ABSTRACT

Building an effective Hotel Revenue Management (HRM) system involves categorizing revenues from rooms into a few revenue buckets (Room Product). Once these are known an HRM analyst can forecast product demands for all buckets and control the room inventories to maximize overall revenue. The literature on HRM is sparse on details of forecasts, controls, and subsequent revenue opportunity and applications to real hotel data. This paper attempts to bridge that gap. The research uses archived hotel data from a major hotel chain to report initial results from the first two steps—room product categorizations and forecasting.

KEYWORDS: Hotel Revenue Management, Hotels, Forecasting, Forecast Monitoring, Revenue Bucket
INTRODUCTION

Revenue Management (RM) for service industries is the process of selling the right type of capacity to the right customer, at the right price, at the right time (Smith, Leimkuhler, & Darrow, 1992). The process involves rationally pricing and controlling reservations of perishable assets across market segments to maximize revenue (Cross, 1997; Baker & Murthy, 2005). A number of service firms reported significant increases in profitability due to RM. American Airlines reported approximately a 4-5% increase (1.4 billion USD over three years) in revenue (Smith et al., 1992; Cook, 1998). The ability to effectively implement RM strategies in different industries is subject to the various combinations of duration control and variable pricing that exist within each industry (Kimes, 2010).

The hotel industry was one of the early adopters of RM practices, based on initiatives in the airline industry (McGill and van Ryzin, 1999). Goldman et al. (2002) published a few stochastic models and techniques for hotel revenue management for accepting reservations for stays in a hotel. Vinod (2004) describes various RM techniques for controlling hotel rooms. Aslani et al. (2013) address the multiple night stay in hotel revenue management using a decomposition model. Ferguson and Smith (2014) more recently chronicled the changes in practices and responsibilities of the hotel revenue managers.

With all these efforts on hotel HRM a comprehensive approach is still missing in published literature. A comprehensive approach consists of the following component models: revenue opportunity model, forecasting model, and optimization model (Mukhopadhyay et al. 2014). All component models need be tested on real data before publication. Published studies on hotel RM examine only individual components. Further, most proposed HRM models, while providing analytical insight, do not report results using real data. Creating a comprehensive approach and using real data to examine revenue opportunities will bridge this gap in the literature.

This study can benefit both academic researchers and practitioners. Academic researchers can use the results as benchmark to compare results from their models, as Bodea et al (2009) suggest. For practicing analysts, the key role includes determining the different price/product groups that need to be created, and using appropriate forecasting and optimization models. Specifically, they must dynamically adjust the forecasting process to maximize the revenue. In this paper, we illustrate one such adjustment - that of using a risk factor involved in holding out rooms close to the consumption date. Illustrating this complexity can help practitioners make more effective decisions in this regard.

Our study proposes a two-step approach. First we develop and build all the component models for an effective HRM and apply them to real data. Next, we build a revenue opportunity model (ROM) to assess the overall revenue impact. The proposed study will fill the existing gap in the current literature. We will use archived data (Bodea et al, 2009) from a hotel chain for five hotels. This paper addresses the first two component models – room product categorization and forecast models. Subsequent future research will address the room inventory optimization modules and ROM.
LITERATURE REVIEW

HRM Framework
This section builds on the proposed HRM framework of Mukhopadhyay et al. (2014). Their theoretical framework extended traditional airline RM concepts to HRM. An airlines RM comprises two main modules – forecasting of passenger demand and optimized allocations of seats in each flight. We propose to use the framework along with real data to define product buckets and create forecasts. We give below conceptual overview of three HRM modules – room product descriptions, room product forecasting, and room allocation models.

