for future reservations. Obtain the mean (or median) for these paths for each of the arrivals variable and the occupancy variable (mean or median over the K paths for each time step). These are the forecasts.

MISCELLANEOUS ASPECTS

Level of aggregation

Many RM systems prefer to have the demand forecast segmented by category (Vinod, 2004). As detailed before, a pricing policy on the basis of various guest categories is at the heart of a successful RM system. Categories, such as by rate, guest type, room type and LOS are usually considered in most hotels. One procedure (disaggregate forecasting) is to consider each category's guest flow separately, and develop a separate forecasting model for each category. The alternative procedure (aggregate forecasting) is to develop an aggregate model and then disaggregate by breaking down the aggregate forecast into its disaggregate constituents in some reasonable way (Weatherford and Kimes, 2003). In our simulation-based model, the latter could be performed as follows. We estimate a common set of parameters for all categories using all the available historical data set. When forecasting, each category has its own set of reservations at hand. Starting from these reservations, we perform a simulation forward in time to obtain the forecasts for each category (the forecasts will generally be different for each category because of the different starting reservation numbers).

Each of the disaggregate and the aggregate approaches have their own strong and weak points. For example, the advantage of the disaggregate approach is the specificity of the estimated set of parameters to the considered category. The disadvantage, however, is that by considering each category separately, the data could get considerably diluted to the extent that the parameter estimates would-be of suspect accuracy. Of course, a middle ground could be taken, by combining some categories into a few major ones that are known to possess parameter sets different from each other. Then, we disaggregate further in the forecast step. For example, we could categorize as business guests versus leisure, and/or high rate versus low rate. These dichotomies would probably have different reservation curves, and different cancellation and LOS profiles.

Unconstrained demand forecasting

The reservations data that are typically recorded in the hotel's books do not entirely gauge the whole amount of room demand. Some would-be guests have attempted to reserve, but were turned down because of lack of availability or because of the booking limit imposed on the relevant category: these are called *denied reservations*. Unconstrained demand is the total demand including these denied reservations. In other words, it is the total amount of reservations that would have come, had the hotel accepted every arriving reservation attempt. (An analogous definition applies for arrivals as well.) For the purpose of RM, unconstrained demand is the more relevant quantity to consider. The problem is that denial data are usually not recorded in the hotels, or if recorded, they are considered to be unreliable. Even if there were attempts to record denials, it will be very hard to determine if an inquiry about room availability would have eventually led to an actual reservation, had the answer been positive instead of negative. For these reasons, the problem of forecasting of unconstrained demand is a very challenging issue.

The dominant approach in the literature has been to assume certain distributions for the variables and use the concept of censoring (see Lee (1990) and Weatherford and Kimes (2003)). For example, Lee (1990) modeled airline reservation arrivals and cancellations as Poisson processes. Subsequently, he used the maximum likelihood approach to obtain the parameters of the distribution, taking into account denials as censored information when expressing the likelihood. Liu *et al* (2002) developed a parametric regression model to estimate unconstrained demand in hotels in the presence of censoring. Another approach is to use the concept of detruncation. Skwarek (1996) used pick-up detruncation whereas Wickham (1995) used booking curve detruncation. These approaches are based on estimating the number or fraction of reservations that were denied. Another detruncation approach was proposed by Zeni (2001), who used the expectation maximization (EM) algorithm to iteratively apply the 'E'-step (replacing censored information by their expectation), and the 'M'-step (re-estimating the parameters of the distribution, given the newly unconstrained data). For a comparison and discussion of various unconstraining methods for hotels refer to the work of Queenan *et al* (2009).

