Forecasting Demand for Lodging Properties at a Resort: A Comparison of Methods

by

Dylan Roth

B.Comm., Queen's University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN BUSINESS ADMINISTRATION

IN THE FACULTY OF GRADUATE STUDIES

FACULTY OF COMMERCE AND BUSINESS ADMINISTRATION

DEPARTMENT OF MARKETING

We accept this thesis as conforming to the required standard

(Dr. Daniel Putler)

(Dr. Charles Weinberg)

(Dr. John Claxton)

(David Jenkins)

THE UNIVERSITY OF BRITISH COLUMBIA

March 2003

Department of Commerce

The University of British Columbia
Vancouver, Canada

Date March 21/2003

Abstract

Demand forecasts are the most important piece of information used to make revenue management decisions for lodging properties. High demand forecasts may lead to increases in room rates and stay restrictions while low demand forecasts may result in price decreases and easing of stay restrictions. A number of demand forecasting methods, both long-term (more than 90 days prior to a target date) and short-term (within 90 days of a target date) were modelled and compared for the lodging properties at a major North American ski resort. Long-term forecasting methods included random walk, multiplicative Holt-Winters, ARIMA (autoregressive integrated moving average), linear regression, and nonlinear regression. Short-term models included the five long-term forecasting methods as well as additive pickup and a regression-based booking curve model. In terms of long-term forecasts, the nonlinear regression method was found to be superior while capacity was trending upward but after a capacity shock (unexpected loss in capacity) the random walk method proved optimal. In terms of short-term forecasts, the regression-based booking curve model was optimal in-sample and data was not tested out of sample. Further, the long-term nonlinear regression model and short-term regression-based booking curve model explicitly defined seasonal periods. These statistically defined seasonal periods should help management set seasonal rate targets as well as better understand typical booking patterns among periods.

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

Forecasting demand is a critical component of revenue management for lodging operators (hotel and rental properties). Lodging units are perishable inventory since revenue from a lodging unit on a certain date is lost forever if the unit is not filled. Given that the majority of lodging costs are fixed, lodging operators must take appropriate actions to maximize lodging revenue. The demand forecast is the piece of information upon which revenue management decisions are made. A demand forecast that is higher than expectation may lead to increases in price, stay controls (i.e. minimum two night stay) and other restrictions. On the other hand, a demand forecast lower than expectation may trigger promotions, price discounts, and a lifting of restrictions. In this paper, demand is used synonymously with bookings and is defined as occupied room nights. For example, a reservation for 2 units at 7 nights is equivalent to 2 room nights per day for a total of 14 room nights.

Lodging demand estimates were calculated for a major North American ski resort. Long-term estimates were calculated more than 90 days prior to a target date, while short-term estimates were calculated within 90 days of a target date. Standard forecasting models including linear regression, multiplicative Holt-Winters, ARIMA (autoregressive integrated moving average), and random-walk were used to create long-term forecasts. These were compared to a nonlinear regression model built specifically to capture the resort's yearly demand trend and seasonality (e.g. regional school holidays). Within the four year sample period, the nonlinear regression provided superior estimates to the other models. However, out of sample, the nonlinear regression model only provided a marginal improvement over other models during the first four months of the out of sample period. At this 4-month point an underlying assumption of the nonlinear model (linearly increasing yearly demand) was violated as the resort experienced a large decrease in capacity when a major lodging property switched reservation management providers. After this capacity shock, random walk provided the best estimates for the remainder of the out of sample period (and had provided the second best estimates among long-term forecasting methods prior to the assumption violation). As a result, a 'customized' nonlinear regression model is a recommended long-term forecasting method for resorts which are experiencing predictable yearly shifts in demand (increasing or decreasing) which can be approximated by a functional form such as a linear trend or modified exponential curve. If changes in yearly demand are sporadic or small the random walk method is recommended for long-term estimates since it is simple and robust.

Short-term estimates (estimates within 90 days of a target date) typically come in two varieties. The first variety includes the same models used for long-term forecasts but with a shorter forecasting horizon. In other words, these models use past complete stay information to forecast future complete stays. The second variety of short-term models includes models that incorporate current bookings for future dates (bookings to date). Additive pickup (AP) models are similar to random walk methods; they use current bookings and add last year's pickup (number of reservations made in the prior year from Y days out up until the target date) to come up with a short-term forecast. The other tested model incorporating bookings to date is a 'customized' booking curve model. This model creates a baseline booking curve (pattern of bookings over time for a particular target date) and then compares actual bookings to the baseline bookings in order to project demand for the target date. This projected demand is then combined with a long-term non-linear estimate; the weight between estimates determined by the number of days (lead time) from the target date.

Within the four year sample period, the 'custom' booking curve model provided superior estimates to all other models (AP and long-term models). Further, the booking curve model and long-term nonlinear model explicitly define seasonal periods based on demand. These statistically significant seasonal periods provide management with valuable information for setting room rate targets since room rates are set to correspond to distinct demand levels. As well, the booking curve model provides management with expected booking curves; the systematic build-up in bookings for a particular target date. These expected booking curves quantify the relationship between lead-time and demand for a certain period, helping management to identify the likely extent of last minute bookings versus reservations in advance. However, while the nonlinear model and booking curve model have many benefits and are likely to increase the accuracy of forecasts, the benefit of this additional accuracy is directly related to the amount of excess capacity. A large amount of excess capacity, as is the case in the resort studied, leads to a low cost of demand inaccuracy since all reservations can be accommodated regardless of final demand. Constrained capacity environments, on the other hand, have a large opportunity cost of demand forecast inaccuracy since high-value reservations (e.g. reservations with high daily room rates and long length of stay) should be prioritized above low-value reservations. If the forecasts for high-value reservations and low-value reservations are inaccurate in a situation of constrained capacity, then reservation management will make sub-optimal decisions with respect to pricing, stay controls, and appropriate mix of market segments.

2 FORECASTING APPROACH

Weatherford, Kimes, & Scott (2001) provide a useful framework for forecasting demand for hotel properties. They contend there are seven decision factors that must be determined prior to a lodging forecast and these are outlined in Table 1, as well as the approach taken in this paper.

Table 1: Forecasting choices made in resort lodging models

Weatherford et al. forecast choices	Forecast choice for resort estimates
1) What to forecast a) Arrivals b) Room nights	**1. b)** Room nights
2) Level of aggregation a) Fully aggregated b) Aggregated by rate category with length-of-stay probability distributions c) Aggregated by length of stay with rate-category probability distributions d) Fully disaggregated (by rate category with length of stay)	**2.** The approach taken deviates slightly from the choices stated by Weatherford et. al. (2001). Booking data was aggregated by market segment (independent traveler, group, and owner) as well as by bedroom (one bedroom (including suites), two bedroom, and three plus bedrooms). Forecasts were provided for the independent traveller segment by bedroom type.
3) Unconstraining method a) None b) Denials data c) Mathematical models i) Pickup ii) Booking curve iii) Projection	**3. c)** There is no unconstraining method for long-term models. For short-term models, both pickup and booking curve methods are used.
4) Number of periods to include in forecast a) All b) Selected number	**4. a)** All
5) Which data to use a) Only complete stay-nights b) All data (complete and incomplete stay-nights)	**5. b)** All data; only complete stay-nights are used for long-term forecasts while short-term forecasts used all data.
6) Outliers a) Included b) Not included	**6. a)** Outliers included
7) Level of forecast accuracy a) Aggregated forecasts, errors reported as average across all reading days b) Aggregated forecasts, errors reported for each individual reading day c) Disaggregated forecasts, errors reported as average across all reading days d) Disaggregated forecasts, errors reported for each individual reading day	**7. a)** Aggregated forecasts…while models are calculated at a disaggregate level by reading day (as in d), decisions about the model's efficacy are reported at an aggregate level.

An effective revenue management system uses estimates for both guest arrivals and room nights in order to maximize revenue. Predicted arrival distributions are important so that the resort can implement effective strategies for specific arrival days (i.e. price changes and stay controls).

However, if capacity is not expected to be surpassed then stay controls are never used. In situations of capacity slack, predicted room nights alone, rather than predicted room nights by arrival segment, are generally adequate for revenue management. Room nights for the independent traveller segment were determined to be the most important estimates for the studied resort since independent travellers pay higher room rates than group reservations, and their bookings are made closer to the target date than owner or group reservations. Since the resort rarely sold out (4 days in the most recent year), and therefore estimating the number of rooms likely to be occupied was revenue management's primary concern.

Weatherford et al. (2001) found that summing disaggregated hotel demand forecasts produced a more accurate forecast than a single aggregate demand forecast. As a result, the resorts' booking data was disaggregated as much as possible. Room night forecasts were only created for the independent traveller segment as these were the most predictable bookings and did not suffer from data inconsistency problems at the resort level. Group bookings were often excluded from the reservation management system until shortly prior to a target date making it problematic to determine when reservations/cancellations were actually made. The owner bookings were generally flat (did not change much from 90 days out up until the target date) due to incentives for owners to claim vacation dates far in advance. As a result, demand forecasts were not created for group and owner segments. Furthermore, denials (requests for unavailable lodging units) and turndowns (customers refusing a room type at a certain price or stay control) were not tracked by reservation agents for historical data. As a result, it was deemed problematic to disaggregate by rate class in forecasts since estimating appropriate probability distributions for different rate classes would be contrived. Furthermore, due to the tremendous seasonality of the resort (nearly 100% occupancy during Christmas period and often less than 10% during shoulder periods – e.g. early November and early May) it was hypothesized that seasonality alone would explain most of the demand variation.

The third decision factor cited by Weatherford et al. is unconstraining method. In other words, what technique is used to separate demand from capacity? Quite simply, it is impossible to occupy more than 100% of the resort's lodging units, yet this does not limit demand to 100% of capacity. For all the long-term forecasting models there is no unconstraining method. Complete stay night information, by definition, is constrained by the resort's capacity so these models do not capture demand above capacity. However, as mentioned earlier, due to the infrequent nature of sellouts at the resort, this was not seen as a major problem. The short-term forecasting models do provide unconstrained forecasts. The additive pickup method may forecast demand above capacity if last year's pickup plus current bookings are above capacity. However, this method may underestimate total unconstrained demand if either current bookings have been limited by capacity or if last year's pickup was limited by capacity. The short-term nonlinear regression model uses booking curves (pattern of bookings observed in similar days past) as a baseline to gauge future demand. However, similar to the additive pickup model, there is potential to underestimate unconstrained demand if either the booking curve developed from prior year data was constrained by capacity or if current bookings are constrained by capacity. Due to the infrequent nature of resort-wide sellouts, both the additive pickup model and short-term nonlinear regression model were expected to be very close approximations of unconstrained demand.

In terms of data used for forecasting, the entire four year sample was utilized in the model analysis and this includes bookings that never materialized due to cancellations and no-shows. By including cancellations and no-shows, the models are better able to project future demand given current bookings. As well, the data was not scrubbed to exclude outliers since large variation in demand (due to weather, promotions, randomness, etc.) is typical in lodging forecasting. Readers interested in a more detailed description of the data preparation and transformation process used in the models are referred to Appendix A.

A reading day is defined by Weatherford et al. (2001) as the day when the number of reservations on hand for a particular arrival day is updated. At the resort studied, reading days were generally updated on a weekly basis within 90 days of a target date and updated daily in the week prior to a target date. Since resort revenue management's approach to forecasting was random walk (last year's occupancy figure for long-term forecasts and additive pickup for short-term forecasts) it was straightforward to compare model forecasts to likely management forecasts for any given day of historical data. In order to provide maximum accuracy in forecast comparisons, each day in the 90 day window was treated as a reading day.

3 LONG-TERM MODELS

3.1 Resort Overview

Revenue management at the resort studied manages roughly 50% of the bed base on-mountain. The fraction of rooms under management has remained roughly constant in the past five years as the development of on-mountain properties by the resort's real estate division has largely matched development by external hotel chains. For resort managed properties, the revenue management division is responsible for managing reservations as well as setting prices and stay restrictions. The following long-term and short-term models are demand forecasting methods for FIT (free independent traveller) segments only.

The long-term models consist entirely of past demand information. The reader is referred to Figure 1A and Figure 1B for a sample of room nights in the 01/02 season. As can be seen from Figure 1A, there is tremendous seasonality throughout the year. The first shoulder period from late April until mid June is very slow, as the ski hill is closed for downhill skiing and school is not yet out for the summer. The summer period has very high occupancy, as the resort has many summer travelers and on-hill activities, with demand peaking in the first weekend of August. In the second shoulder season, demand gradually declines from early September until mid November, until the hill opens for downhill skiing in late November. The bookings then ramp up until the Christmas period peaking at New Year's. From January until late March the hill is again very high occupancy with peaks for regional holidays such as school breaks, as well as the weekends surrounding Martin Luther King Day and President's Day. Figure 1B shows the daily variation in demand, with Friday and Saturday nights commanding greater demand than weekdays. However, the daily variation changes dramatically by seasonal period, with weekend nights making up a greater proportion of room nights in shoulder seasons while daily variation is more evenly distributed during high occupancy periods. Both weekly seasonality as well as daily variation in demand are much greater at resort hotels than at business-oriented hotels. As a result, the variation in pricing is also much more cyclical at a resort hotel than at a business-hotel which tends to have higher average occupancy levels.

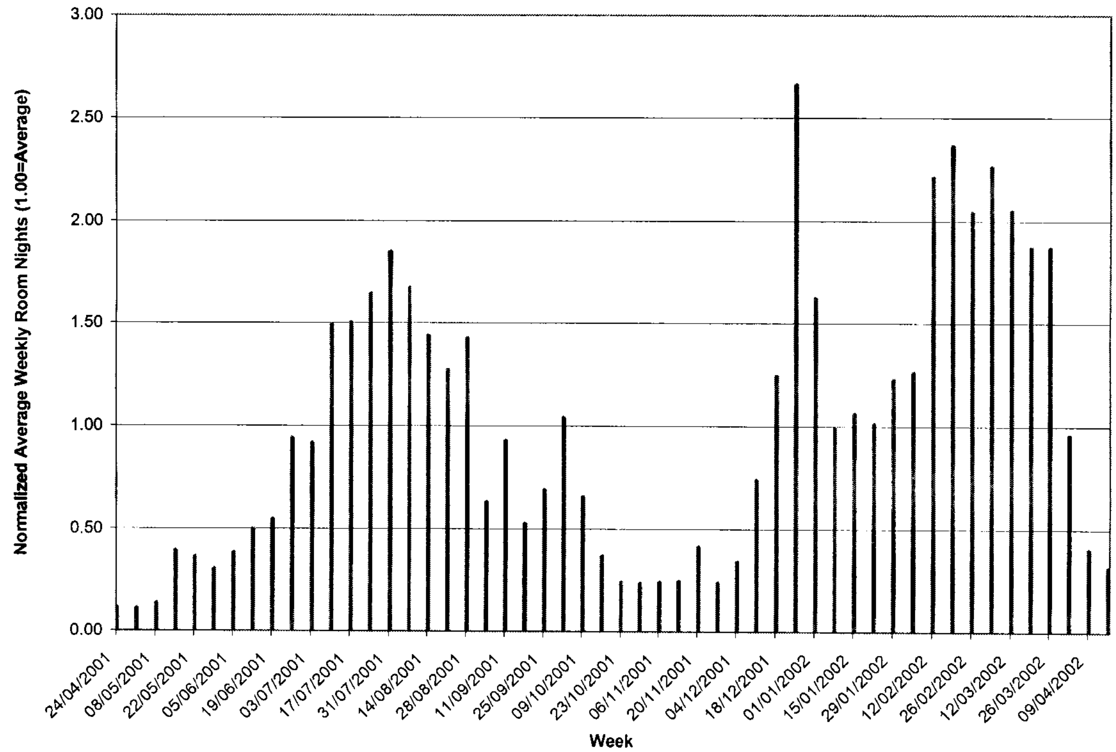

Figure 1A: Normalized average daily room nights for the 01/02 season by week

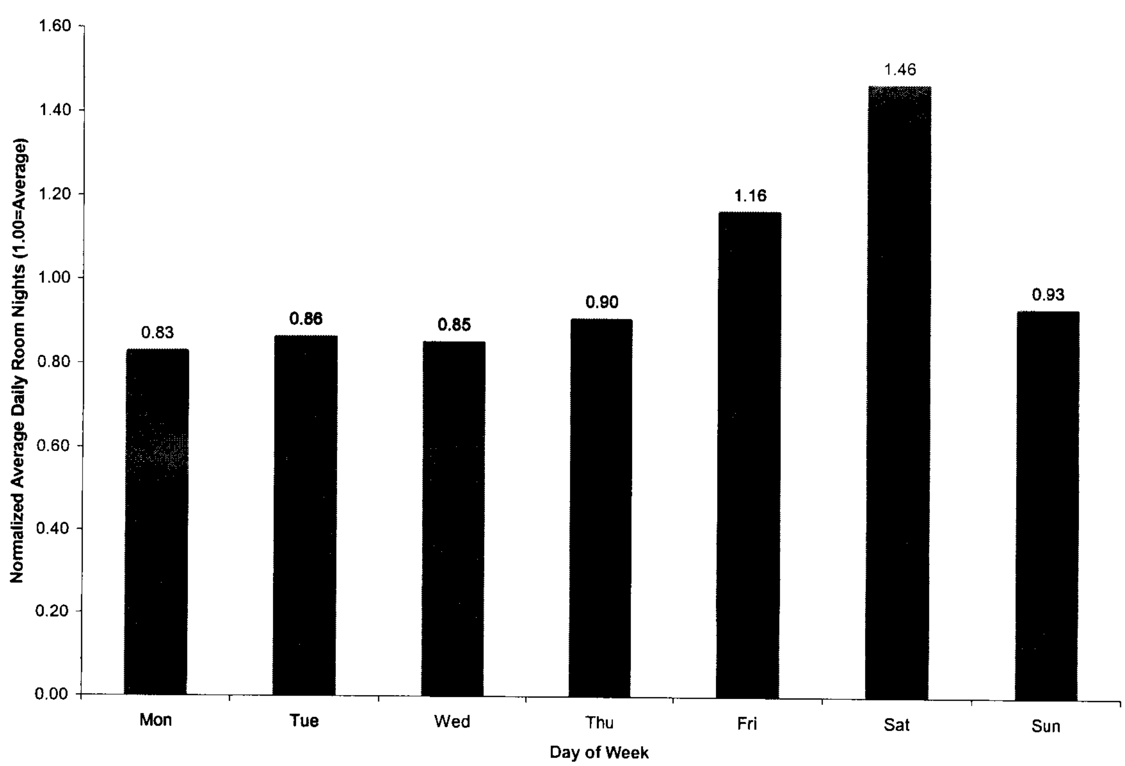

Figure 1B: Normalized average daily room nights for the 01/02 season by day of week