Room Product Descriptions
The hotel industry functions in many ways like an airline industry. An airlines passenger demand forecasting system forecasts the number of passengers expected to fly on each combination of itinerary and fare category (Colville, 1996). A major airline can have thousands of unique fare categories or products. On a very fundamental level, values to variables (such as city-pairs, number of stops, advance purchase requirements, Saturday restrictions, time of day [red-eye flights, etc.], sale fares, competitor pricing, and other market conditions) give rise to thousands of unique fare products. At the most detailed level, one distinct fare product is known in the airline industry as fare basis code (FBC). Airlines RM ensures optimal allocation of different products (seat-mix problem) for all flights to maximize as much as possible the overall revenue. However, to avoid computational complexity and time consuming reservation decision making, airlines do not deal with the thousands of unique FBCs. Instead, airlines cluster these FBCs into a more manageable number of classes, called fare buckets, in each cabin based on the fare product’s economic value. From the airline’s point of view, due to differences in attributes of itinerary such as origin and destination, time of the day, and so on, passengers in the same fare class on the same flight are not equally valuable.

We use very similar concepts in hotel revenue management. HRM generates forecasts and room availability numbers at bucket levels. Hotels will have similar products like special rooms, general rooms, suites, etc.

Model Development
In determining the room fare product we employ k means type algorithm (k means algorithm) summarily explained below. Given a set of room prices (Xp1, Xp2, ..., Xpn), where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k (≤ n) sets S = {S1, S2, ..., Sk} that minimizes the within-cluster sum of squares. Lloyd (1982) proposed an algorithm for application in pulse-code modulation. Given an initial set of k means m(1)i, m(1)j, the algorithm proceeds by iterating between two steps:

Step 1: Assign each observation to the cluster whose mean yields the least within-cluster sum of squares.

\[ S_i^{(t)} = \left\{ x_p : \| x_p - m_i^{(t)} \|^2 \leq \| x_p - m_j^{(t)} \|^2 \forall j, 1 \leq j \leq k \right\}, \] (1)

where each \( x_p \) is assigned to exactly one \( S_i^{(t)} \).
**Step 2**: Calculate the new means for centroid of the observations in the new clusters.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

The formula above minimizes the within-cluster sum of squares. The means become the average prices for room fare buckets. We explain the concept of room fare bucket next in the forecasting section.

**Demand Forecasting and Forecast Accuracy**

Usually, forecasts for a room reservation can be made on the basis of *unconstrained demand* (an estimate of true demand) during a set of time interval periods for a group of similar room products. If a room product is open for sale, the observed booking for the product during the period is equal to unconstrained booking.

We assume that a typical room product stays for about 56 days prior to the check-in-date \(d\) in the reservation system. The 56-day span comprises a set of control time intervals. Usually, at the end of each interval, HRM analysts updates reservation and other controls for all \(d\) in future. The total time can consist of 56 days of equal seven-day intervals. However, intervals do not have to be of equal length. Historical room bookings used in forecasting demand for a room check-in-date \(d\) come from both past and future check-in-dates. Days left (DL) to \(d\) is then \((d-fd)\) where \(fd\) is the date when the forecast is made. The final forecast for a specific product for a specific \(d\) for a hotel will be sum of the bookings on hand up to \(fd\) and forecast for remaining bookings to come from \(fd\) to \(d\). The forecast for the remaining bookings to come is the sum of the incremental booking forecasts for all time intervals from \(fd\) to \(d\). We assume day-of-week seasonality since Monday bookings look like Monday bookings for all \(ds\) falling on Mondays and so on.

Let us say that each set of time intervals is indexed by a set \(i \in I\). For example, the set \(\{0, 7, \text{and} 14\}\) would index two periods that correspond to intervals DL = 7 to DL = 0 and DL = 14 to DL = 7. The final forecast for DL = 14 will be the sum of the bookings on hand up to DL = 14 from DL = 56 and the forecasts for remaining incremental bookings from DL = 14 to DL = 7 and DL = 7 to 0. HRM analysts determine the set of check-in-dates \(d_i\) to be used in the computation of the forecasts for the interval \(i\), such as:

\[
d_i \in D(I) \quad \text{iff} \quad (d_i - \text{Min DL in interval } i) < (fd - 1)
\]

and

\[
d_i \geq \text{maximum of } (d_i \in D(I)) - 7N
\]

The constraint (3) above ensures that all the bookings data for the interval \(i\) forecast have been collected. Constraint (4) ensures that historical incremental bookings are used from the most recent similar check-in-dates. The parameter \(N\) is multiplied by 7 in constraint (4) to accommodate day-of-the-week (DOW) seasonality. A value of \(N = 8\) spans over approximately the most recent two months’ similar check-in-dates bookings.