We extended the proposed Monte Carlo approach to the unconstrained forecasting case. The approach follows more a detruncation-type methodology, as the other censoring approach will become analytically intractable. The steps of the proposed method are as follows:

Algorithm Unconstrained Forecasting:

1. Estimate the parameters, distributions and so on, exactly as detailed in Estimation of the system's components section using the constrained historical data. Let this set of parameter values be *S*.
2. Simulate Monte Carlo paths as detailed in The overall Monte Carlo simulation system section (Algorithm Demand Forecasting), steps 1–3 based on the parameter values *S* that are obtained in the previous step (using the constrained data). Simulate these Monte Carlo paths on the historical (or in-sample) period (*not* on the forecasting period). We implement steps 1–3 exactly except that we assume unlimited hotel capacity. This means no simulated reservations will be denied.
3. Re-estimate the parameters as detailed in Estimation of the system's components section using the reservations data generated in the previous step (step 2 of this algorithm). Note that unlike how it is done in step 1 in this algorithm or in Estimation of the system's components section, we have here many Monte Carlo paths (in other words reservation paths). We make use of all these reservation scenarios in the parameter estimation step. This can only enhance the accuracy (relative to an estimate using only one reservation scenario).

THE PLAZA HOTEL CASE STUDY

We applied the proposed forecasting model to the problem of forecasting the arrivals and the occupancy of Plaza Hotel, Alexandria, Egypt, as a detailed case study. Plaza Hotel is a mid-sized four-star sea-side hotel, located on the Mediterranean Sea. It has 134 rooms partitioned into 11 single rooms (studio type), 91 double rooms, 32 suites (12 junior suites and 20 deluxe suites). Its clientele is a mix of business, leisure and foreign tourist guests. The type of business guests covers conferences, government, sporting clubs, corporations and so on.

Modern RM approaches (the type discussed earlier) have recently started to attract some interest from hotels in Egypt. A good fraction of five star hotels in Egypt apply some form of RM (Shehata, 2005). On the other hand, the majority of four star hotels and lower do not apply any form of RM. Plaza Hotel plans to implement a RM system. As a first step, an arrivals and occupancy forecasting model needs to be developed. In collaboration with the hotel, we have applied our proposed forecasting model to the hotel's data.

SIMULATION RESULTS

The hotel data

We have applied the proposed Monte Carlo simulation forecasting model on the data of Plaza Hotel. We have obtained a full set of data covering the period from 1 October 2006 until 1 March 2008. The breadth of the data is extensive, and they include all aspects of the reservations, with all its details, such as room type, customer category, rate category and so on.

We have considered a 3-month-ahead forecast, using an expanding window approach for the estimation or in-sample set. In that set up, we forecast 3 months ahead at three snapshots during the last 5 months of the data. Table 1 shows the in-sample periods and the forecast periods for the three snapshots. We considered both daily forecasting (that is, forecasting every day of the 3-month forecast horizon), as well as weekly forecasting (forecasting only week-by-week in the 3-month forecast horizon). In practice, the hotel will probably be

Table 1: The in-sample and the 3-month-ahead forecast periods for the three forecasted snapshots

Snapshot no.	*In-sample period*	*Forecast period*
1	1 October 2006 – 30 September 2007	1 October 2007 – 31 December 2007
2	1 October 2006 – 31 October 2007	1 November 2007 – 31 January 2008
3	1 October 2006 – 30 November 2007	1 December 2007 – 29 February 2008

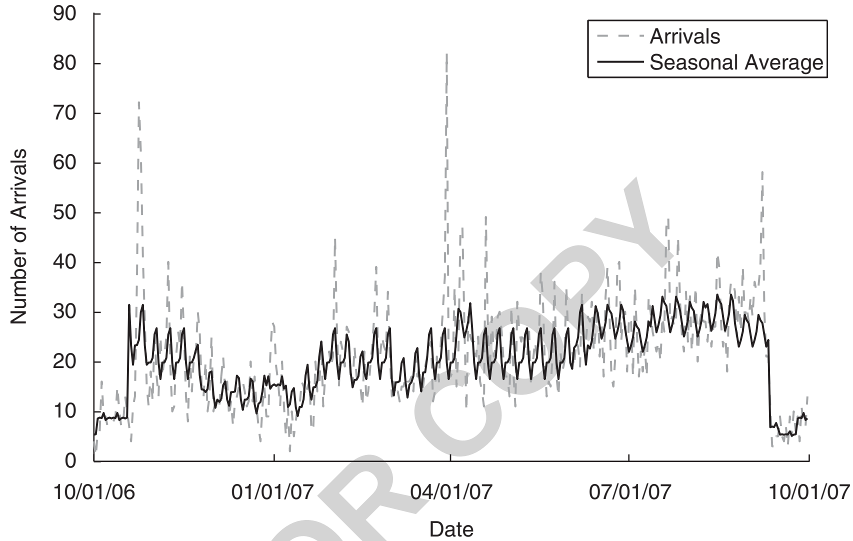

Figure 4: The arrivals time series for the in-sample period, together with the seasonal average.