There are numerous approaches to long-term daily demand forecasting and final forecasts are often a combination of statistical estimates and managerial judgement. This paper focuses on a statistical approach to forecasting while readers interested in formal approaches to the integration of managerial judgement and statistical forecasts are referred to Ghalia & Wang (1999). Before the different long-term forecasting models are described and compared, however, appropriate criteria for model efficacy must be chosen.

3.2 Model Efficacy Criteria

Traditionally, mean square error (MSE) is a standard error measure for statistical models. Specifically, the objective of most parameter estimation algorithms is to minimize MSE (as is the case of all models to be tested in this paper except for ARIMA models). However, MSE is found by many researchers to be a poor measure of forecast validity. Armstrong and Collopy (1992) in their oft cited work "Error Measures For Generalizing About Forecasting Methods" tested error measures against a number of criteria including reliability, construct validity, sensitivity to small changes, protection against outliers, and relationship to decision-making. They recommend using deviants of two different error measures, the relative absolute error (RAE) and absolute percentage error (APE), in order to choose among forecasting methods. The RAE (Equation 1) for a single estimate is the ratio of the absolute error of a particular forecasting method (e.g. Holt-Winters' method) divided by the error of the random-walk method. The APE for a single estimate (Equation 2) measures the absolute error as a percentage of the actual observation. For a single horizon Armstrong & Collopy recommend using the median relative absolute error (MdRAE) when a small number of time series are available and the median absolute percentage error (MdAPE) when there are a large number of series (Equations 3-4). To compare series over a long horizon, they recommend the cumulative relative absolute error (CumRAE) for a single series and median cumulative relative absolute error (MdCumRAE) for multiple series (see Equations 5-6).

$$RAE_{m,h} = \left| \frac{F_{m,h} - A_h}{F_{rw,h} - A_h} \right| \tag{1}$$

$$APE_{m,h} = \left| \frac{F_{m,h} - A_h}{A_h} \right| \tag{2}$$

$$MdRAE_{m,h} = Median(RAE_{m,h,s}) \text{ for all series } s \tag{3}$$

$$MdAPE_{m,h} = Median(APE_{m,h,s}) \text{ for all series } s \tag{4}$$

$$CumRAE_m = \frac{\sum_{h=1}^{H} \left| F_{m,h} - A_h \right|}{\sum_{h=1}^{H} \left| F_{rw,h} - A_h \right|} \tag{5}$$

$$MdCumRAE_m = Median[CumRAE_m] \text{ for all series s} \tag{6}$$

where:

m	Forecasting method (e.g. Holt-Winters, ARIMA, etc.)
h	Horizon (lead time) being forecast (*h*>90 for long-term forecasts)
s	Forecast series
$F_{m,h}$	Method *m* forecast for horizon *h*
$F_{rw,h}$	Random walk forecast for horizon *h*
A_h	Actual observation at horizon *h*
$RAE_{m,h}$	Relative absolute error of method *m* at horizon *h*
$APE_{m,h}$	Absolute percentage error of method *m* at horizon (lead time) *h*
$MdRAE_{m,h}$	Median relative absolute error of method *m*, horizon *h* for all series *s*
$MdAPE_{m,h}$	Median absolute percentage error of method *m*, horizon *h* for all series *s*
$CumRAE_m$	Relative absolute error (RAE) of method *m* summarized across all *h* horizons of a particular series
$MdCumRAE_m$	Median CumRAE of method *m* for all series *s*

Now that appropriate error metrics have been chosen the long-term model estimates can be compared and evaluated. The five different long-term estimation methods include random walk (RW), linear regression (LR), multiplicative Holt-Winters (HW), autoregressive integrated moving average (ARIMA), and nonlinear regression (NL). The models were calibrated using the entire four year sample period from May 15, 1998 to April 29, 2002 and forecasts compared in year four

(April 28, 2001 to April 27, 2002). In other words, the entire four year sample was used to determine the functional form of each model (number and type of parameters), but depending on the model, the entire four years may not have been used to calculate the parameter estimates for the in-sample period. Specifically, the LR and NL models used the entire four year sample to calculate parameter estimates and the same parameter estimates were used in year four. In contrast, the HW and ARIMA models used only sample data prior to the in-sample forecast to calculate parameter estimates; so data from years one to three were used to calculate parameter estimates for year four forecasts.

Using the ARIMA model as an example, it was determined using the entire four year sample that an ARIMA$(2,0,2)(1,0,0)_7(1,1,0)_{364}$ functional form for one bedrooms best fit the entire sample dataset, yet the actual parameter values for the ARIMA$(2,0,2)(1,0,0)_7(1,1,0)_{364}$ were different for year four. Further, since a long-term forecast is defined in this paper as any forecast made more than 90 days prior to a target date, the HW and ARIMA models began to forecast from January 28, 2001 in order provide long-term estimates for the year four forecast period (April 28, 2001 to April 27, 2002). The out of sample period was from July 29, 2002 to January 22, 2003. The out of sample period did not begin until July 29, 2002 to allow for a 90 day period from the most recent in-sample date (April 29, 2002) used to parameterize the models.

3.3 Random Walk (RW)

Random walk simply means to make predictions of future demand using past demand directly (without any modelling process). In this paper, in order to obtain the same seasonal period and day of week, the final demand from 364 days prior (52 weeks) is used as an estimate for future demand. For example, the final long-term demand estimate for July 30, 2002 is taken from the final demand for July 31, 2001.

3.4 Multiplicative Holt-Winters (HW)

A standard statistical demand forecast is a simple exponential moving average model. While a simple exponential smoothing model is not an appropriate method for daily demand forecasting when seasonality is present, it is a good base upon which to understand more complex smoothing models such as HW and ARIMA. The basic form of an exponential smoothing model is shown in Equation 7 (as derived from smoothing model presentations in SAS ETS User's Guide, 1999 and Chatfield, 1989). As can be seen, the weights decrease in a constant proportion, thereby giving more weight to recent observations and less weight to past observations. Exponential smoothing is the process by which the weights are calculated recursively in order to minimize the squared error. The error term in exponential smoothing is shown in Equation 8, and so (7) can be restated in error-correction form as Equation 9 or alternatively as Equation 10. ARIMA models, to be explained in the proceeding section, are a large class of models expressed in error-correction form. The simple exponential moving average model is expressed as an ARIMA(0,1,1) in Equation 11.

$$\hat{Y}_t = \alpha Y_{t-1} + \alpha(1-\alpha)Y_{t-2} + \alpha(1-\alpha)^2 Y_{t-3} + \ldots \tag{7}$$

$$e_t = Y_t - \hat{Y}_t \tag{8}$$

$$\hat{Y}_t = \alpha e_{t-1} + \hat{Y}_{t-1} \tag{9}$$

$$\hat{Y}_t = \alpha e_{t-1} + \alpha e_{t-2} + \alpha e_{t-3} + \ldots = \alpha \sum_{j=1}^{T-1} e_{t-j} \tag{10}$$

$$(1-B)Y_t = e_t(1-\theta B) \tag{11}$$

where:

Y_t	Observation at time t
$\hat{Y}_t$	Estimated observation at time t
α	Smoothing parameter for time-varying mean term
e_t	Error (disturbance) term at time t
B	Backward shift operator (e.g. $(1-B)y_t = y_t - y_{t-1}$)
T	Total number of time periods for which observations exist

The multiplicative Holt-Winters model is based on an exponential smoothing model but includes parameters to account for trend and seasonality. The multiplicative version was used since the additive version can be expressed as an ARIMA(0,1,p+1)(0,1,0)$_p$ model, and a multiplicative model seemed more appropriate since variation in demand is likely to increase with an increase in yearly demand. In the hotel industry, the multiplicative Holt-Winters three parameter exponential smoothing method is an industry standard (Baker & Collier (1999)). The HW model actually has more than three parameters, but it is referred to as a three parameter model as it has three smoothing parameters; alpha smoothes the time varying mean-term, gamma smoothes the time-varying slope, and delta smoothes the time-varying seasonal contribution. The estimate of the HW model is shown in Equation 12, with the separate elements of (12) detailed in Equations 13-15. For comparison, the simple exponential smoothing model is described as an HW model in Equation 16. It should be noted that HW is multiplicative since the time-varying mean and slope terms are multiplied by the seasonal term. This results in seasonal variation increasing as the trend or slope terms increase, whereas the additive HW model maintains constant seasonal variation around the trend and slope terms.

$$\hat{Y}_t(h) = (L_t + hT_t)S_{t-p+h} \tag{12}$$

$$L_t = \alpha(Y_t / S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{13}$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \tag{14}$$

$$S_t = \delta(Y_t / L_t) + (1-\delta)S_{t-p} \tag{15}$$

$$\hat{Y}_t(h) = L_t \qquad \text{since } T_t = 0,\ S_{t-p+h} = 1 \tag{16}$$

$$\hat{Y}_t(h) = L_t + hT_t + S_{t-p+h} \tag{17}$$

where:

Y_t	Observation at time t
$\hat{Y}_t$	Estimated observation at time t
h	Forecast horizon
α	Smoothing parameter for time-varying mean term
γ	Smoothing parameter for time-varying slope
δ	Smoothing parameter for time-varying seasonal contribution
L_t	Smoothed level that estimates the time-varying mean term
T_t	Smoothed trend that estimates the time-varying slope
S_{t-j}	Smoothed trend that estimates the time-varying seasonal contribution for one of the *p* seasons (j=0,…,p-1)

The additive HW model (Equation 17) can be expressed as an ARIMA(0,1,p+1)$(0,1,0)_p$ whereas the multiplicative HW model cannot be expressed as an ARIMA model. Note that the components of the multiplicative HW model (Equations 13-15) are not the same as the components of the additive HW model. Further, in the additive HW model the sum of the seasonal terms is zero while in the multiplicative HW model the average of the seasonal terms is one. Models were created for one bedroom demand and two/three bedroom demand. Two and three bedrooms were combined into a single model since the HW algorithms require non-zero elements and there were many zero value days for the three bedroom time-series.

To determine starting values of the trend component, SAS software allows either a constant estimate, linear trend estimate, or quadratic trend estimate of the starting value. At the resort studied, the constant trend estimate provided the lowest MSE for one-step ahead forecasts.

Further, in the HW model estimated for the resort, the seasonal term is actually the product of two terms. The first seasonal term is a weekly term, thus there are 52 seasonal week parameters. The second seasonal term is a day of week term, thus there are seven day of week parameters. As a result, there are 364 (52X7) unique seasonal factors derived from 59 (52+7) seasonal parameters. The multiplicative HW parameter estimates for the entire four year sample (separate models for one bedroom as well as two/three bedrooms) are shown in Appendix B while the summary model results are shown below in Table 2. The HW parameter estimates can be compared directly to the normalized room night values of the past season (Figure 1A and Figure 1B). The HW weekly parameters can be seen to be quite similar to the normalized weekly values although not as extreme in high periods, while the HW daily parameters have more variation than normalized daily demand values.

Table 2: Long-term Holt-Winters multiplicative model results (May 15, 1998 to April 29, 2002)

Model	Type of model	# of parameters	Classes of parameters	R^2	# of observations
1 bedroom	Holt-Winters multiplicative	62	• Smoothing parameters (3) • Day of week parameters (7) • Weekly parameters (52)	.57*	1,446
2/3 bedroom	Holt-Winters multiplicative	62	• Smoothing parameters (3) • Day of week parameters (7) • Weekly parameters (52)	.57*	1,446

*The R^2 is calculated from 1 step-ahead forecasts for the entire sample period.

3.5 Autoregressive Integrated Moving Average (ARIMA)

A slightly more complex time-series approach than either RW or HW for modelling daily demand is an ARIMA (autoregressive integrated moving average) model. ARIMA models, as discussed previously, are models described in error-correction form. Generally, data is differenced (often by year or by some other seasonal period) to induce stationarity, and then the pattern of movement around the mean term is estimated. The pattern of movement about the mean is estimated using polynomial based models. Polynomial based models are effective since they allow a large amount of variation in the weighting of past observations by using a minimum number of parameters. For example, a small number of parameters in the numerator and denominator of the error structure (right hand side of an ARIMA specification) can interact to form a complex weighting pattern that can be applied to an infinite number of observations.

The process of estimating an ARIMA model can be described as analyzing the residuals (error terms) of a time-series process and adding appropriate parameters until there is no longer a systematic component in the residuals. Once the systematic component (or signal) has been sufficiently modelled, the new residuals are said to have been reduced to white noise. White noise means a stochastic process with mean zero. ARIMA models are not automatic (independent of modeller judgement and specification) as they require the modeller to analyze autocorrelation, inverse autocorrelation, and partial correlation plots of the error terms in order to determine appropriate ARIMA parameters. As mentioned, parameters are deemed appropriate if they are statistically significant (i.e. significant t-statistics at a 95% level of confidence) and generally added until the residuals are deemed to be white noise (as tested by chi-square statistics at a 95% level of confidence). Since the ARIMA process is so flexible, the same weighting functions can be achieved by a variety of ARIMA specifications. Therefore, parsimony is extremely valuable in ARIMA models, and the Akaike Information Criterion (AIC) is often used to judge the appropriateness of different ARIMA specifications (and was the objective used in modelling ARIMA models of daily demand for the resort studied).

The strength and weakness of ARIMA models is that they are often able to capture patterns not immediately apparent to the researcher. In the best instances, they allow discovery of new data patterns and hence provide better forecasts of future observations. In the worst instances, they result in a model that cannot be interpreted or a model that has simply overfit the sample data. Overfit models are overly complex and do not provide better forecasts than simpler more interpretable models. In spite of ARIMA model reservations, these models have been used extensively in financial analysis (e.g. prediction of stock market data) and are often combined with econometric models to further specify the error terms generated by a regression-based analysis.