The unadjusted initial room demand forecast (UIF) for interval \(i\), DOW \(w\), booking product \(b\), check-in-date \(d\), and forecasting date \(fd\) is:

\[
UIF_{i,w,b,d,fd} = \frac{\sum_{n=\text{d} \in D(I)} M_n * IB_{i,w,b,d,fd,n}}{\sum_{n=\text{d} \in D(I)} M_n}
\]
Where $M_n = \text{smoothing weights} = F(a, n)$, where 
\[ a = \text{hotel revenue management analyst (HRMA)} \text{ chosen smoothing parameter}, \]
\[ n = 1, \ldots, N, \] and $IB$ is incremental bookings for interval $i$ of similar flight departures.

The Unadjusted final forecast (UFF) for $w, b, d, \text{ and } fd$ is:
\[
U_{w,b,d,fd} = BH_{w,b,d,fd} + \sum_{i} U_{w,b,d,fd}^i
\]

Where $BH_{w,b,d,fd}$ is current bookings on hand up to $fd$ in booking product $b$ for date $d$, day-of-week $w$.

**Accuracy Measurement Levels**

Let us say, for example, the forecast date $fd$ is May 4, 2015, Monday, for a check-in-date in 14 days on May 18, 2015. The RM model provides forecasts for two incremental intervals: days 14-7 and days 7-0. If RMAs choose the value of $N = 8$, historical bookings from eight most recent check-in-dates (May 11, May 4, Apr 27, Apr 20, Apr 13, Apr 06, Mar 30, and Mar 23, all in 2015) will be used for 14-7 incremental forecast. The incremental forecast for the interval 7-0 is generated from historical bookings from eight most recent check-in-dates (May 4, Apr 27, Apr 20, Apr 13, Apr 06, Mar 30, and Mar 23, Mar 16), all in 2006. The system computes final forecasts by adding the incremental forecasts of intervals 14-7 and 7-0 with bookings on hand up to May 4, 2015 all the way from $DL = 56$.

This research adopts the concept of **critical booking threshold** (CBT) from Mukhopadhyay et al. (2007) for measuring forecast accuracy. Monitoring forecast accuracies based on products could sometimes involve low demand numbers, because some of the high revenue products may not observe any bookings at all for many nights. Percentage forecast error for these products will be high even if there is only one room difference between the actual value and the forecasted value. Therefore, it is better to evaluate forecast accuracy in terms of the sum of the bookings in a product and all higher-valued products above. The calculation for the thresholds starts with calculating the cumulative percent of bookings for each $b$. To demonstrate the idea of buckets, we reproduce the example from Mukhopadhyay et al (2007) here. In their example, they created 16 buckets or price products, $B_1$ through $B_{16}$, with $B_1$ being the highest-revenue-yielding product and $B_{16}$ the lowest. This is shown in Table 1. The total or cumulative bookings for $a$ for a service provider, e.g. an airline or a hotel, at a specific $d$ are accumulated at the product level from $B_1$ to $B_{16}$. Using these bookings, the cumulative percent of bookings in each $b$ is calculated.

Next, the products with the top percentiles in terms of revenue to be measured are shown. The top percentile thresholds are defined as 10%, 25%, 50%, 75%, 90%, and 100% of hotel bookings. The percentile levels may vary from hotel to hotel depending on RMA experience. The booking type selected is the first product that is greater than or equal to the percentile to be measured (fifth column in Table 1). This also means that when a product spans two thresholds, it will be selected for both.
Using a similar structure and using real hotel data (Bodea et al, 2009), we created four (arbitrarily chosen target number) buckets for analysis, described in the data and results section. We propose to measure the Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE) for all forecasts at the CBT levels.