interested in daily forecasting for the short period ahead, followed by weekly forecasting for the farther period ahead. We have considered forecasting aggregate room demand (arrivals as well as occupancy), rather than disaggregate by room type or other category. For Plaza Hotel, the partition of rooms by type is not very rigid, as rooms get frequently converted from one type to another according to need.

Seasonal analysis

As mentioned in Introduction section, we have used a tri-level seasonal regime classification as a basis, with the levels being: high season, low season and very low season. The reason for the third 'very low season' is that the period of the month of Ramadan is exceptionally low in demand. The month of Ramadan is the holy month in the Islamic calendar that precedes one of the major Islamic feasts, the 'Eid El-Fetr'. The appendix lists the dates of the different seasonal regimes, as identified by the Plaza Hotel managers. Typically, the dates of most events whether religious feasts, conventions or sporting events are well-known months in advance.

Plaza Hotel also possesses a distinct day of the week pattern of arrivals. Alexandria is about 230 km away from Cairo, the capital and most populous city in Egypt. It therefore attracts visitors over the weekend (the weekend holiday in Egypt is Friday and Saturday).

Figure 4 shows the arrival time series for the in-sample period, together with the estimated seasonal average. When we perform the forecast, we check the classification of the regions of the forecast period, and use the seasonal averages for the corresponding regions of the in-sample period. For example, if part of the forecast period includes the month of Ramadan, we use the seasonal average of Ramadan as estimated from the in-sample set. As mentioned in Seasonality subsection, the best seasonality analysis approach is an interactive approach, whereby we make use of the information provided by the hotel managers, and pose it in a statistical framework. For example, we made use of the classification into the three seasonal regimes supplied by the hotel managers. But, as mentioned in Seasonality section, we augmented that with a finer subpartition of these three seasonal regimes that is determined by the data. The reason is that some of the high season period cannot be treated equally (for example, Eid El-Fetr typically exhibits a higher hotel traffic than other high season periods). Also, the low season is too large to simply have a one-level seasonal average. Nevertheless, the classification provided by the hotel managers is indispensible. Consider, for example, a convention that is held every year, but not at an exactly fixed date. The hotel manager would identify these dates, and the seasonal index of the future convention would be based on the level of arrivals of the past conventions. Another example is an expected non-repeatable event, that hotel managers would flag as similar to some other past high-season events. A purely mechanical approach would not be able to get the seasonality right in these cases.

The compared models

To obtain a comparative idea about the relative performance of the proposed forecasting model, we have applied to the same data five competing forecasting models. As mentioned in Related work section, there are two major categories of forecasting models. The first one, the time series approach, considers only the arrivals or the occupancy time series, and applies time series forecasting models on these. The second one, the reservations-based or the pick-up approach, models the amount of reservations to be 'picked up' until arrival day.

We have considered one model from the first category, and four models from the second category. The five compared models are as follows:

1. A time series model using Holt's exponential smoothing using the maximum likelihood approach by Andrawis and Atiya (2009) for estimating the parameters (see Trend estimation section for an overview over Holt's exponential smoothing). We used a deseasonalization strategy similar to the one we used in our proposed model (as described in The hotel data section).
2. Additive classical pick-up using simple moving average.
3. Additive advanced pick-up using simple moving average.
4. Additive classical pick-up using exponential smoothing.
5. Multiplicative classical pick-up using exponential smoothing.