Two different ARIMA models (one, two/three bedrooms) were specified for the resort. Three bedroom data was combined with two bedroom data as an ARIMA model built on three bedroom data model alone did not provide good estimates due to many zero values. In fact, a two/three bedroom model provided better estimates than the sum of an independent two bedroom model and a three bedroom model. The results of the final ARIMA models were favourable in that few parameters were required; the one bedroom model required only seven parameters (including mean term) and the two/three bedroom model required six parameters. The two models can be described as $ARIMA(2,0,2)(1,0,0)_7(1,1,0)_{364}$ and $ARIMA(3,0,1)(1,0,0)_7(0,1,1)_{364}$ respectively. The parameter estimates of the one bedroom model are shown in Equations 18 and the parameter estimates of the two bedroom model are shown in Equation 19. Parameter estimate detail and model fit statistics shown in Appendix C. It should be noted that negative forecasts were replaced with zero for all ARIMA forecasts.

$$(1-B^{364})Y_t = 28.939 + \left(\frac{1-.209B^2+.116B^4}{(1-.875B+.118B^3)(1-.175B^7)(1+.381B^{364})} \right) e_t \tag{18}$$

$$(1-B^{364})Y_t = \left(\frac{(1-.921B)(1-.468B^{364})}{(1-1.868B+1.044B^2-.169B^3)(1-.094B^7)} \right) e_t \tag{19}$$

where:

Y_t Observation at time t

B Backward shift operator (e.g. $(1\text{-}B^2)y_t = y_t - y_{t-2}$)

e_t Random disturbance (error) at time t

3.6 Linear Regression (LR)

Linear regression is a common correlation-based statistical technique that has at least one input variable, and calculates coefficients for each input variable so that the model estimate (response variable) is a linear combination of the input variables. If a linear combination of data inputs is not appropriate, often the variables can be transformed so that estimates are still possible within a linear regression framework (e.g. taking logs of the data or taking z-scores of the data). For univariate time series data, the modeller often creates separate binary input variables to specify mutually exclusive seasonal periods. For example, if a modeller wanted to calculate regression coefficients for 12 periods (months) within a dataset, she may create 11 new input variables (one month being the default month to prevent perfect collinearity among input variables). In this example, a specific monthly input variable (say February) would be one if the observation was taken from this month, and zero otherwise. In this manner, each observation would have at most one monthly variable that was non-zero.

To continue the example, assuming positive observations, if January was taken to be the default month, the calculated regression coefficients for the other 11 months can be interpreted as the difference between the specified month and January. If the coefficient for February was positive, then the expected seasonal impact of February on observed data values would be higher than that for January. Conversely, if the coefficient for February was negative, one would expect lower observed values for February than that of January. In this manner, binary variables were created to represent specific seasonal periods for the resort.

The resort managers provided 13 different periods they viewed as distinct. Sample periods included seven winter periods (winter is defined as all dates in which the ski hill is open for downhill skiing) and six summer periods. A period could be defined both by hard dates (e.g. December 20 to January 4 – *Holiday Period*) and soft dates (e.g. Friday prior to President's Day to the following Saturday – *President's Week*). Binary variables were included in the input dataset to represent these periods (e.g. the binary variable for *President's Week* is one if an observation falls within that week and zero otherwise). The 13 periods were broken down further by specifying weeks within periods and the regression was run to see if the additional parameters were significant at a p=.05 level. In this way periods were further segmented or combined until each parameter was significant.

The final one bedroom model contained 26 seasonal period parameters while the final two bedroom model included 29 seasonal period parameters. The data was also partitioned by day of week, with a separate parameter for each day of the week if significant at a p=.05 level. The final one bedroom model contained one day of week parameter while the final two bedroom model contained two day of week parameters. Next, partitions for day of week seasonal period interactions were created. After some testing, only a weekend-seasonal period interaction (weekend defined as a Friday or Saturday night) was found to be significant and for only some of the seasonal periods. The final one bedroom model contained nine weekend period interaction parameters while the final two bedroom model contained only one weekend period interaction parameter. Finally, the model included a year term to capture broad-based yearly trend.

A linear regression was not appropriate to model three bedroom demand as it would lead to heteroskedasticity since small count data violate the assumption of normality necessary for linear regression. One and two bedrooms, on the other hand, have count data that are large enough to adequately approximate a normal distribution. In order to overcome the heteroskedasticity problem inherent in small count data a Poisson regression model was used to model three bedroom demand. Poisson regression employs a quasi-maximum likelihood technique which finds conditional probabilities based on values of the explanatory variable (see Woolridge, 1999 for a full discussion of Poisson regression analysis). Essentially, the benefit of using a Poisson distribution is that it can be fully described by the mean term alone, and this is exploited to form a log-likelihood function in order to calculate parameter estimates. Mathematically, the probability that demand equals a specific value (conditional on input variables is shown in Equation 20). Interpretation of the parameter estimates themselves is quite similar to linear regression. However, rather than the $x\beta$ terms predicting y directly as in linear regression, $\exp(x\beta)$ predicts y in a Poisson regression.

$$P(y = k \mid x) = \exp[-\exp(x\beta)][\exp(x\beta)]^k / k! \quad (20)$$

where:

y	three bedroom demand
k	value for three bedroom demand (k = 0,1,...)
β	input data coefficients
x	data input values (i.e. seasonal binary variables)

The input data for the Poisson regression was very similar to the input data for the linear regression. The final model included 20 seasonal period parameters, 1 day of week parameter, and 8 weekend period interaction parameters. Furthermore, no yearly trend in the number of units booked was observed for the three bedroom model so no yearly trend component was included in the model. The linear regression and Poisson regression results for the in-sample period are shown in Table 3, with detailed parameter estimates and model fit statistics for the one and two bedroom models shown in Appendix D, and the parameter estimates and model fit statistics for the three bedroom model shown in Appendix E.

Table 3: Long-term linear regression model results (May 15, 1998 to April 29, 2002)

Model	Type of model	# of parameters	Classes of parameters	R^2	# of observations
1 bedroom	Linear regression	38	• General intercept (1) • Period intercepts (26) • Day of week intercepts (1) • Demand trend (1) • Weekend period interactions (9)	.77	1,446
2 bedroom	Linear regression	37	• General intercept (1) • Period intercepts (32) • Day of week intercepts (2) • Demand trend (1) • Weekend period interactions (1)	.72	1,446
3+ bedroom	Poisson regression	29	• Period intercepts (20) • Day of week intercepts (1) • Weekend period interactions (8)	.30*	1,446

*Minimizing SSE (sum of square errors) is not the objective function of a Poisson regression; however, a linear regression was run with the same parameters to get an approximate R^2.

3.7 Non-Linear Regression (NL)

The most important decision to be made with respect to a customized long-term model was an appropriate functional form that would specifically capture the resort's demand situation (rather than say a more automatic model such as a HW model). Initially it had been thought that demand as a percentage of capacity may be a good measure of demand. Over the four year period for which data had been provided, the lodging capacity had increased each year in roughly a linear trend. However, the observed occupancy rates for those periods were not constant. What was happening was that the number of units occupied increased when capacity increased, but not in the same proportion as the increase in capacity. For instance, the highest demand period of the year, the New Year's holiday, would generally be close to capacity regardless of the absolute increase in capacity for the year. On the other hand, slow shoulder periods (e.g. early May and early November) showed almost no increase in demand year over year regardless of newly added capacity. Other periods, defined as mid to high season, showed an increase in demand year over year, but not in the same proportion as the increase in capacity. As a result, in an attempt to capture the idiosyncratic demand elements of the resort studied, a nonlinear regression model was estimated with two components (Equation 21).

$$UNITS_{day,year} = DEMAND_{year} * SHARE_{day} \tag{21}$$

$$DEMAND_{year} = \beta_0 + \beta_1 YEAR \tag{22}$$

$$\text{SHARE}_{\text{day}} = \text{f(seasonal period, day of week, seasonal period day of week interaction)} \tag{23}$$

$$SHARE_{day} = \frac{1}{1+\exp(-x\beta)} \tag{24}$$

where:

β input data coefficients

x data input values (i.e. seasonal binary variables)

The first component is an estimate of the maximum potential daily demand in a given year (Equation 22 and shown graphically for one bedroom units in Figure 2). The second component is a logistic function that determines the share of maximum daily demand (up to 100%) based on seasonal factors (Equation 23 and shown graphically for one bedroom units in Figure 3). A logistic function is appropriate as a share of demand function since it is bounded between zero and one (Equation 23 is expressed mathematically in Equation 24). The β coefficients are thus estimated so that the linear $x\beta$ terms are extremely positive in high demand periods and extremely negative in low demand periods.

Estimating daily demand as a share of estimated maximum potential daily demand was expected to provide a better estimate of demand than demand as a share of capacity. This hypothesis was supported by analyzing data during the four year sample period; increases in yearly capacity often did not lead to a proportionate increase in yearly demand. A logistic share of demand component creates a multiplicative seasonal component rather than an additive seasonal component as in linear regression. A multiplicative model is more intuitive since the variation in demand among seasonal periods is likely to increase with overall yearly demand rather than staying constant. Stated differently, an additive model is based on the assumption that the difference in units occupied between high and low demand periods remains constant from year to year. A multiplicative model, on the other hand, is based on the assumption that the difference in units occupied between high and low demand periods is a proportion of overall maximum demand. In a multiplicative model, as overall yearly demand increases, the difference in units occupied between high and low demand periods increases.

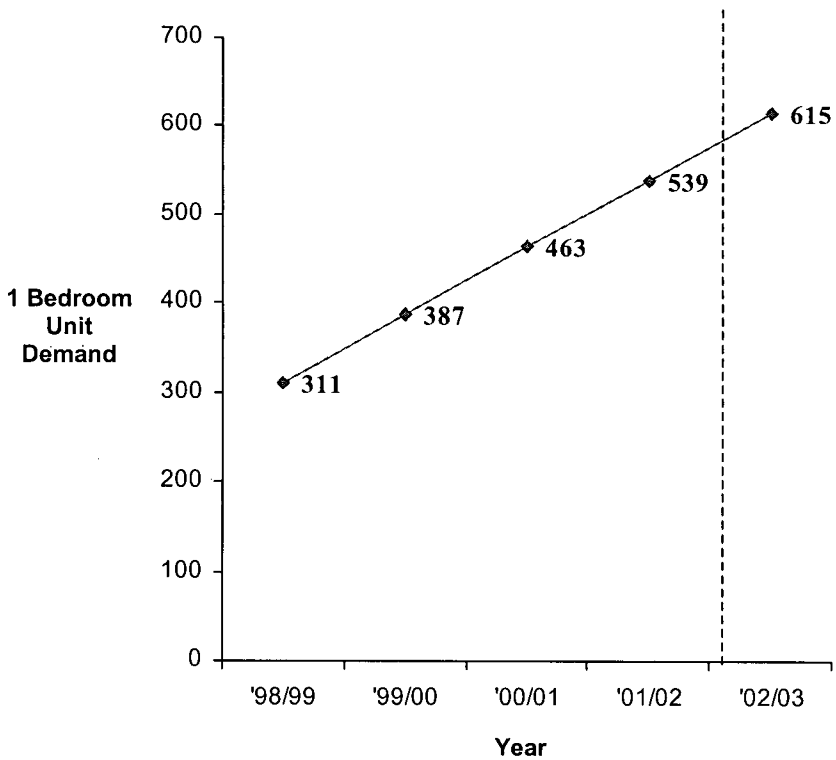

Figure 2: Estimated maximum daily demand for one bedroom units by year

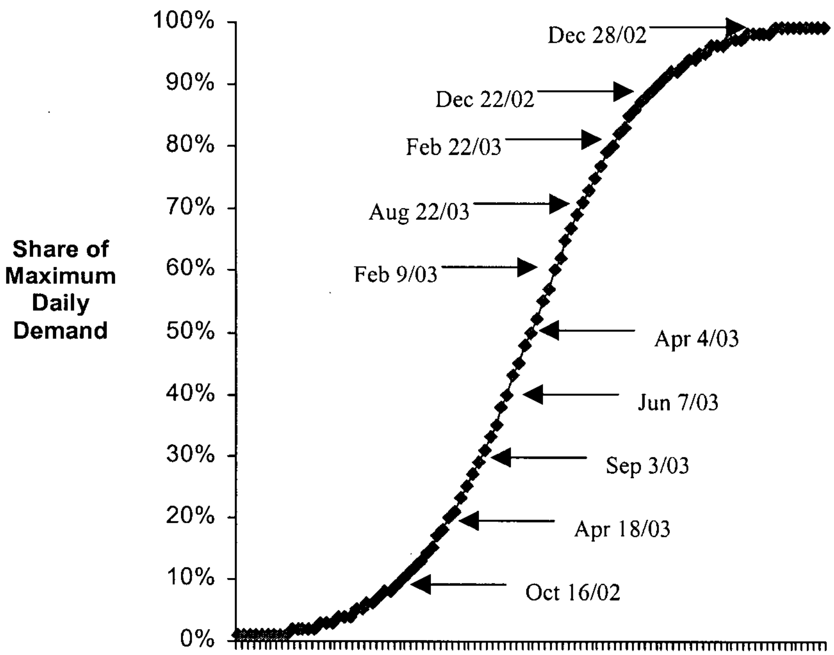

Figure 3: Predicted share of maximum daily demand for one bedroom units on selected dates in 02/03 (615 units = 100%)

Three bedroom demand, as mentioned earlier, had small count data (capacity less than 15 with an average number of occupied units less than 3). Similar to linear regression, a nonlinear regression would have had heteroskedasticity problems with such count data. As a result, the same three bedroom Poisson regression estimates that were combined with linear regression estimates were combined with nonlinear regression estimates for aggregate demand estimates.

The input variables for the nonlinear regression model belong to the same categories of input variables included in the linear regression model: seasonal period parameters, day of week parameters, seasonal period weekend interaction parameters, and a yearly trend parameter. The difference is that all the inputs excluding the yearly trend input were included in a logistic function (Equation 23) that was then combined with a yearly demand function based on year. The final one bedroom logistic function contained 32 seasonal period parameters, 2 day of week parameters, and 9 weekend period interaction parameters. The final two bedroom logistic function included 33 seasonal period parameters, 5 day of week parameters, and 3 weekend period interaction parameters. Nonlinear model results are shown in Table 4 while detailed parameter estimates and model fit statistics are shown in Appendix F. Results of the three bedroom Poisson regression are shown in Appendix E.

Table 4: Long-term nonlinear regression model results (May 15, 1998 to April 29, 2002)

Model	Type of model	# of parameters	Classes of parameters	R^2	# of observations
1 bedroom	Nonlinear regression	45	• Period intercepts (32) • Day of week intercepts (2) • Demand trend (2) • Weekend period interactions (9)	.82	1,446
2 bedroom	Nonlinear regression	43	• Period intercepts (33) • Day of week intercepts (5) • Demand trend (2) • Weekend period interactions (3)	.78	1,446
3+ bedroom	Poisson regression	29	• Period intercepts (20) • Day of week intercepts (1) • Weekend period interactions (8)	.30*	1,446

*Minimizing SSE (sum of square errors) is not the objective function of a Poisson regression; however, a linear regression was run with the same parameters to get an approximate R^2.

3.8 Long-Term Model Comparison

The long-term models were created to forecast demand more than 90 days prior to a target date. As a result, the five long-term models (RW, HW, LR, ARIMA, NL) were compared within an in-sample period as well as within an out of sample period. Appropriate functional forms for all long-term models were constructed using the entire four year sample. The model estimates were then forecast out for year 4 within sample and the results compared. For the LR and NL model, the entire sample was used to calculate parameter estimates (input coefficients) and these same coefficients were used for the in-sample forecasts. The HW and ARIMA models also used the entire four year sample to determine model structure (e.g. number and type of parameters for the ARIMA models). However, parameter estimates for these models vary by day, so the estimates for year 4 were based on data up to year 3 and then forecast out for year 4. RW is not based on any model, and demand estimates were simply taken from 364 days prior.

The results of the five models in the in-sample period are shown in Table 5. Three different error measures are shown: MSE (mean square error), MdAPE (median absolute percentage error), and CumRAE (cumulative relative absolute error). For the in-sample period, the NL model is shown to be superior on all error measures although Armstrong & Collopy suggest CumRAE is the most robust error metric for model comparison in this instance. A CumRAE value of .738 indicates that the NL model contains 73.8% of the cumulative error of the RW method, thereby indicating a 26.2% improvement over RW. A CumRAE value of 1.356 for the HW model indicates estimates that are 35.6% more inaccurate than RW.