**Room Inventory Model**

The room booking of hotel management problem is probabilistic because there is uncertainty about the ultimate number of booking requests for rooms on a future date \( d \). There will be stochastic variation in demand around the expected values among similar nights sampled consistently over a homogeneous period of time similar to airline seat bookings assumed in a past research (Belobaba, 1987). The goal of the room inventory management is to protect a few rooms for late bookings with high revenue, even though very similar rooms might have gone to customers who booked at a lower rate earlier.

Past analyses have generally assumed a Gaussian (normal) distribution of product demand in the transportation industry. Similarly one may assume a Gaussian distribution for demand density in the hospitality industry. Means and variances depend on the hotel being studied and on the nature of its bookings. Demand densities for different products are not correlated significantly and the numbers of requests for various products during different periods before the night of stay are not correlated. The goal of this HRM module is to determine how many rooms to protect for the current product from the next lower revenue generating room products. We call this protection level (PL) of the current room product. PL for the current product is:

\[
PL = \text{Demand Forecast for the current product} + \text{Risk Factor due to Forecast Accuracy}
\]

**Risk Factor due to Forecast Accuracy**

Depending on the forecast accuracy HRM analysts will add or subtract a few rooms from the forecasted number to compute the PL. We assume that PL is the ideal expected demand level of the current product at which the expected revenue by protecting a room (marginal room number) for the current product from the next lower level product is equal to the marginal revenue (revenue of the current product minus the revenue of the next lower valued product). The optimal solution for the allocation problem is given by:

### Table 1: Example of CBT

<table>
<thead>
<tr>
<th>Products</th>
<th>Bookings</th>
<th>Cumulative Bookings</th>
<th>Cumulative Percentage</th>
<th>Thresholds</th>
<th>Sum CBT Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>10%</td>
<td>7</td>
</tr>
<tr>
<td>B3</td>
<td>2</td>
<td>9</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B4</td>
<td>2</td>
<td>11</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B5</td>
<td>0</td>
<td>11</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6</td>
<td>5</td>
<td>16</td>
<td>27</td>
<td>25%</td>
<td>16</td>
</tr>
<tr>
<td>B7</td>
<td>5</td>
<td>21</td>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B8</td>
<td>5</td>
<td>26</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B9</td>
<td>5</td>
<td>31</td>
<td>52</td>
<td>50%</td>
<td>31</td>
</tr>
<tr>
<td>B10</td>
<td>5</td>
<td>36</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B11</td>
<td>5</td>
<td>41</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B12</td>
<td>5</td>
<td>46</td>
<td>77</td>
<td>75%</td>
<td>46</td>
</tr>
<tr>
<td>B13</td>
<td>2</td>
<td>48</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B14</td>
<td>2</td>
<td>50</td>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B15</td>
<td>5</td>
<td>55</td>
<td>92</td>
<td>90%</td>
<td>55</td>
</tr>
<tr>
<td>B16</td>
<td>5</td>
<td>60</td>
<td>100</td>
<td>100%</td>
<td>60</td>
</tr>
</tbody>
</table>
\[ \text{PL}_{b,(d-fd)} = \text{DF}_{b,(d-fd)} + \sigma_b \times \text{Norm Inverse of } (1 - R2/R1) \]  
\[(7)\]

Where \( R1 \) = average revenue from the current higher valued room bucket product, 
\( R2 \) = revenue from the next lower valued room bucket product, 
DF is the demand forecast for the current product and 
\( \sigma_b \) = standard deviation of bucket demand \( b \).

The second part of the right hand side of equation 7 is denoted as the "risk factor." The risk factor associates with forecast accuracy. The risk factor goes up with decreased differential revenue between the higher valued current product and the next lower-valued product. In that case, protecting too many rooms for the higher valued product will cost the hotel with unoccupied room inventories. On the contrary, the risk in protecting fewer rooms for the current product will cost the hotel in achieving suboptimal total revenue.