The main idea of the pickup method is to estimate the amount of reservations (net of cancellations) that are expected to come from now till arrival day (the day to be forecasted). This expectation is computed from the historical reservations data. One version of the pick-up is the so-called additive, whereby the increments of reservations are assumed to add up linearly. Therefore to obtain the forecast, we simply add the expected number of net reservations to be picked up to the current number of bookings. The other version is the multiplicative pick-up. This method assumes that future net reservations are proportional to the current booking. Therefore, to get the forecast, current bookings are multiplied by the average pickup ratio. Another classification is classical versus advanced. By classical, we mean that we use only completed booking curves in computing the average number of reservations to be picked up. Conversely, by advanced we mean that we use not-yet-completed booking curves. A completed booking curve means the cumulative net reservations of previous arrival days, as they are complete and no more reservations are expected to add up. On the other

Table 2: The overall forecast error for the out-of-sample periods for the proposed Monte Carlo model and the five competing models for the case of daily forecasting

Model	*Arrivals SMAPE*	*Occupancy SMAPE*
Proposed Monte Carlo	43.9	37.7
Pick-up (Add, Class, Simple)	48.3	–
Pick-up (Add, Adv, Simple)	47.9	–
Pick-up (Add, Class, Exp)	54.8	–
Pick-up (Mul, Class, Exp)	97.9	–
Exp Smoothing	63.8	61.1

Table 3: The overall forecast error for the out-of-sample periods for the proposed Monte Carlo model and the five competing models for the case of weekly forecasting

Model	*Arrivals SMAPE*	*Occupancy SMAPE*
Proposed Monte Carlo	21.5	23.4
Pick-up (Add, Class, Simple)	23.1	–
Pick-up (Add, Adv, Simple)	22.5	–
Pick-up (Add, Class, Exp)	35.1	–
Pick-up (Mul, Class, Exp)	100.7	–
Exp Smoothing	49.2	41.2

hand, not-yet-completed refers to future arrival days (which are expected to still accumulate more reservations). Simple/exponential smoothing corresponds to the way we compute the average reservations to be picked up. Note that Zakhary *et al* (2008) conducted a comparison between the different variations of the pick-up approach and found models (2), (3) and (5) have been among the top three models in forecasting performance. The disadvantage of the pick-up approach is that it applies only to arrival forecasting, but *not* to occupancy forecasting.

We used as an error measure the symmetric mean absolute percentage error, defined (whether for the arrivals or occupancy time series) as

$$\mathrm{SMAPE} = \frac{1}{M}\sum_{j=1}^{3}\sum_{m}\frac{|\hat{y}_m^{(j)} - y_m^{(j)}|}{(|\hat{y}_m^{(j)}| + |y_m^{(j)}|)/2} * 100 \qquad (20)$$

where $y_m^{(j)}$ and $\hat{y}_m^{(j)}$ are, respectively, the actual time series value and the forecast for forecast period j (as mentioned there are three-month-ahead forecast periods, listed in the last column of Table 1). Also, M is the total number of points that are forecasted (the sum of the points in the three-month forecast periods).

Results

Table 2 shows the SMAPE error measure for the proposed Monte Carlo model and the five competing models for the case of daily forecasting. Similarly, Table 3 shows the SMAPE error measure for the proposed model and the competing models for the case of weekly forecasting. Also, Figure 5 shows the forecast of the proposed Monte Carlo model versus actual arrivals for the first 3-month-ahead forecast period. Also shown are the one standard deviation confidence bands. Figure 6 shows the forecast and the actual (with the confidence bands) for the occupancy for the first 3-month-ahead forecast period (also for the proposed Monte Carlo model).

One can deduce the following observations. The proposed Monte Carlo model beats all the competing models for both the arrival forecasting and the occupancy forecasting