Table 5: In-sample long-term model comparisons (April 28, 2001 to April 27, 2002)

Model	MSE	CumRAE	MdAPE
Random walk (RW)	12,711	1.000	25.7%
Nonlinear regression (NL)	6,333	.738	20.4%
ARIMA	10,431	.993	31.8%
Holt-Winters multiplicative (HW)	22,175	1.356	37.5%
Linear regression (LR)	9,113	.992	33.9%

The models were also compared out of sample. Since long-term estimates are forecasts more than 90 days prior to a target date, the models forecast demand more than 90 days after the last date of in-sample data (April 29, 2002). As a result the out of sample period was July 29, 2002 to November 30, 2002 and the results are shown in Table 6. As can be seen in Table 6, the NL model is still superior, but by a much narrower margin of improvement (3.3%) than in-sample (26.2%). As at December 1, 2002 the resort unexpectedly lost 105 units of capacity due to a hotel property switching reservation management provider. As a result, the assumption underlying the NL and LR models was violated, and the quality of estimates significantly deteriorated. The loss of capacity also increased the error in the other long-term forecasting methods (see MdAPE measures) but since the other methods were not based on an increasing yearly trend in demand they were not as adversely affected. The out of sample period post December 1, 2002 is shown in Table 7. In this period, the RW method is far superior to other long-term methods; providing a minimum 38% improvement over all other long-term models.

Table 6: Out of sample long-term model comparisons (July 29, 2002 to November 30, 2002)

Model	MSE	CumRAE	MdAPE
Random walk (RW)	10,987	1.000	23.1%
Nonlinear regression (NL)	7,187	.967	30.8%
ARIMA	9,932	1.098	32.0%
Holt-Winters multiplicative (HW)	24,273	1.610	35.6%
Linear regression (LR)	10,475	1.398	72.0%

Table 7: Out of sample long-term model comparisons (December 1, 2002 to January 22, 2003)

Model	MSE	CumRAE	MdAPE
Random walk (RW)	3,325	1.000	13.8%
Nonlinear regression (NL)	11,176	1.698	28.5%
ARIMA	5,269	1.377	27.3%
Holt-Winters multiplicative (HW)	37,106	3.033	35.6%
Linear regression (LR)	8,580	1.667	26.0%

4 SHORT-TERM MODELS

Most papers on hotel forecasting employ one of two approaches: a long-term forecast or a short-term forecast. The long-term forecast uses past years' data on daily occupancy to predict daily occupancy in the future. Long-term forecasts ignore the buildup of bookings for dates in the future (they are not adjusted for actual bookings to date). Short-term forecasts, on the other hand, analyze the build-up of bookings for a single future date (target date), and then project final demand for that target date based on actual bookings to date. Some short-term forecasts treat target dates in isolation, ignoring the final occupancy figures of years past while others integrate both actual bookings to date as well as final occupancy figures from years past. This paper analyzes integrative short-term forecasts as they use all available information and will be shown to provide better estimates than either models based entirely on past occupancy data or models based entirely on bookings to date. The two short-term methods to be studied include additive pickup (AP) and a customized booking curve (BC) model which is based on a non-linear model nearly identical to the long-term NL model.

4.1 Additive Pickup (AP)

AP is a simple yet robust short-term forecasting method which automatically integrates prior year occupancy data as well as actual bookings to date. It can be thought of as a detailed random walk. The AP estimate for a target date is bookings to date plus expected pickup. AP is used extensively in the airline industry for forecasting passenger pickup (short-term passenger demand); for specific model specifications see Harris & Marucci, 1983 and L'Heureux, 1986. Often a deviant of a direct AP method is used where an exponential moving average of a subset of flights' pickup is used to predict pickup for a current flight. The subset of appropriate flights may be based on day of week, seasonal period, or operating environment such as a fare sale.

For the resort studied, expected pickup was defined as the pickup that was experienced in the year prior (364 days prior so that the pickup is from the same day of week and seasonal period). As an example, suppose it is 15 days prior to a target date of December 20, 2003 and there are 500 bookings to date. To find the expected pickup one would look at last year's bookings for the target date December 21, 2002 in the 15 day period prior to the target date; let's say there were 300 bookings in that period. In this case, the AP estimate for December 20, 2003 is 800 (500 bookings to date + 300 expected pickup). The one complication that should be mentioned is that some of the bookings to date will cancel. The method used in this paper was to include all cancellations as part of the expected pickup. For example, suppose the bookings to date Y days prior to a target date were 500. Further suppose that in the year prior, Y days before the target date, 400 new bookings were made in the Y day interval and 100 cancellations were made in the

Y day interval. Then there would be an expected pickup of 300 units (400 new bookings less 100 cancellations).

4.2 Booking Curve Estimate (BC)

In order to understand the BC estimate it is important to explore the concept of a booking curve. The typical booking curve (pattern of bookings over time) for a specific date in the future (target date) is generally a convex curve, with the most bookings occurring in the week immediately prior to a target date (see Figure 4).

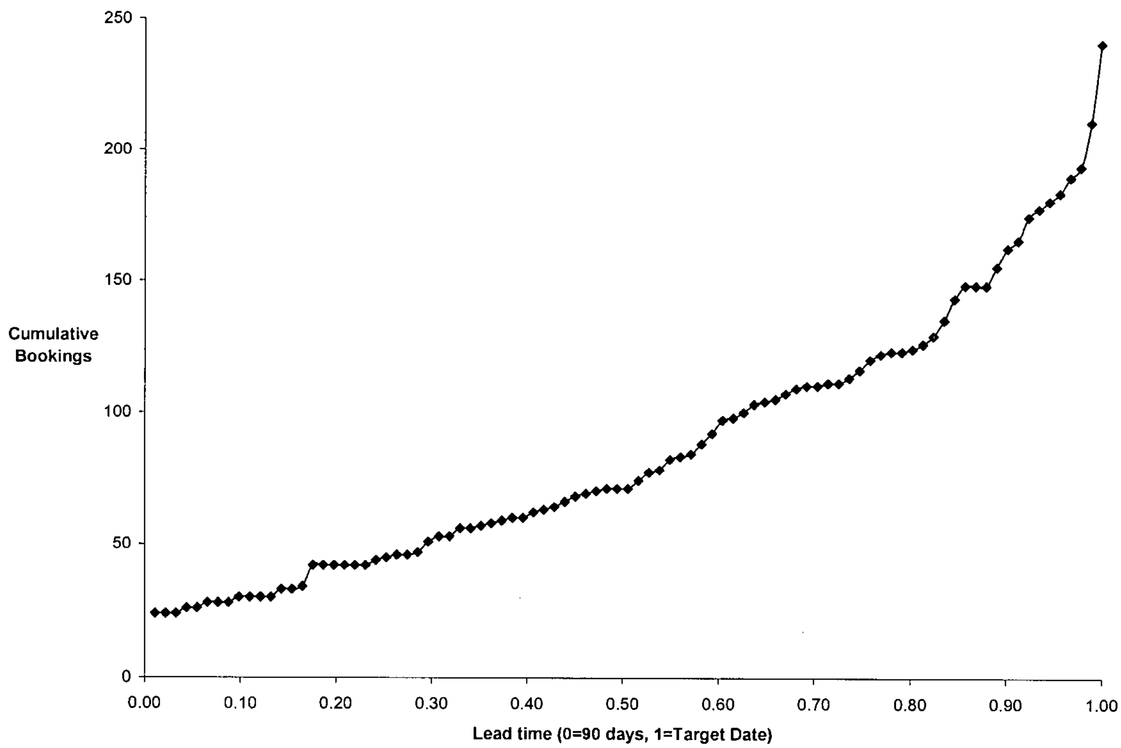

Figure 4: Typical booking curve (booking curve for target date of July 5, 2001)

However, the booking curve changes in an absolute sense (overall number of bookings) and relative sense (shape of booking curve) depending on the time of year. High demand periods generally yield booking curves that are concave with high overall bookings while low demand periods produce curves that are convex with low overall bookings (see Figure 5A and Figure 5B). Resort hotels tend to display more seasonality than business-oriented hotels and as such the variation of booking curves for a resort hotel tend to be larger than that of a business hotel.

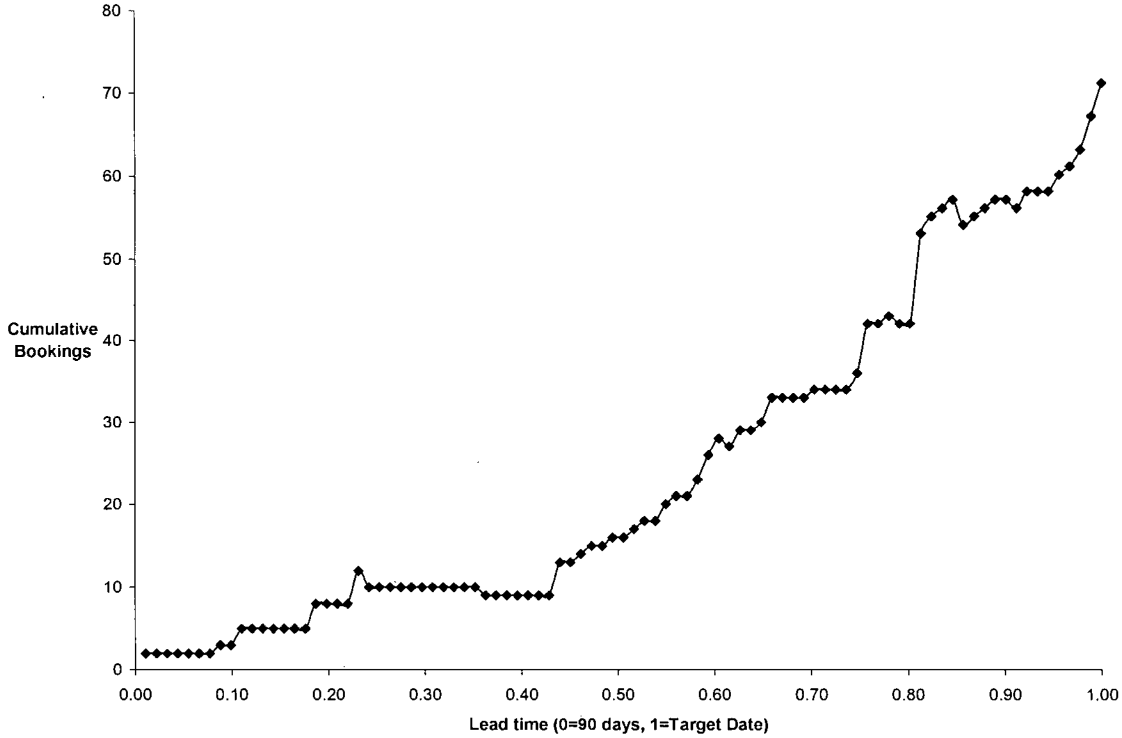

Figure 5A: Low demand period – convex booking curve (booking curve for target date of December 6, 2001)

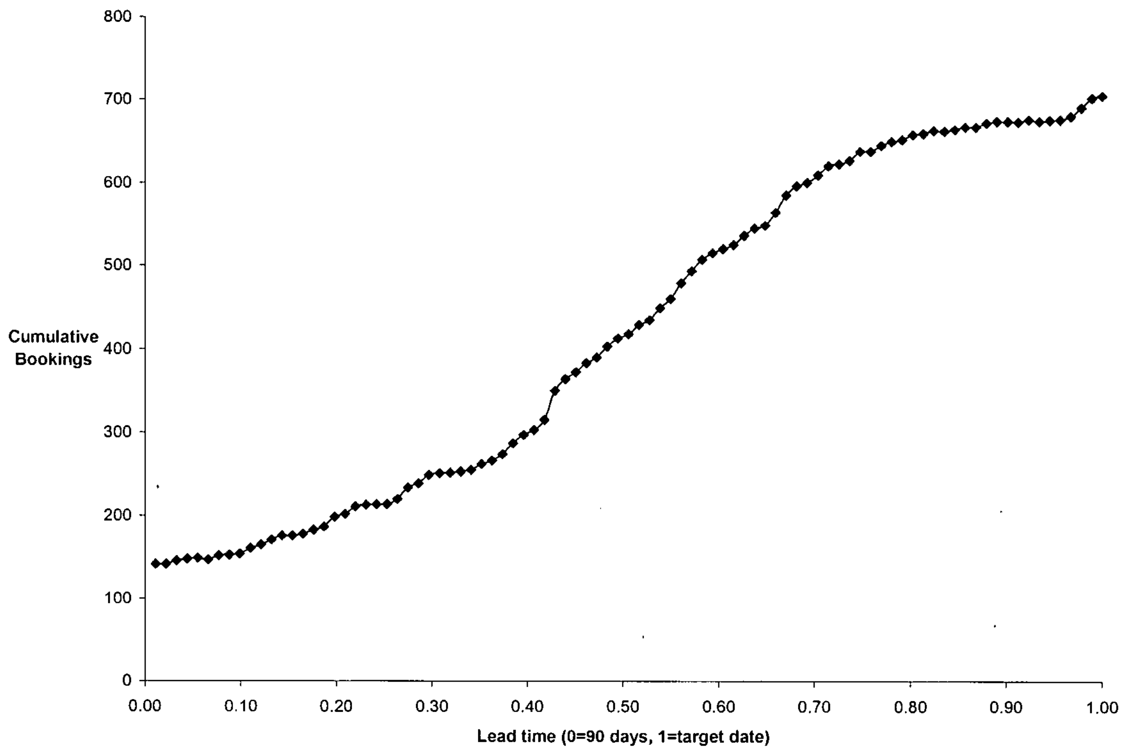

Figure 5B: High demand period – concave booking curve (booking curve for target date of December 28, 2001)

This paper utilizes an approach similar to that used by Rajopadhye et al. (1999) in which long-term estimates are used to predict future demand, and these estimates are continually adjusted based on bookings to date. The process to achieve the resort's short-term booking curve estimate is outlined in Figure 6.

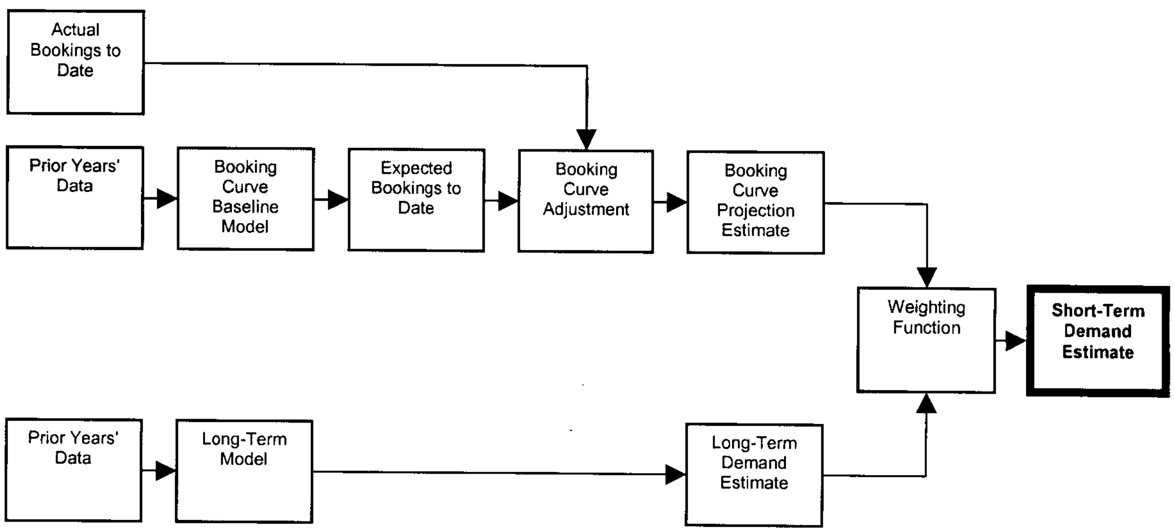

Figure 6: Flowchart of forecasting process for booking curve short-term estimate

The short-term demand estimate is a weighted average of the long-term demand estimate and the booking curve projection estimate. The long-term estimate is derived from a model based on final daily occupancy figures in years past. In this paper the NL model is used to produce a long-term estimate (although theoretically any of the long-term models could be used). The booking curve projection estimate, on the other hand, is composed of two steps. First, a baseline booking curve model uses the pattern of bookings to date in year's past to create expectations of current bookings to date. Second, expected bookings to date and actual bookings to date are input into an adjustment function that creates an estimate of final demand (booking curve projection estimate). Finally, a weighting function (based on lead time) combines the long-term estimate and booking curve projection estimate to come up with the short-term demand estimate. In this way, the short-term demand estimate is continually updated as new bookings are made, existing bookings are cancelled, and the target date approaches.

4.2.1 Booking curve baseline model

The booking curve baseline model is similar to the long-term NL model in its structure (two component nonlinear regression for one and two bedrooms and Poisson regression for three plus bedrooms). However, rather than provide a single point estimate for a target date, the model provides an estimate for each of the 90 days prior to a target date as well as the target date itself. The major difference between the short-term BC model and the long-term NL model is the inclusion of a lead time element. A lead time element enables the model to account for the increase in bookings as the target date approaches. The lead time element is also interacted with seasonal period binary variables so that the shape of the booking curve can vary by period (i.e. concave for high demand days and convex for low demand days). The lead time parameters as well as the lead time seasonal period interaction parameters are captured in the logistic component of the model. The logistic function for share of demand was used since it is a good representation of the booking curve (see Figure 7). An approaching target date is equivalent to following the logistic curve from left to right for a specified interval.