**DATA AND INITIAL RESULTS**

We used the hotel bookings data published by Bodea et al (2009). They describe data collected from five U.S. properties of a major hotel chain that can be used for revenue management (RM) research. The data described in their paper is publicly available at the journal's website at http://msom.pubs.informs.org.

We carefully recreated the reading days and room bucket differentiation. We used PROC FASTCLUS procedure of the Statistical Analytical Systems [(SAS) software version 9.2] to generate the average fare bucket values. In our initial runs we decided to use four buckets. The average fares for buckets B1 – B4 are 510, 428, 380, and, 299. We have determined 6 reading days in the initial runs. However, a more in-depth analysis is required to optimally determine the right number of products and the reading days. Figure 1 shows a typical booking pattern of fare bucket over 60 days prior to check-in-date up to the check-in-date. The graph shows the time series in reverse, with the check-in date as 0 on the X axis. As expected, because of cancellations there is a dip near the check-in-date.

**Figure 1: A sample booking profile of room product from 60 days prior to check-in-date**

The booking profiles across reading days for all four buckets are shown in Figure 2.
As expected the shape of the booking profiles are similar whereas the magnitudes are decreasing from higher bucket number (cheaper price) to lower bucket numbers (more expensive).

Table 2 shows a sample of cumulative bookings and forecasts over reading day where bkg1 means cumulating bookings up to the reading days in the bucket b for a specific check date. Fcst2 is the forecast made at the beginning of reading day 2 for a specific bucket and checkdate combination. Bkg 1 is the total bookings at the checkdate. We compute forecast errors based on the checkdate bookings and each reading day forecast. Table 3 shows the errors in forecasting for all reading days across all buckets.

Table 3: Forecast accuracy across all reading days and buckets in the sample

<table>
<thead>
<tr>
<th>MAE</th>
<th>38.16</th>
<th>50.20</th>
<th>55.15</th>
<th>57.85</th>
<th>58.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>54.42</td>
<td>78.63</td>
<td>82.95</td>
<td>86.22</td>
<td>86.65</td>
</tr>
</tbody>
</table>

As expected, the forecasts are more accurate for the reading days closer to the check-in-date (Figure 3).
CONCLUSION AND FUTURE RESEARCH

RM models generate a lot of interest to service industry practitioners as well as to the academic community. In the hotel industry, RM involves holding out a certain number of rooms in the hope of selling them at a higher price as the consumption date nears. Determining the number of rooms to hold out depends on demand sensitivity to its price. This problem is dynamic in nature, and the forecasts must be continually adjusted based on prior demand and risk factor. As described in the paper, a revenue management system clusters room types into different price/product groups. The actual RM then involves closing or releasing rooms from each bucket depending on the demand at hand and the assessed risk factor in order to maximize revenue.

The revenue management analyst adjusts the forecasting as well as the risk factor based on their experience and the latest information on booking. RM systems for Airlines traditionally have three components – creating/defining revenue products, forecasting, and optimization. Although this knowledge exists in the literature, an extension of the airline model to the hotel industry is lacking. The main contribution of this paper is the application and extension of traditional RM methods to the hotel industries. Specifically, we show the logic behind and an illustration of bucket creation as a pricing strategy, followed by forecasting multiple times until the consumption date. Future research will report results from the remaining modules – optimization and revenue opportunity. Since we use real data that is publicly available, the results in this paper can serve as benchmark for future researchers, in line with the expectation of Bodea et al (2009).

Effective HRM in the future will involve moving it to a separate department in the organization and making sure that the revenue management analysts have the necessary analytical and communications skills to apply models effectively (Kimes, 2010). In recent years, several firms have created a separate analytics function (including revenue management) within their organization as the complexity of data and models continues to grow. Practitioners in the hotel industry can apply the methods shown in this paper to create their own customized solutions for optimal revenue management.
REFERENCES