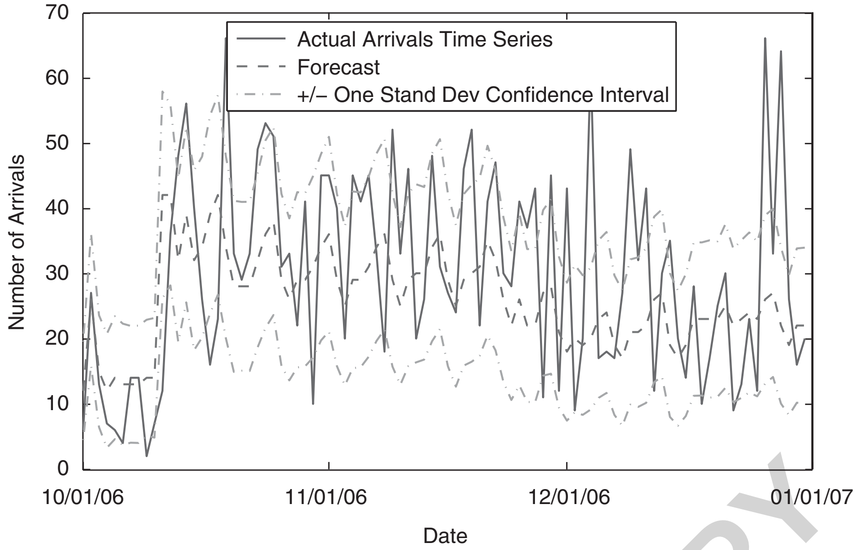

Figure 5: The forecast versus the actual for the arrivals for the first 3-month-ahead forecast period.

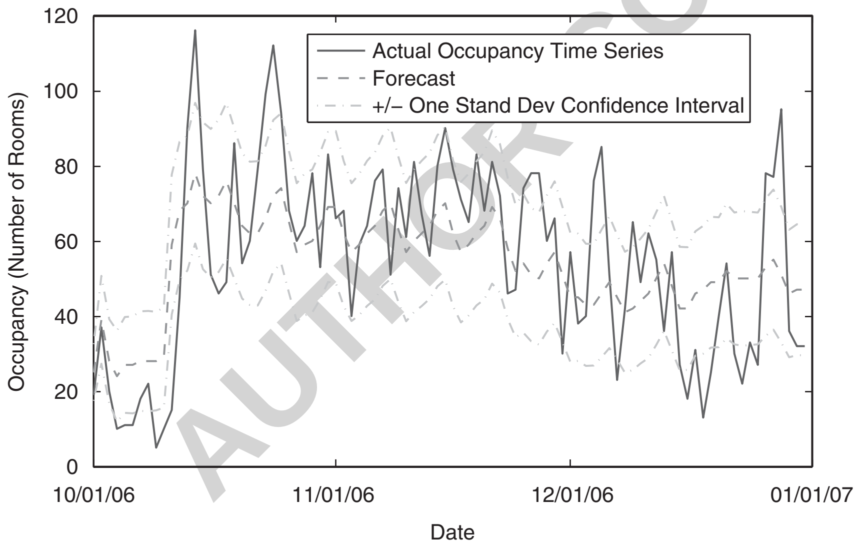

Figure 6: The forecast versus the actual for the occupancy for the first 3-month-ahead forecast period.

problems, and for the daily forecasting, as well as the weekly forecasting problems. For the arrival forecasting, the outperformance is considerable, when compared with the Holt's exponential smoothing model and with two of the pick-up models. The two most competitive pick-up models, the additive classical pick-up using simple moving average and the additive advanced pick-up using simple moving average, are still about 4 per cent behind the proposed model in SMAPE for the daily arrival forecasting, and 1–1.5 per cent behind for the weekly arrival forecasting.

As seen in the table, the proposed model is the undisputed winner for the occupancy forecasting case. Even though the pick-up approach is simple, it has some severe limitations. Its major problem is its inapplicability to the occupancy forecasting problem. Occupancy is one of the central quantities in a RM system.

Its role becomes even more pronounced when dealing with overbooking. As mentioned before, other than forecasting accuracy the advantage of the proposed model is its versatility, and this warrants the additional complexity. It can basically obtain many of the quantities of interest or many desired computations, for example: unconstrained demand, a forecast of the number of denials, the impact of a particular overbooking strategy, the probability of reaching hotel capacity, and the density of the forecasted quantities. All these computations are highly desirable for a RM professional. In addition, it allows the RM professional to perform some scenario analysis, and assess the impact of possible policies on revenue. For example, Figures 7 and 8 show the distributions of the forecasted arrivals and occupancy