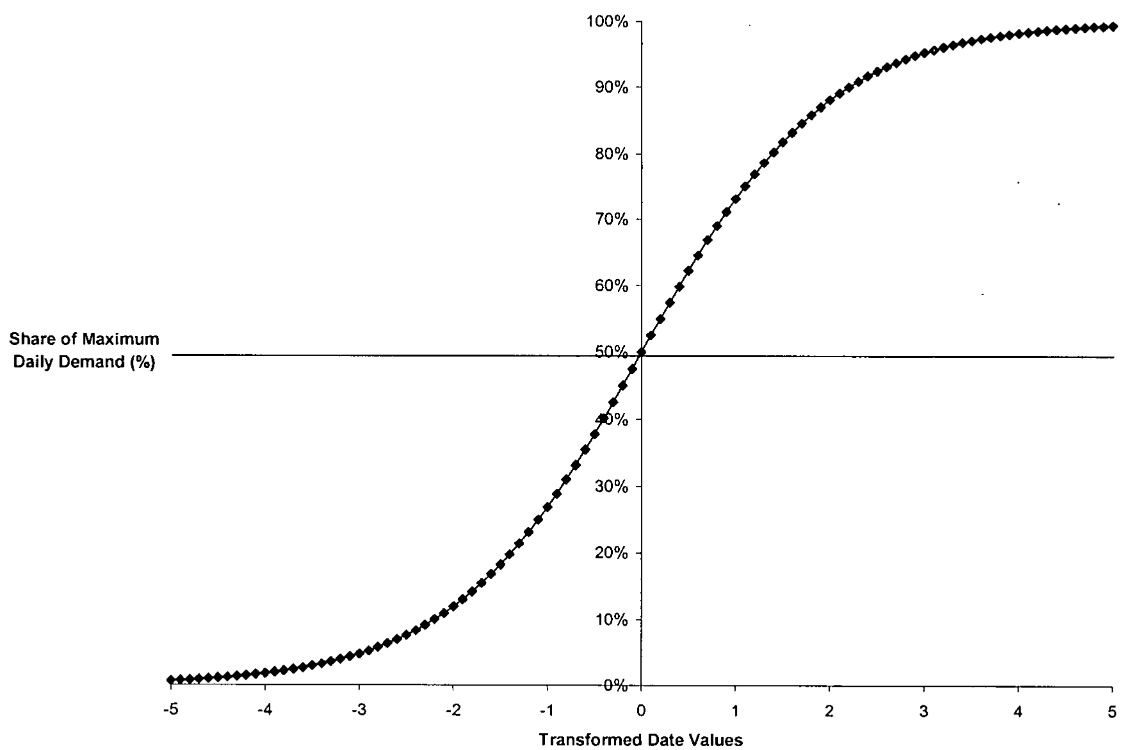

Figure 7: Generic logistic curve

The left hand side of the logistic function closely resembles the 90-day booking curve for most days with a traditional convex build up (imagine Figure 5A superimposed on the left hand side of Figure 7). The right hand side of the logistic function closely resembles a 90-day booking curve for a high demand day (imagine Figure 5B superimposed on the right hand side of Figure 7). Therefore, choosing an appropriate intercept along the logistic function for a specific target date (to mark the beginning of a specific time interval) as well as including lead time seasonal period interactions provides a flexible functional form to approximate booking curves for a specific target date and lead time. The large amount of variation explained by the booking curve baseline regression models is evidence of the appropriateness of the logistic functional form within the nonlinear regression model (Table 8). The baseline regression parameter estimates and model fit statistics are shown in Appendix G.

Table 8: Booking curve baseline regression model results (May 15, 1998 to April 29, 2002)

Model	Type of model	# of Parameters	Classes of parameters	R^2	# of observations
1 bedroom	Nonlinear regression	112	• Period intercepts (57) • Day of week intercepts (5) • Demand trend (2) • Lead-time elements (2) • Period lead-time interactions (24) • Weekend period interactions (22)	.84	131,586
2 bedroom	Nonlinear regression	85	• Period intercepts (47) • Day of week intercepts (5) • Demand trend (2) • Lead-time elements (2) • Period lead-time interactions (15) • Weekend period interactions (14)	.82	131,586
3+ bedroom	Poisson regression	80	• General intercept (1) • Period intercepts (50) • Day of week intercepts (5) • Lead-time elements (1) • Period lead-time interactions (12) • Weekend period interactions (11)	.41*	131,586

*Minimizing SSE (sum of square errors) is not the objective function of a Poisson regression; however, a linear regression was run with the same parameters to get an approximate R^2.

4.2.2 Booking curve adjustment

Expected bookings to date for a specific target date and lead time from the booking curve baseline model is used as a baseline figure to be compared with actual bookings to date. A booking curve projection of final demand (number of units demanded at the target date when lead time equals zero) is thus the expected bookings for the target date adjusted by a function of the actual bookings to date. Five different approaches for a booking curve projection were attempted. The idea was to adjust the projection by an amount proportional to the deviation (actual less expected bookings) at a certain lead time (see Appendix H for calculations and notation for the first four approaches). The fifth approach was somewhat different in that it employed an ARIMA model to estimate the pattern of booking curve errors to date, and then projected that pattern to the target date. The results of the ARIMA model were mixed, and the approach was ultimately discarded due to additional complexity in computation and implementation on resort (see Appendix I for results of the ARIMA approach). Of the five approaches, the direct multiplicative approach was the most straightforward approach and led to the greatest reduction in squared error although all of the first four methods provided very similar improvements. In the direct multiplicative approach, the booking curve projection results from multiplying the baseline estimate by the ratio of actual to expected bookings to date (see Equation 25).

$$\left(\frac{AB_{LT=Y}}{EB_{LT=Y}}\right)EB_{LT=0} = BCE_{LT=Y} \tag{25}$$

where:

$AB_{LT=Y}$	Actual bookings to date Y days prior to the target date (lead time = Y)
$EB_{LT=Y}$	Expected bookings to date Y days prior to the target date (from baseline booking curve)
$EB_{LT=0}$	Expected bookings on the target date (from baseline booking curve)
$BCE_{LT=Y}$	Booking curve estimate of final demand at Y days prior to target date

4.2.3 Short-term weighting function

The short-term estimate is a weighted average of the booking curve projection and the long-term estimate. Econometric literature provides many examples of situations where combined forecasts provide superior results to single forecasts. In fact, combined forecasts will always be optimal as long as forecasts are unbiased (Min & Zellner, 1993). However, Min & Zellner go on to prove that combining biased forecasts does not necessarily provide superior forecasts. As a result, a linear regression model (no intercept) was used to combine forecasts at each lead time as this was a method that would minimize the squared error regardless of whether or not bias was present. This model allowed one estimate to be weighted between 0% and 100% depending on its contribution to MSE. In fact, bias was likely for the short-term model given the underlying yearly trend in demand was likely to either overestimate or underestimate actual yearly demand, which would then bias all daily estimates.

The short-term model weight changes at different lead times since the error of the booking curve projection is not consistent across lead times. Instead, booking curve estimates at long lead times (i.e. 90 days prior to a target date) have higher errors than booking curve estimates at short lead-times. This is because booking curve projections at long lead times have fewer bookings to date in which to make a forecast and must forecast further out. Long-term estimates, on the other hand, do not vary based on bookings to date as they are constructed entirely from prior years' data. Figure 8 compares the mean square error of long-term NL estimates and short-term booking curve projection estimates at different lead times over a three year in-sample period (May 14, 1999 to April 29, 2002).

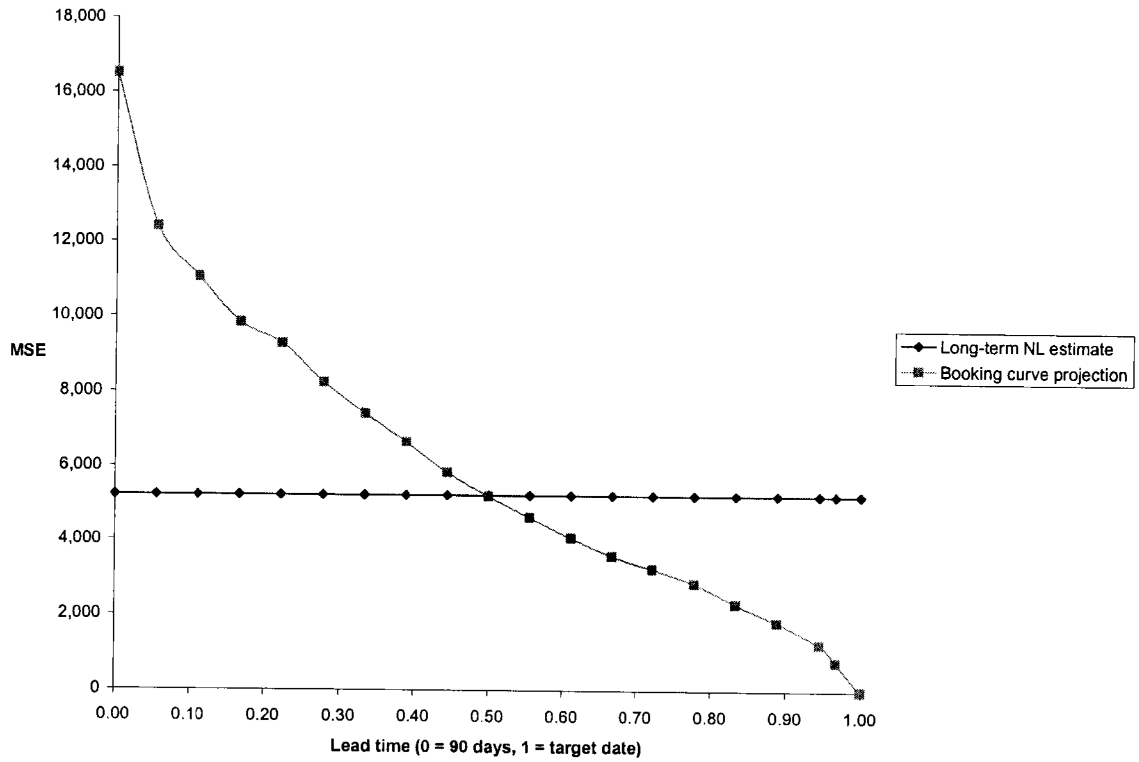

Figure 8: Mean square error of one bedroom long-term NL estimates and short-term booking curve projections at different lead times

In order to integrate the time-dependent error of booking curve projections into better short-term estimates, a weighting function is used to balance the contribution of long-term estimates and booking curve projections. The weighting function between the long-term estimate and the booking curve projection in this paper is similar to the approach taken to projected demand estimates in Rajopadhye et al. (1999). Rajopadhye et al. update their weighting function based on the mean square error (MSE) of a short-term ARIMA forecast and MSE of a long-term ARIMA forecast. Since short-term forecasts typically have smaller MSE than do long-term forecasts as the target date nears, the short-term forecasts are weighted more heavily closer to the target date. Similar to the MSE ratio calculated in Rajopadhye et al., the weighting function in this paper is based on a linear regression (no intercept) of the sample days (1,446 days) at each lead time (from 90 days out to 1 day out) to determine the optimal weighting between long-term NL estimate and booking curve projection (see Equation 26 and Equation 27).

$$STE_{LT=Y} = \hat{\alpha}_{LT=Y} BCE_{LT=Y} + (1 - \hat{\alpha}_{LT=Y}) LTE \tag{26}$$

$$\hat{\alpha}_{LT=Y} = \frac{AB_{LT=0} - LTE}{BCE_{LT=Y} - LTE} \tag{27}$$

where:

$AB_{LT=0}$	Actual bookings on the target date (LT=0)
$STE_{LT=Y}$	Short-term estimate Y days prior to target date
$BCE_{LT=Y}$	Booking curve projection Y days prior to target date
LTE	Long-term NL estimate (does not change across lead times)
$\hat{\alpha}_{LT=Y}$	Regression weighting parameter Y days prior to target date

Once the weighting parameter, alpha, was determined for each lead time, alpha was estimated as a general function of lead time (T) in order to remove any idiosyncratic effect that may have occurred at a specific lead time. Alpha as a general function of lead time (T) explained over 99% of the variation in the original alpha estimates and hence the general alpha function was used as the weighting for all short-term estimates. See Equations 28-30 for the general alpha functions and Figure 9A to 9C for a graphical representation of the weighting function. Figure 10 is identical to the MSE reported in Figure 8 but with the addition of the short-term estimate. As can be seen from Figure 10, the short-term estimate provides much-improved forecasts across all lead times.

$$\hat{\alpha}(1\_bedroom) = .1805 + .6476T - .5900T^2 + .7435T^3 \quad (28)$$

$$\hat{\alpha}(2\_bedroom) = .2377 + .8530T - .5088T^2 + .3981T^3 \quad (29)$$

$$\hat{\alpha}(3\_bedroom) = .3321 + .7028T - .1320T^2 + .1102T^3 \quad (30)$$

where:

$\hat{\alpha}$ Short-term estimate weighting parameter

T Lead time expressed between 0 and 1. Y days prior to a target date; T = (91-Y)/91. Therefore, at Y=0; T=1, at Y=90; T=.011.