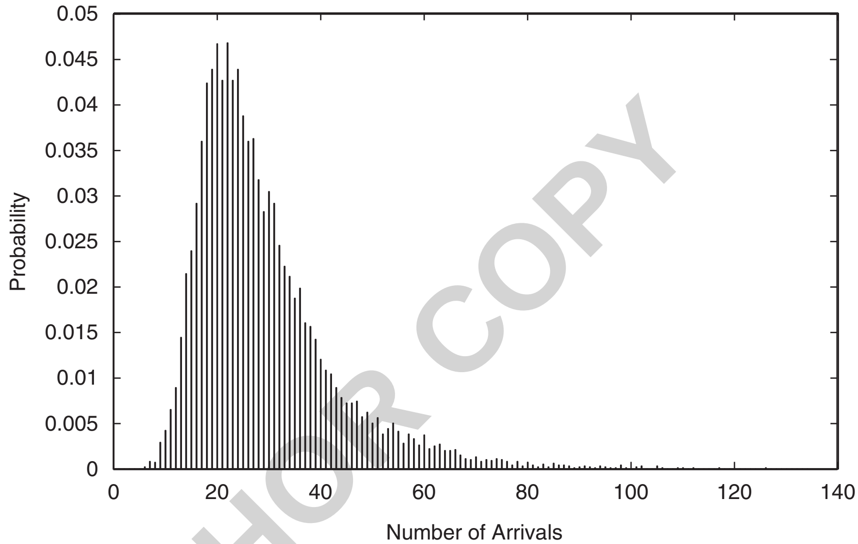

Figure 7: The forecast distribution of the arrivals for 30-day snapshot (each bar corresponds to a specific number of rooms).

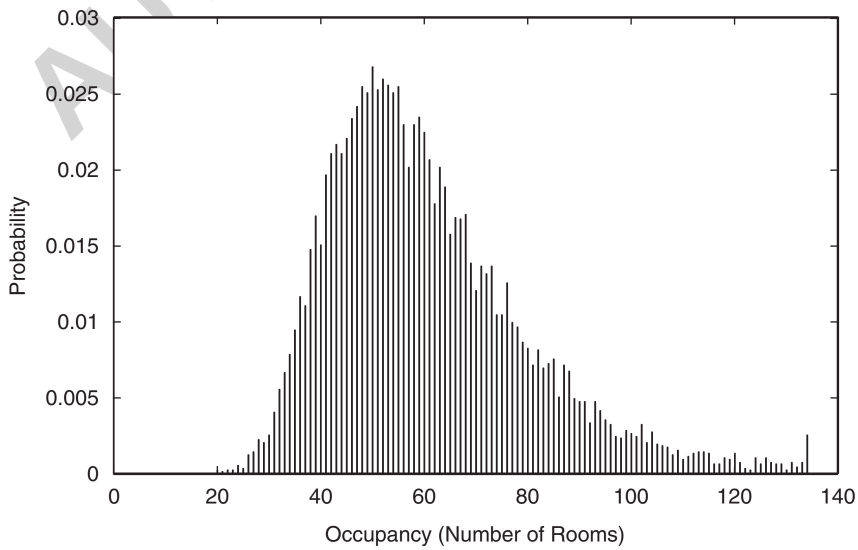

Figure 8: The forecast distribution of the occupancy for 30-day snapshot (each bar corresponds to a specific number of rooms).

(respectively) at a 30 day-ahead snapshot. From these figures one can discover interesting facts, for example the right skewness of the distribution, and the very low probability of getting less than seven or eight reservations (and an occupancy less than 20).

CONCLUSIONS

In this article, we have proposed a new model for hotel arrivals and occupancy forecasting using Monte Carlo simulation. The proposed model has two main phases. In the first phase, we estimate the parameters related to the reservations process. In the second phase, we simulate the reservations process forward in time, making use of the estimated parameters obtained in Phase 1. We considered as a case study the Plaza Hotel of Alexandria, Egypt. The proposed forecasting model achieves good forecasting accuracy and beats other competing forecasting models. It also exhibits other nice features, such as obtaining densities for any variable of interest. In other words, it estimates the whole picture of what will happen in the future for all processes, and in a probabilistic way.

ACKNOWLEDGEMENT

We acknowledge the help of Hossam Shehata of Alexandria University. His help has been invaluable in giving us information and insights about the hotel business, and in interfacing with Plaza Hotel managers. We acknowledge the help of Professor Hanan Kattara of Alexandria University (and the owner of Plaza Hotel), for her generous help and willingness to supply all Plaza Hotel's data. We thank Emad Mourad, the manager of Plaza Hotel, for his assistance. We acknowledge the help of Robert Andrawis of Cairo University, who has developed the maximum-likelihood-based exponential smoothing code. We also acknowledge the useful discussions with Professor Ali Hadi of the American University of Cairo and Cornell University. This work is part of the *Data Mining for Improving Tourism Industry Revenue in Egypt* research project within the Data Mining and Computer Modeling Center of Excellence in Egypt. This work is supported by the Information Technology Industry Development Agency (ITIDA) in Egypt through the Centers of Excellence Program.

REFERENCES

Andrawis, R. and Atiya, A.F. (2009) A new Bayesian formulation for Holt's exponential smoothing. *Journal of Forecasting* 28: 218–234.

Andrew, W., Cranage, D. and Lee, C. (1990) Forecasting hotel occupancy rates with time series models: An empirical analysis. *Hospitality Research Journal* 14: 173–181.

Bailey, N. (1964) *The Elements of Stochastic Processes with Applications to the Natural Sciences*. New York: Wiley.

Ben-Akiva, M. (1987) Improving airline passenger forecasts using reservations data. In Presentation at Fall ORSA/TIMS Conference, St. Louis, MO.

Ben Ghalia, M. and Wang, P. (August 2000) Intelligent system to support judgmental business forecasting – The case of estimating hotel room demand. *IEEE Transaction on Fuzzy Systems* 8(4): 380–397.

Bitran, G. and Caldentey, R. (2003) An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* 5: 203–229.

Bitran, G. and Gilbert, S. (1998) Forecasting for Airline Network Revenue Management: Revenue and Competitive Impacts. MIT Flight Transportation Lab Report R98-4.

Bitran, G. and Mondschein, S. (1995) An application of yield management to the hotel industry considering multiple day stays. *Operations Research* 43: 427–443.

Chiang, W.-C., Chen, J.C.H. and Xu, X. (2007) An overview of research on revenue management: Current issues and future research. *International Journal of Revenue Management* 1(1): 97–128.

Chow, W.S., Shyu, J.-C. and Wang, K.-C. (1998) Developing a forecast system for hotel occupancy rate using integrated ARIMA models. *Journal of International Hospitality, Leisure Tourism Management* 1: 55–80.

Franses, P.H. (1998) *Time Series Models for Business and Economic Forecasting*. Cambridge, UK: Cambridge University Press.

Gardner, E.S. (2006) Exponential smoothing: The state of the art – Part II. *International Journal of Forecasting* 22: 637–666.

Hyndman, R.J., Koehler, A.B., Ord, J.K. and Snyder, R. (2008) *Forecasting with exponential smoothing: The state space approach*. Berlin, Germany: Springer Verlag.

Ingold, A., McMahon-Beattie, U. and Yeoman, X.I. (eds.) (2003) *Yield Management: Strategies for the Service Industries*, 2nd edn. London and New York: Continuum.

Kimes, S.E. (1999) Group forecasting accuracy for hotels. *Journal of the Operational Research Society* 50: 1104–1110.

Lee, A. (1990) *Airline Reservations Forecasting*. Erewhon, NC: Prentice-Hall.

L'Heureux, E. (1986) A new twist in forecasting short-term passenger pickup. In: Proceedings of the 26th Annual AGIFORS Symposium.

Liu, P., Smith, S., Orkin, E. and Carey, G. (2002) Estimating unconstrained hotel demand based on censored booking data. *Journal of Revenue & Pricing Management* 1: 121–138.

Liu, S., Lai, K.K., Dong, J. and Wang, S.-Y. (2006) A stochastic approach to hotel revenue management considering multiple-day stays. *International Journal of Information Technology & Decision Making* 5: 545–556.