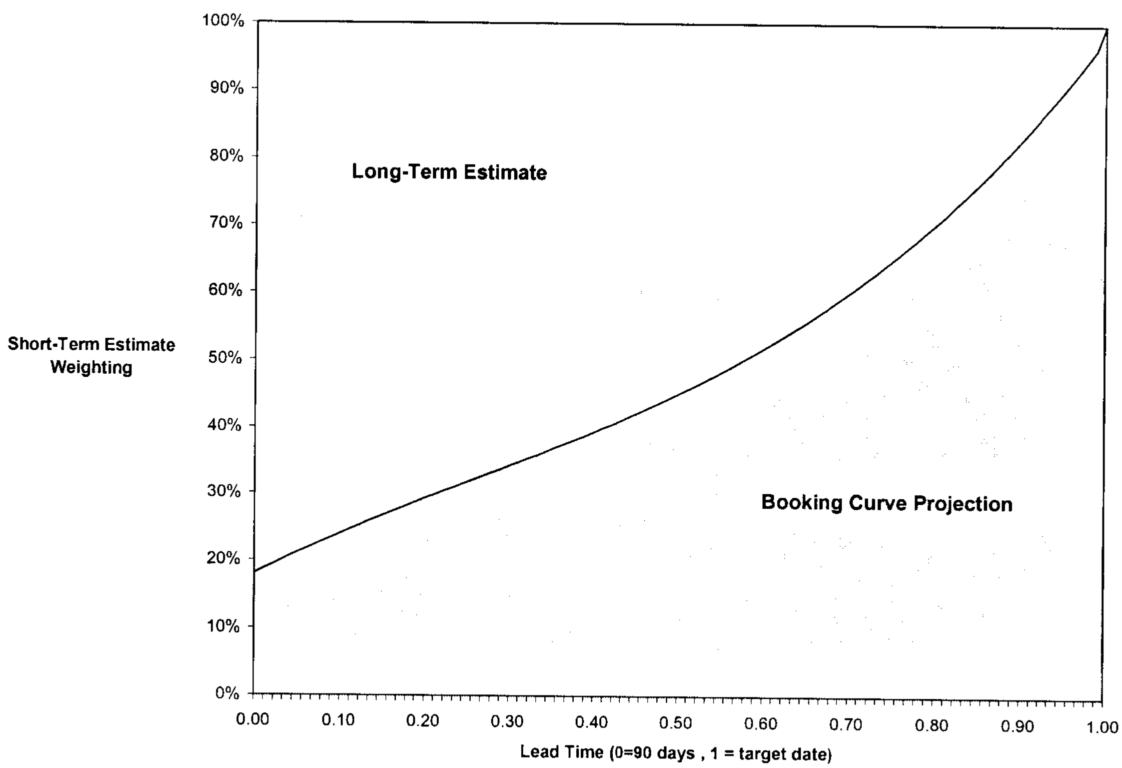

Figure 9A: One bedroom weighting function for short-term estimate

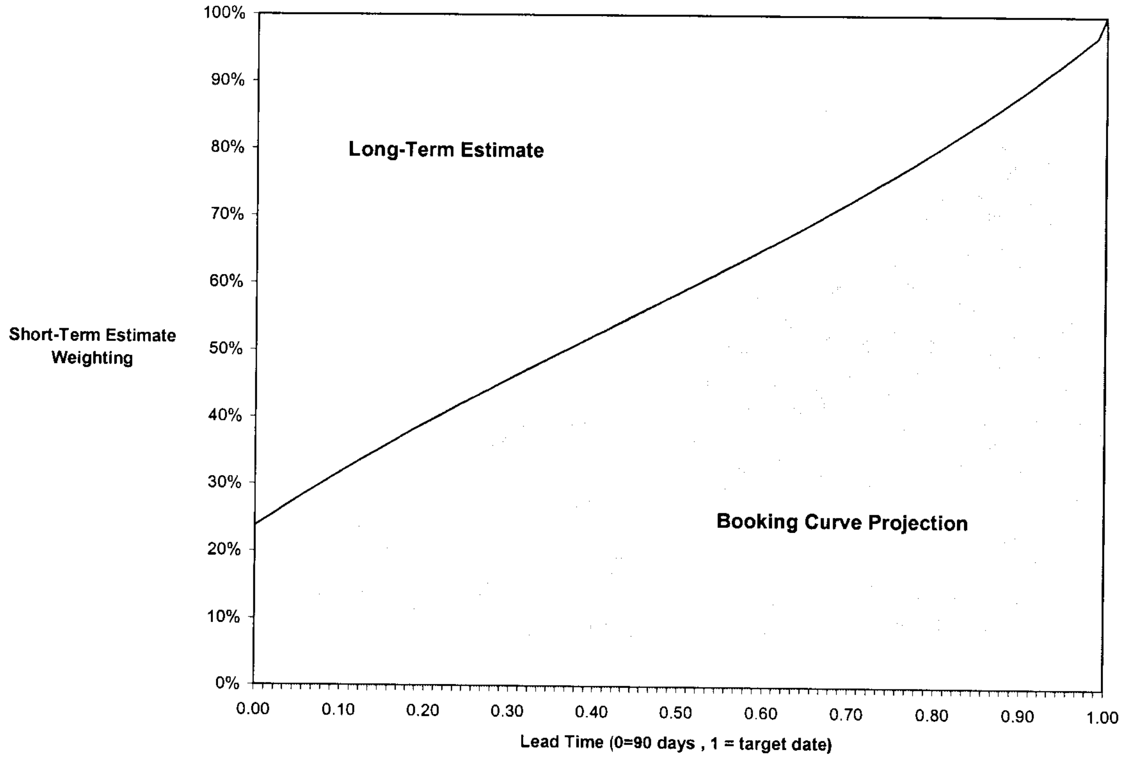

Figure 9B: Two bedroom weighting function for short-term estimate

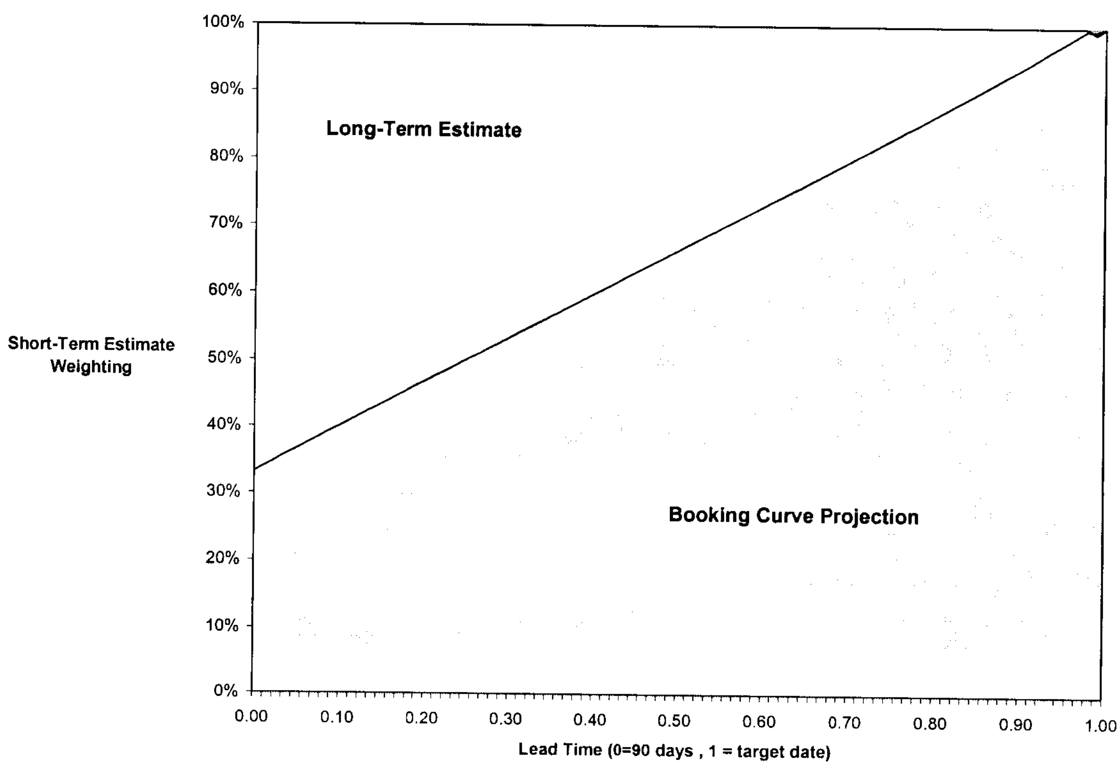

Figure 9C: Three plus bedroom weighting function for short-term estimate

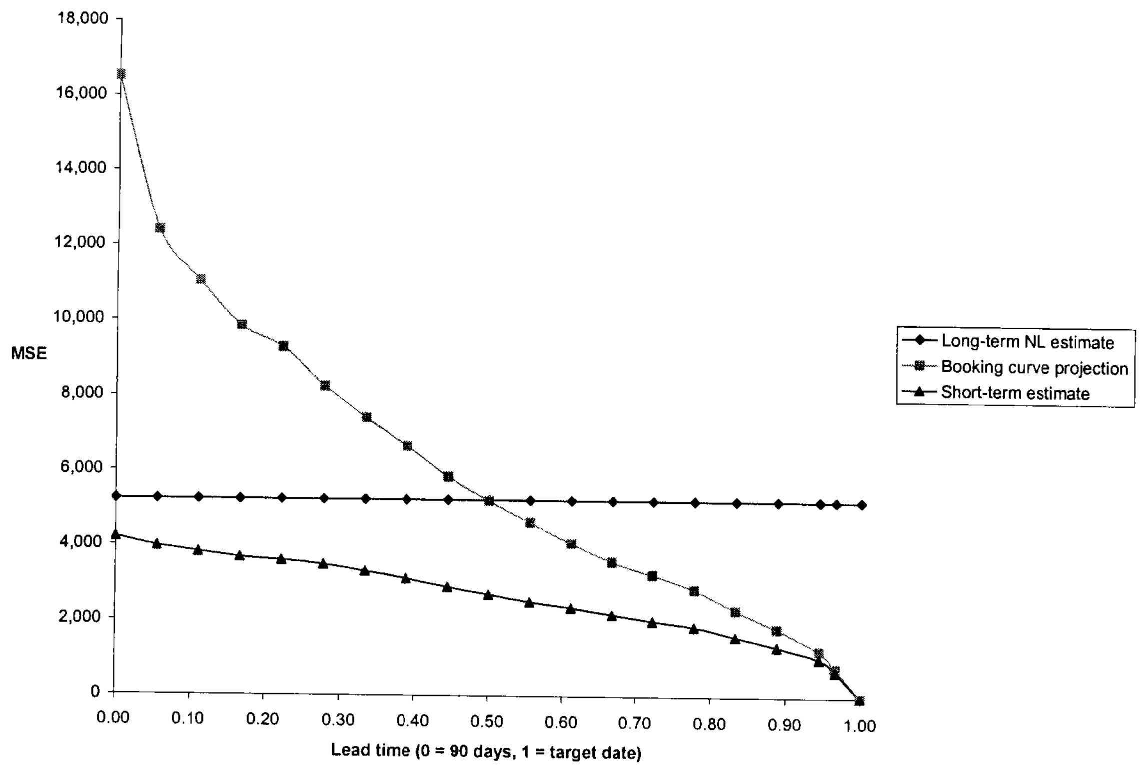

Figure 10: Mean square error of one bedroom long-term NL estimates, short-term booking curve projections, and short-term booking curve estimates at different lead times

4.3 Comparing Short-Term Forecasts

Table 9 compares short-term forecasts for the two short-term forecasting methods as well as the five long-term methods during a 2 year in-sample period. Each method was forecast out 90 days and there were seven different 90 day forecast periods (630 days) in the sample period (August 6, 2000 to April 27, 2002). Therefore, mean error measures are the mean errors across 90 lead times across seven different forecast periods. Conceptually one might think of the median error measure as the error 45 days prior to a target date for a typical forecast period. Since each day is considered a separate data series (due to the use of bookings to date), the MdAPE error measure as recommended by Armstrong & Collopy (1992) is the most appropriate (see Equation 2 and Equation 4 for MdAPE calculation). As can be seen, the improvement of the BC method over the AP method is significant (29.4%).

Table 9: In-sample short-term model comparisons (August 6, 2000 to April 27, 2002)

Model	MSE	MdCumRAE	MdAPE
Models using complete stay information only:			
Random walk (RW)	12,882	1.00	31.2%
Nonlinear regression (NL)	6,381	.69	21.1%
ARIMA	8,605	.84	27.0%
Holt-Winters multiplicative (HW)	24,397	1.41	37.3%
Linear regression (LR)	8,255	.78	29.6%
Models using both complete stay information and bookings to date:			
Additive pickup (AP)	4,781	.57	17.7%
Booking curve (BC)	3,180	.49	12.5%

The error of the seven forecasting methods also changes across lead times, with the two short-term forecasting methods providing clearly superior forecasts closer to the target date (see Figure 11). Short-term forecasts were not compared out of sample due to the time effort required and the minimal managerial benefit given a capacity-slack environment.

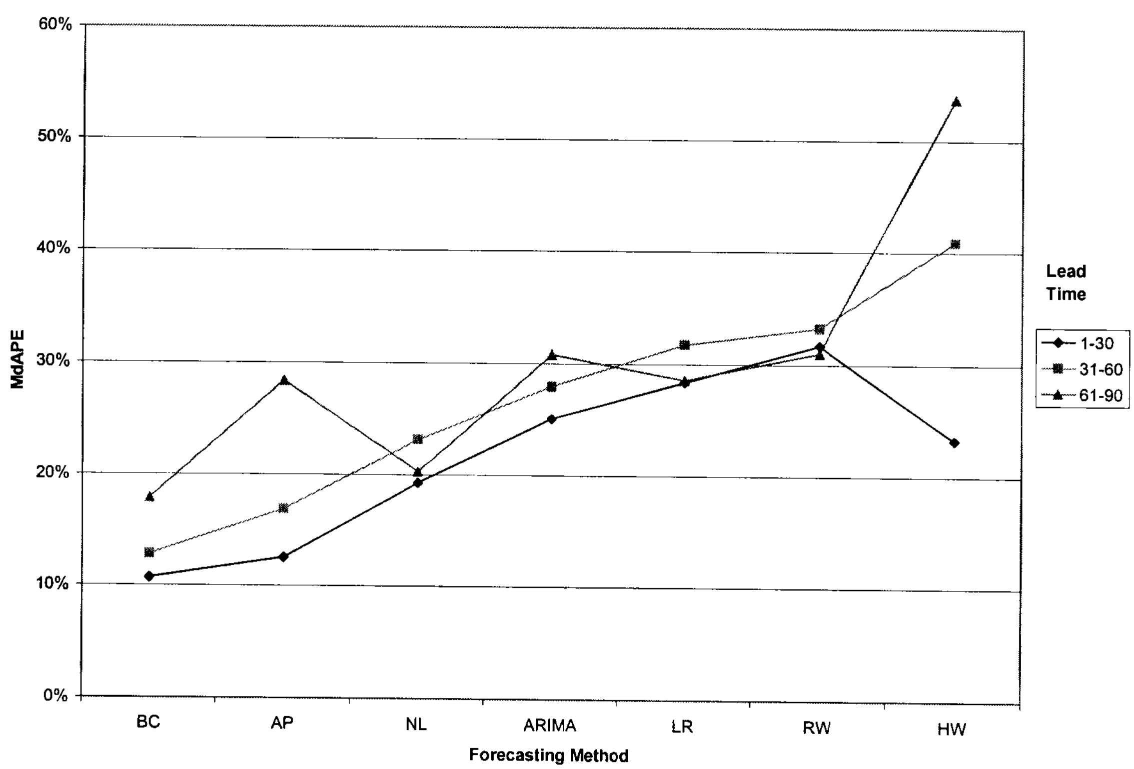

Figure 11: In-sample median absolute percentage error (MdAPE) for forecasting methods across lead times

4.4 Booking Curve Decision Support System

A decision-support system (DSS) based on the BC model (which was calibrated using SAS statistical software) was built in Microsoft Excel so that the resort could calculate short-term estimates in the current season on an ongoing basis. The DSS consists of an input worksheet which is linked to the resort reservation management system that provides the number of room nights booked to date (excluding group and owner bookings) for the next 90 days by bedroom type. These booking figures are automatically used in BC model calculations (with appropriate values based on target date: seasonal period, day of week, year) to forecast short-term demand estimates expressed in the output worksheet (see Figure 12). The output page provides forecasts by bedroom as well as in aggregate, and provides expected pickup between bookings to date and final demand estimates so that managers can scrutinize and monitor actual versus expected pickup. The DSS also has charts of final demand estimates and expected booking curves that are automatically updated from the resort reservation system.

90-DAY SHORT-TERM ESTIMATES

Date of data extract	**09-Dec-02**
1 bedroom capacity	**836**
2 bedroom capacity	**147**
3 bedroom capacity	**27**
Total capacity	1,010

Target Date	**Day of Week (1 = MONDAY)**	**Lead Time (Days)**	**TOTAL UNITS** Long-Term Estimate (before booking data)	Bookings to Date	**Short-Term Estimate**	**Expected Pickup (units)**	Forecast as % of capacity
09-Dec-02	1	0	75	92	**92**	**0**	9%
10-Dec-02	2	1	75	79	**84**	**5**	8%
11-Dec-02	3	2	75	96	**107**	**11**	11%
12-Dec-02	4	3	77	121	**139**	**18**	14%
13-Dec-02	5	4	230	339	**403**	**64**	40%
14-Dec-02	6	5	339	413	**511**	**98**	51%
15-Dec-02	7	6	138	430	**510**	**80**	50%
16-Dec-02	1	7	118	168	**214**	**46**	21%
17-Dec-02	2	8	118	128	**173**	**45**	17%
18-Dec-02	3	9	118	142	**193**	**51**	19%

...

Figure 12: Portion of decision-support system output page (data as at December 9, 2002)

5 DISCUSSION

For a hotel with fixed capacity, Weatherford, Kimes & Scott (2001) found that four forecasting methods for hotel demand (exponential smoothing, moving average, linear regression, and additive pickup) performed equally well. In the case of the resort studied, with increasing yearly capacity, this was certainly not the case. For long-term forecasts, assuming stable yearly trend, the nonlinear regression was slightly superior to the random walk method, and clearly superior to the ARIMA, linear regression, and multiplicative Holt-Winters models. Further, in a situation of a downward capacity shock, as was experienced at the resort on December 1, 2001, random walk was clearly superior to all other long-term models. To generalize to other resort lodging properties, given a predictable yearly trend in demand, a nonlinear regression model is recommended. The performance of an ARIMA model was also quite good in both capacity situations (predictable and unpredictable capacity) while the performance of a linear regression model and multiplicative Holt-Winters model were clearly inadequate in all capacity situations. In terms of short-term demand forecasting, a booking curve model as developed in this paper performed very well in-sample and can only be assumed to be the case in an out of sample setting with predicable capacity.

In terms of managerial implications, this paper has basically given support to the resort management's practice of random walk for long-term forecasts and additive pickup for short-term forecasts. Nonlinear regression long-term models and short-term booking curve models provide marginal improvements in the resort's demand forecasting given a predictable capacity environment or an upward demand shock. However, given the resort has a large amount of capacity slack, more accurate demand forecasts will likely have a small impact on lodging operations. Rather, the resort should revisit these models if capacity becomes strained. If sell outs become more frequent then demand forecasting accuracy becomes much more important. Furthermore, the booking curve model can be adjusted slightly to provide unconstrained demand estimates by arrival date. Unconstrained demand estimates by market segment, length of stay, and arrival date are critical inputs into intelligent revenue management decisions during periods of constrained capacity.