Pfeifer, P. and Bodily, S. (1990) A test of space – Time ARMA modelling and forecasting of hotel data. *Journal of Forecasting* 9: 255–272.

Queenan, C.C., Ferguson, M., Higbie, J. and Kapoor, R. (2009) Comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management* 16: 729–746.

Rajopadhye, M., Ben Ghalia, M., Wang, P.P., Baker, T. and Eister, C.V. (2001) Forecasting uncertain hotel room demand. *Information Sciences* 132: 1–11.

Sa, J. (1987) Reservations forecasting in airline yield management. PhD thesis, MIT, Flight Transformation Lab, Cambridge, MA.

Schwartz, Z. and Hiemstra, S. (1997) Improving the accuracy of hotel reservations forecasting: Curves similarity approach. *Journal of Travel Research* 36: 3–14.

Shehata, H.S. (2005) Measuring the concepts and practices of revenue management system in Egyptian hotels. Master's thesis, Faculty of Tourism and Hotels, Alexandria University, Egypt.

Sivillo, J., Ahlquist, J. and Toth, Z. (1997) An ensemble forecasting primer. *Weather Forecasting* 12: 809–817.

Skwarek, D.K. (1996) Competitive Impacts of Yield Management System Components: Forecasting and Sell-up Models. MIT Flight Transportation Lab Report R96-6.

Talluri, K.T. and Van Ryzin, G.J. (2005) *The Theory and Practice of Revenue Management*. New York, NY: Springer Science and Buisness Media.

Vinod, B. (2004) Unlocking the value of revenue management in the hotel industry. *Journal of Revenue and Pricing Management* 3(4): 178–190.

Weatherford, L.R. (1997) A review of optimization modeling assumptions and their impact on revenue. In Presentation at Spring INFORMS Conference, San Diego, CA.

Weatherford, L.R. and Kimes, S.E. (January 2003) A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting* 99(19): 401–415.

Wickham, R.R. (1995) Evaluation of forecasting techniques for short-term demand of air transportation. PhD thesis, MIT, Flight Transformation Lab, Cambridge, MA.

Yuksel, S. (2007) An integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling* 46: 1063–1070.

Zakhary, A., El Gayar, N. and Atiya, A.F. (2008) A comparative study of the pickup method and its variations using a simulated hotel reservation data. *ICGST International Journal on Artificial Intelligence and Machine Learning* 8: 15–21.

Zeni, R.H. (2001) Improved forecast accuracy in airline revenue management by unconstraining demand estimates from censored data. PhD dissertation, Graduate School-Newark Rutgers, The State University of New Jersey.

APPENDIX

Table A1 shows the different seasonal periods for Plaza Hotel, as determined by the managers. Shown is the very low season period and the high season periods. Any other period is considered low season. We made use of these periods to determine the seasonal average.

Table A1: The dates of the different seasonal regimes for Plaza Hotel (the rest of the days are low season days)

Start date	*End date*	*Occasion*	*Season type*
05 January 2006	14 January 2006	Eid Adha	High Season
28 March 2006	03 April 2006	Petroleum Convention	High Season
20 April 2006	26 April 2006	Easter	High Season
15 June 2006	15 September 2006	Summer	High Season
19 September 2006	18 October 2006	Ramadan	Very Low Season
19 October 2006	28 October 2006	Eid Fitr	High Season
28 December 2006	07 January 2007	Eid Adha	High Season
05 April 2007	11 April 2007	Easter	High Season
15 June 2007	15 September 2007	Summer	High Season
11 September 2007	10 October 2007	Ramadan	Very Low Season
11 October 2007	20 October 2007	Eid Fitr	High Season
17 December 2007	25 December 2007	Eid Adha	High Season
24 April 2008	03 May 2008	Easter	High Season
18 May 2008	23 May 2008	Petroleum Convention	High Season
01 June 2008	31 August 2008	Summer	High Season
25 August 2008	24 September 2008	Ramadan	Very Low Season
25 September 2008	06 October 2008	Eid Fitr	High Season
04 December 2008	13 December 2008	Eid Adha	High Season