A comparison of the models based on forecast accuracy alone is probably insufficient for a complete evaluation of model effectiveness. Given that the models are used in a business context, the insight that the models may shed on the lodging environment is an important management consideration. The ARIMA models, while providing decent forecasts (especially short-term), are nearly uninterpretable. Even if the ARIMA equations (Equations 18-19) were expressed as a weighting of past observations, the differencing of the data and long time-span required make management insight from these models very unlikely. The RW models, on the other hand, are very straightforward to understand and have provided very good predictions. Unfortunately, other than providing a good estimate of demand in the current period, it is difficult to decipher how much of the RW estimate is due to a systematic seasonal component and how much is due to random flux. The HW model is very good in this regard as it explicitly models the systematic period component and interpretation of these periods is straightforward. For example, one need only to multiply the appropriate week of the year parameter by day of week parameter to see how that day compares to the average day (1.00) or any other day of the year. Unfortunately, the HW model likely simplifies too much as the day of the week effect is not constant throughout the year and grouping the year by chronological week misses important events that span less than one week such as President's day, New Years, and weekend festivals. The simplicity of the HW model, while readily interpretable, is likely responsible for its poor forecasting performance (especially long-term forecasts).

LR, NL, and BC methods explicitly model all seasonal components (period, day of week, yearly trend, interactions between period and day of week) and likely strike the best balance between a straightforward interpretation and a level of sophistication that provides good estimates. All three of these models have statistically tested the significance of seasonal periods and thus provide a reliable base from which management can view periods as being truly distinct. In the development of these models, management claimed to view the lodging season as 13 distinct periods (6 summer periods and 7 winter periods). These 13 periods were used as the starting point for these models but the predictability of FIT demand has allowed further refinement of these original 13 periods into as many as 57 distinct periods in the case of the BC one bedroom model. Further refinement of demand periods should be very helpful for management as it sets rate targets and manages expectations of seasonal demand.

Figure 13A and Figure 13B show the original 13 periods as defined by resort management, and a further refinement of these periods as defined by the NL one bedroom model. Both figures show the average daily demand (averaged by week or period, whichever was smaller) in the 01/02 season. The bold red dashed lines indicate the original 13 periods defined by resort management and the black dashed lines represent sub-periods within the original 13 periods. As evidenced by visual inspection, the additional periods do seem to discriminate truly different demand levels within the original periods. As well, the weekend-period interaction parameters and day of week parameters of the BC, NL, and LR models should further aid management in setting rates by day of week within larger seasonal periods.

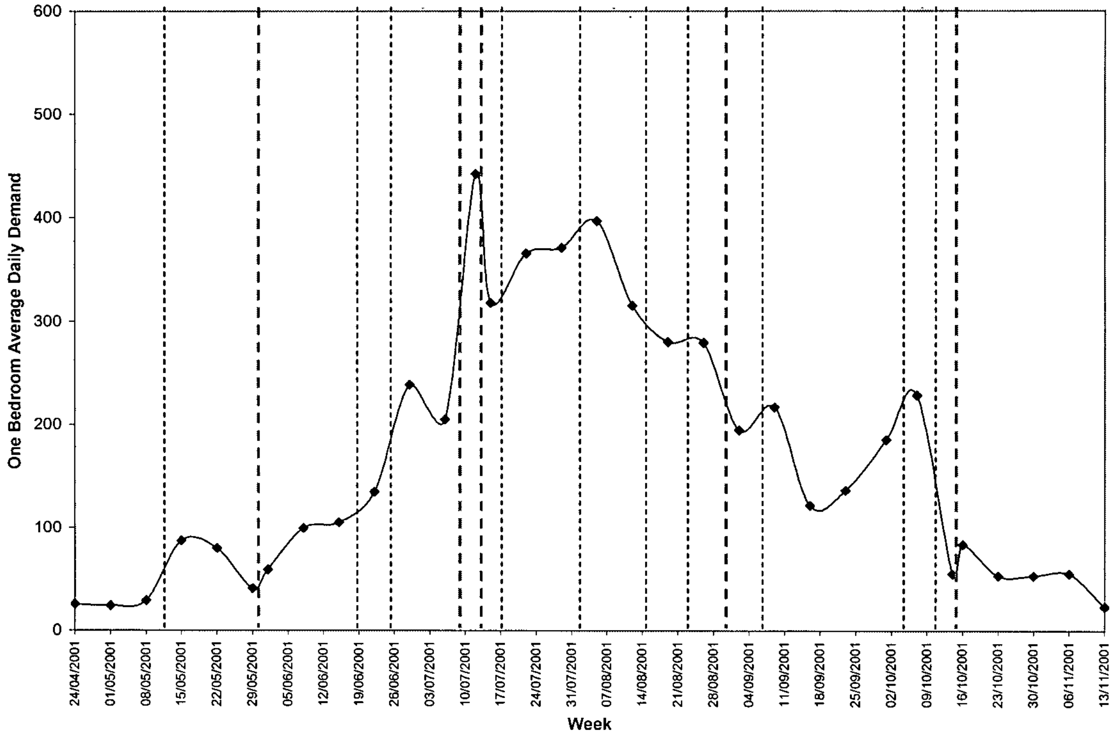

Figure 13A: Average FIT daily demand in 01/02 summer season and corresponding seasonal period classification by week

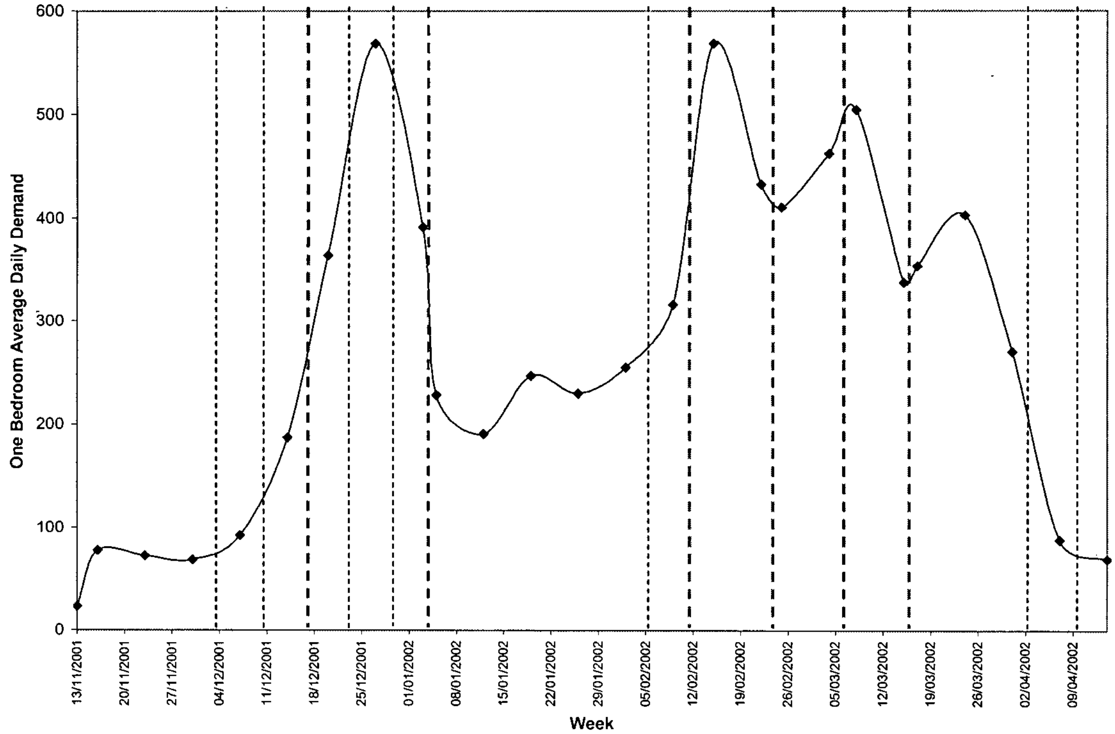

Figure 13B: Average FIT daily demand in 01/02 winter season and corresponding seasonal period classification by week

Beyond seasonal period classification, the BC model provides management with expected booking curves for any target date. This should prove a useful complement to raw pickup numbers taken from AP models. A chart of expected bookings (baseline booking curve) is a compelling visualization of systematic demand build-up. For example, knowing that a certain day of week within a period has consistently shown a large proportion of last-minute bookings should reassure management of its current pricing if room bookings are short of budgeted room nights close to the target date. At the very least, expected booking curves provide another reference point for determining whether last year's pickup (AP model) is representative of historical patterns or whether it may have been an aberration. Figure 14A shows how day of week can have a very large impact on the booking curve as it compares the expected buildup in bookings for a Thursday night and Saturday night within the same week in July 2003. The curves are nearly identical up to about 27 days out from the target date (indicated by a dashed line) at which point the Saturday night is expected to get an acceleration of bookings above and beyond the Thursday night.

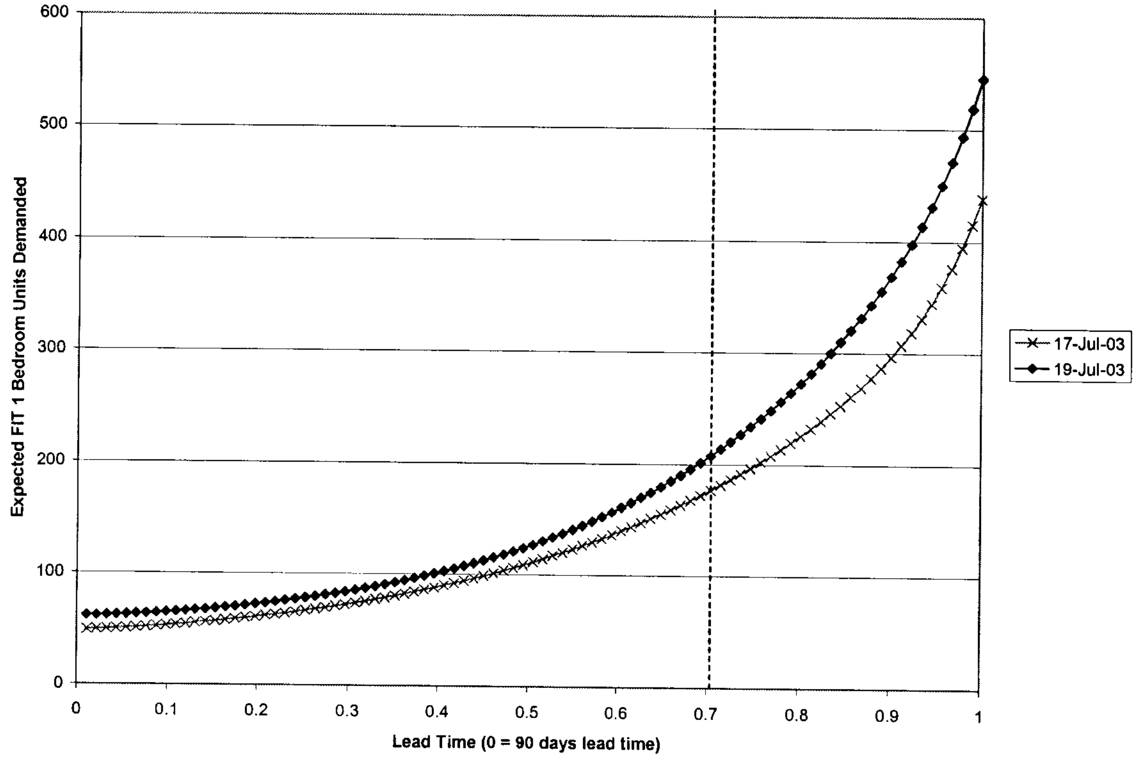

Figure 14A: Expected FIT one bedroom booking curves for a Saturday vs. Thursday in July 2003 (Thursday = July 17, 2003, Saturday = July 19, 2003)

Figure 14B shows an example of how different periods can also result in very different booking curves. The winter season date (March 17, 2003) has an almost linear buildup in bookings while the fall date (September 19, 2003) is expected to get a large proportion of last minute bookings. At 31 days out (marked by a dashed line), there is a difference of 94 rooms booked while the final demand is expected to differ between the two dates by only 6 rooms.

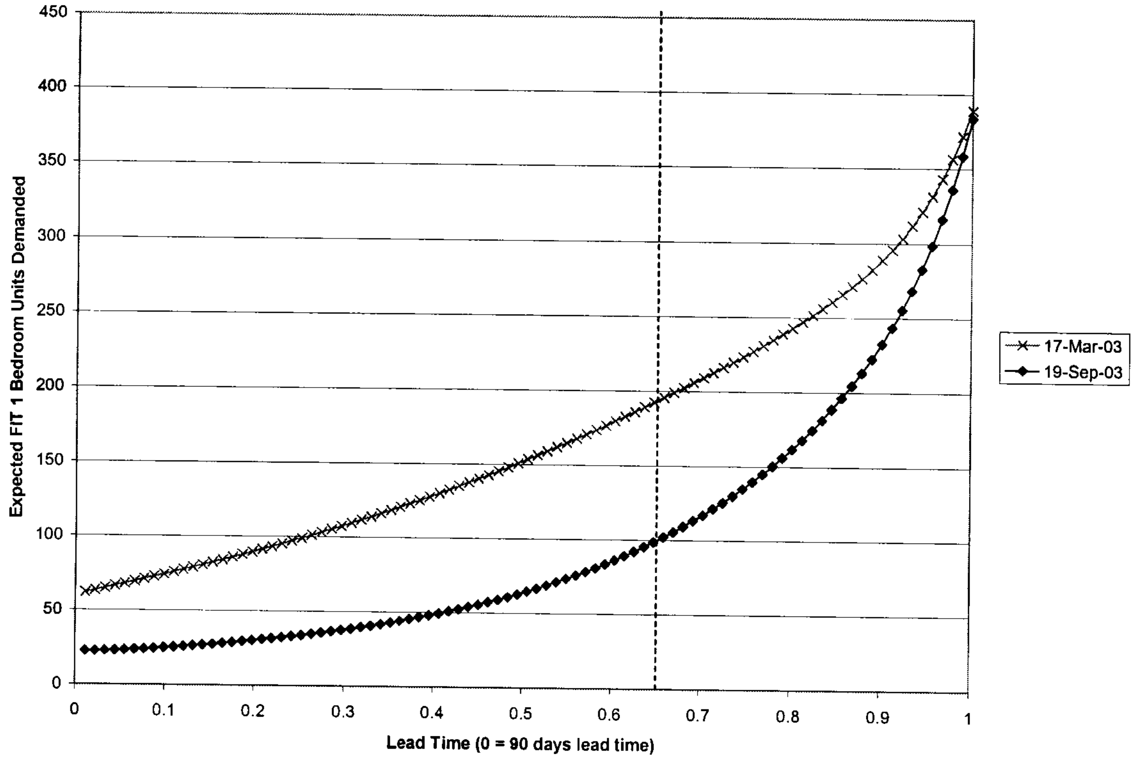

Figure 14B: Expected FIT one bedroom booking curves for a Winter Date vs. Fall Date (Winter Date = March 17, 2003, Fall Date = September 19, 2003)

Mention has thus far been made that identifying distinct seasonal periods may help in setting rates. While the goal of this paper has been to compare methods for estimating demand, the relationship between demand and rates is ultimately the most important issue for revenue management. Figure 15 shows the average one bedroom daily room rate (by week) and average daily demand (by week) with the average daily room rate and average daily demand both normalized to be 1.00. The correlation between average daily rate and average daily demand is .83. As can be seen from the chart, average room rates closely match demand over the winter season while not matching high demand periods in the summer period. Further, the shoulder periods do not see a subsequent decrease in room rates when demand troughs.

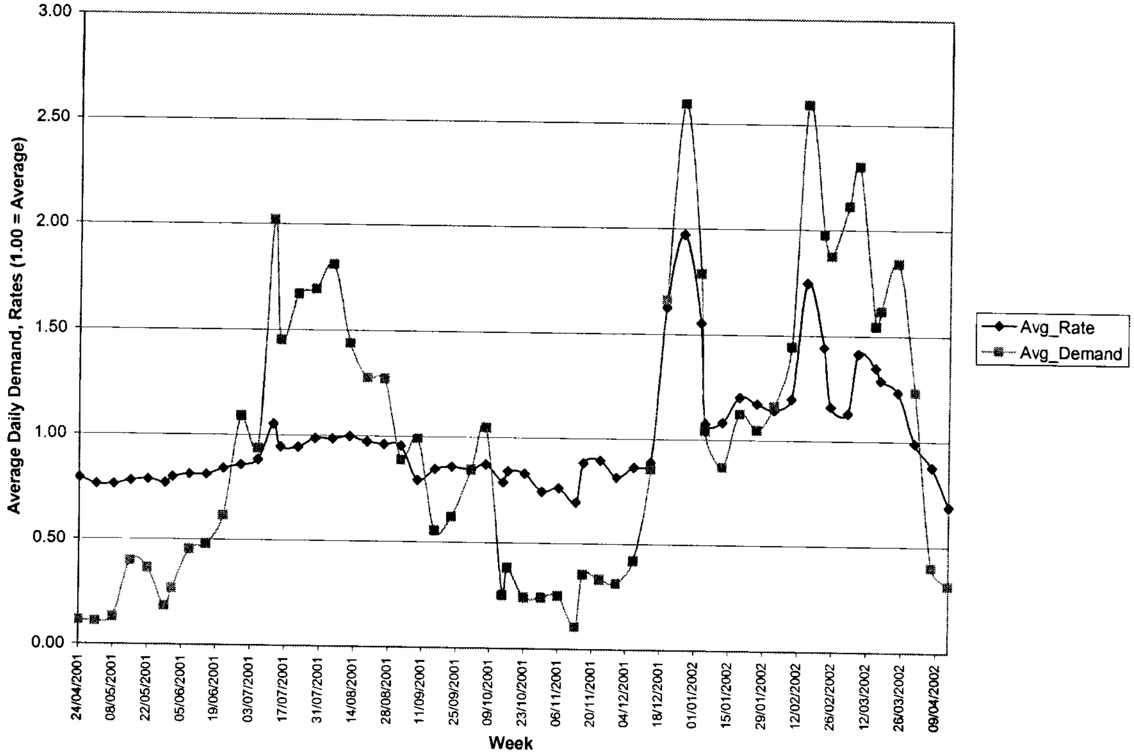

Figure 15: Comparison of one bedroom normalized average daily room rates to normalized average daily demand in the 01/02 season by week (average = 1.00)

Figure 15 does not necessarily imply that rates are inappropriate, as resort managers have stated that the summer demand is comprised of regional guests who are more price sensitive, and as such management has less flexibility to increase room rates when summer demand increases. However, the relationship between demand and room rates should definitely be explored further, and using the seasonal periods defined by the NL and BC models is a good starting point. Resort managers have stated that much of the adjustment of room rates is done on an ongoing basis in conjunction with the resort's call center (which books close to 60% of room bookings). Specifically, the revenue managers monitor conversion rates (calls that end up in bookings) and refusal rates (percent of calls where a specified room is turned down due to price). If conversion rates drop too low (e.g. much below 30%) or refusal rates climb too high (e.g. above 13%) then this is an indicator the current room rates are too high. Conversely, high conversion rates and low refusals may indicate prices are too low. Further analysis of room rates, expected demand, call volume, call classification, and actual demand is beyond the scope of this paper, but appears to be a fruitful area for future analysis. The current season (02/03) is the first season that call center information regarding room demand is being systematically recorded.

5.1 Model Extensions in a Capacity Constrained Environment

In an environment of frequently constrained capacity, additional accuracy in demand forecasts is valuable and worthy of modelling effort. Assuming this capacity constrained environment a number of potential extensions in the short-term booking curve model are proposed. First, the model can be improved by using a larger number of inputs and hence provide more accurate lodging demand estimates. Second, the model can be adapted to integrate more closely with a revenue management system or other optimization engine (although none currently exists at the resort). Third, the model can be extended to include group and owner bookings. Fourth, the model can be extended to include other on-mountain sources of revenue in order to achieve a more global objective of resort revenue maximization.

It is well known (and confirmed by the resort's in-house research) that besides day of week and time of year, weather is the single most important factor in predicting demand at a ski resort. Quite simply, good snow brings crowds. Inputs into the regression models could include snow base (relative to a historical average), projected snowfall and past snowfall (e.g. in a week prior to a target date) for winter seasons. In warmer seasons, while likely to have less of an effect, temperature and rainfall forecasts may also improve lodging demand predictions.

Besides more accurate demand forecasts, forecasts that can be easily integrated with an optimization engine would prove useful. For example, creating complementary models to predict arrival distributions (rather than occupied room nights) and demand by rate class and length of stay would further formalize the revenue management process at the resort. Beginning this year (02/03) the resort is tracking turndown and denial information. This information should prove invaluable in building more disaggregated forecasts and probability distributions that would be classified by room type, market segment, rate class, and length of stay. Only by providing disaggregate estimates in terms of both length of stay, arrivals, and price probabilities can algorithms be developed to optimize revenue.

Demand estimates for independent travelers should be integrated with demand estimates for groups and owners. Owner estimates are important as far as they lower available capacity, while group demand estimates are important in terms of price sensitivity, resort promotion, and analysis of long-term contracts with wholesalers.

In the case of the resort studied, the resort receives revenue from resort operations (ski tickets, rentals, food and beverage, retail) as well as lodging. Therefore, it makes sense to include these sources of ancillary revenue when building revenue maximization models. In other words, since lodging guests will be spending on hill, the objective should be to maximize resort profit rather than lodging profit alone. For example, it may be prudent to lower lodging rates in order to boost lodging occupancy, with the assumption that lost lodging revenue (due to lower lodging prices) would be more than offset by ancillary revenue on mountain. In many situations, the efficacy of rental pool managers as judged by chalet owners is the occupancy rate achieved rather than revenue received (in fact this was the primary factor that led to the loss of units under management in the current season at the resort studied). While revenue received should be the rational economic objective of chalet owners, a focus on occupancy rates may benefit resort management in maximizing resort profit (assuming maximum resort profit comes at the cost of lower lodging profit and higher lodging occupancy). It should be noted that the above hypotheses should be analyzed further, and that other considerations/constraints to resort profit maximization include ski hill capacity, desired clientele / snob appeal, and overall guest experience. As a result, it may not make sense to offer rock bottom lodging rates to attract more skiers to the mountain if it is at odds with the resort's strategy in terms of appropriate target market and atmosphere.

APPENDIX A – DATA PREPARATION AND TRANSFORMATION

Data received from the resort covered the historical period May 15, 1998 to April 29, 2002. The reservation data was received in a raw table format with separate tables for reservations, guest information, and lodging unit information. This reservation data was converted to room night information using SAS statistical software. Essentially, each reservation was classified by market segment (group, independent traveler, owner) and bedroom (one, two, three plus). Once classified, group and owner bookings were excluded. Room night information was then processed using a looping algorithm by counting the number of distinct units to be rented for a specific target date at each of 90 days prior to a target date. A reservation was included in the room night tally for as long as it remained on the books up until the target date. This way, bookings that eventually became cancellations would be included in the sample data for as long as they were on the books. If a booking was cancelled, the reservation was removed from the books upon the date of cancellation.

Consider a reservation for a two-bedroom unit with an arrival date of Feb. 2, 2002, a departure date of Feb. 9, 2002, a reservation date of Jan. 3, 2002, and a cancellation date of Jan. 29, 2002. This reservation is applied to seven different target dates (nights of Feb. 2, 2002 to Feb. 8, 2002) for one two-bedroom unit. Further, the reservation is on the books for 26 days (Jan. 3, 2002 to Jan. 29, 2002) until the cancellation is made. For the target date of Feb. 2, 2002 the booking is included for lead time days 30 to 4 (Feb. 2, 2002 less Jan. 3, 2002 equals lead time day 30; Feb. 2, 2002 less Jan. 29, 2002 equals lead time day 4). The booking for the target date of Feb. 8, 2002 is included for lead time days 36 to 10 (Feb. 8, 2002 less Jan. 3, 2002 equals lead time day 36; Feb. 8, 2002 less Jan. 29, 2002 equals lead time day 10).

APPENDIX B – MULTIPLICATIVE HOLT-WINTERS

Multiplicative Holt-Winters One Bedroom Model

Fit statistics:

Number of observations: 1,446
Sum of squares total: 27,695,441
Mean square error: 8,584
Mean absolute percentage error: 39.3%
Mean absolute error: 54.1
R^2: .57

Degrees of freedom: 1,386
Sum of squares error: 11,989,034
Root mean square error: 92.7
Mean percent error: -18.5%
Mean error: -6.5
Sigma: 92.65

Smoothing parameters:

Alpha (mean-term): .20 Gamma (slope-term): .20 Delta (seasonal-term): .25

Day of week parameters:

Monday: .58	Tuesday: .68	Wednesday: .67	Thursday: .85
Friday: 1.58	Saturday: 1.94	Sunday: .70	

Weekly parameters (approximate beginning date):

18-MAY: .34	25-MAY: .46	01-JUN: .44	08-JUN: .45	15-JUN: .51	22-JUN: .57
29-JUN: .81	06-JUL: 1.12	13-JUL: 1.24	20-JUL: 1.73	27-JUL: 2.09	03-AUG: 2.11
10-AUG: 1.94	17-AUG: 1.74	24-AUG: 1.45	31-AUG: .91	07-SEP: 1.05	14-SEP: .89
21-SEP: .72	28-SEP: .73	05-OCT: .71	12-OCT: .96	19-OCT: .43	26-OCT: .33
02-NOV: .25	09-NOV: .20	16-NOV: .21	23-NOV: .49	30-NOV: .36	07-DEC: .52
14-DEC: .83	21-DEC: 2.02	28-DEC: 2.77	04-JAN: 1.69	11-JAN: .85	18-JAN: .89
25-JAN: .86	01-FEB: 1.02	08-FEB: 1.48	15-FEB: 1.66	22-FEB: 1.87	29-FEB: 1.58
07-MAR: 1.31	14-MAR: 2.12	21-MAR: 1.45	28-MAR: 1.13	04-APR: .90	11-APR: .62
18-APR: .49	25-APR: .28	02-MAY: .19	09-MAY: .20		

Multiplicative Holt-Winters Two/Three Bedroom Model

Fit statistics:

Number of observations: 1,446
Sum of squares total: 819,838
Mean square error: 252
Mean absolute percentage error: 51.4%
Mean absolute error: 9.7
R^2: .57

Degrees of freedom: 1,386
Sum of squares error: 349,034
Root mean square error: 15.9
Mean percent error: -28.6%
Mean error: -1.2
Sigma: 15.87

Smoothing parameters:

Alpha (mean-term): .20 Gamma (slope-term): .20 Delta (seasonal-term): .25

Day of week parameters:

Monday: .77	Tuesday: .78	Wednesday: .68	Thursday: .79
Friday: 1.53	Saturday: 1.77	Sunday: .68	

Weekly parameters (approximate beginning date):

18-MAY: .30	25-MAY: .54	01-JUN: .50	08-JUN: .40	15-JUN: .50	22-JUN: .50
29-JUN: .70	06-JUL: 1.36	13-JUL: 1.34	20-JUL: 1.90	27-JUL: 1.95	03-AUG: 1.85
10-AUG: 1.68	17-AUG: 1.52	24-AUG: 1.54	31-AUG: 1.12	07-SEP: .94	14-SEP: .72
21-SEP: .70	28-SEP: .72	05-OCT: .83	12-OCT: .88	19-OCT: .41	26-OCT: .33
02-NOV: .39	09-NOV: .34	16-NOV: .36	23-NOV: .78	30-NOV: .37	07-DEC: .48
14-DEC: .89	21-DEC: 1.98	28-DEC: 2.44	04-JAN: 1.65	11-JAN: .87	18-JAN: .85
25-JAN: .98	01-FEB: 1.23	08-FEB: 1.50	15-FEB: 1.43	22-FEB: 1.72	29-FEB: 1.64
07-MAR: 1.27	14-MAR: 1.81	21-MAR: 1.60	28-MAR: 1.30	04-APR: .97	11-APR: .76
18-APR: .60	25-APR: .27	02-MAY: .15	09-MAY: .13		

APPENDIX C – ARIMA

ARIMA One Bedroom Model

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
Mean Term					
MU	28.93936	5.06860	5.71	<.0001	0
Moving Average Terms					
MA1,1	0.20889	0.04011	5.21	<.0001	2
MA1,2	-0.11588	0.03713	-3.12	0.0018	4
Autoregressive Terms					
AR1,1	0.87504	0.03003	29.14	<.0001	1
AR1,2	-0.11799	0.02957	-3.99	<.0001	3
Weekly Autoregressive Term					
AR2,1	0.17547	0.03131	5.60	<.0001	7
Yearly Autoregressive Term					
AR3,1	-0.38091	0.03654	-10.42	<.0001	364

Constant Estimate	8.005583
Variance Estimate	2204.724
Std Error Estimate	46.95449
AIC	11407.18
SBC	11442.09
Number of Residuals	1082

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MU	MA1,1	MA1,2	AR1,1	AR1,2	AR2,1	AR3,1
MU	1.000	0.002	-0.002	0.001	-0.004	-0.002	-0.015
MA1,1	0.002	1.000	-0.223	0.655	-0.444	0.052	-0.023
MA1,2	-0.002	-0.223	1.000	-0.144	0.518	-0.145	0.003
AR1,1	0.001	0.655	-0.144	1.000	-0.651	0.092	-0.031
AR1,2	-0.004	-0.444	0.518	-0.651	1.000	-0.261	0.032
AR2,1	-0.002	0.052	-0.145	0.092	-0.261	1.000	-0.037
AR3,1	-0.015	-0.023	0.003	-0.031	0.032	-0.037	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.00	0	<.0001	-0.004	0.001	0.015	0.001	-0.029	0.005
12	6.24	6	0.3965	-0.014	0.032	-0.016	0.015	0.013	0.052
18	14.61	12	0.2632	-0.039	0.067	-0.015	0.018	-0.030	-0.008
24	23.82	18	0.1610	-0.022	0.052	0.064	0.013	0.029	-0.009
30	32.67	24	0.1112	0.006	0.026	-0.039	0.061	-0.026	-0.037
36	39.31	30	0.1190	0.004	0.035	-0.057	-0.006	0.033	0.017
42	42.14	36	0.2224	-0.008	0.030	-0.024	-0.008	0.003	0.031
48	54.12	42	0.0996	0.010	-0.089	-0.035	0.014	-0.024	-0.023

Model for variable FIT_1

Estimated Mean	28.93936
Period(s) of Differencing	364

Autoregressive Factors

Factor 1: 1 - 0.87504 B**(1) + 0.11799 B**(3)
Factor 2: 1 - 0.17547 B**(7)
Factor 3: 1 + 0.38091 B**(364)

Moving Average Factors

Factor 1: 1 - 0.20889 B**(2) + 0.11588 B**(4)

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
Moving Average Term					
MA1,1	0.92053	0.03577	25.73	<.0001	1
Yearly Moving Average Term					
MA2,1	0.46835	0.03620	12.94	<.0001	364
Autoregressive Terms					
AR1,1	1.86845	0.04501	41.51	<.0001	1
AR1,2	-1.04364	0.06061	-17.22	<.0001	2
AR1,3	0.16905	0.03101	5.45	<.0001	3
Weekly Autoregressive Term					
AR2,1	0.09361	0.03248	2.88	0.0040	7

Variance Estimate	62.67246
Std Error Estimate	7.916594
AIC	7553.798
SBC	7583.717
Number of Residuals	1082

* AIC and SBC do not include log determinant.

Correlations of Parameter Estimates

Parameter	MA1,1	MA2,1	AR1,1	AR1,2	AR1,3	AR2,1
MA1,1	1.000	-0.003	0.742	-0.372	-0.231	0.296
MA2,1	-0.003	1.000	0.059	-0.089	0.089	0.010
AR1,1	0.742	0.059	1.000	-0.858	0.319	0.205
AR1,2	-0.372	-0.089	-0.858	1.000	-0.759	-0.107
AR1,3	-0.231	0.089	0.319	-0.759	1.000	-0.064
AR2,1	0.296	0.010	0.205	-0.107	-0.064	1.000

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	0.00	0	<.0001	0.004	-0.021	0.022	0.009	-0.026	0.043
12	11.37	6	0.0776	-0.004	0.004	-0.078	0.012	-0.025	-0.004
18	17.58	12	0.1291	0.023	0.034	0.016	-0.011	-0.053	-0.027
24	23.02	18	0.1897	0.013	0.015	0.045	-0.015	0.046	0.014
30	24.91	24	0.4106	0.003	-0.000	0.036	0.018	-0.001	0.010
36	32.42	30	0.3481	-0.033	-0.007	0.023	-0.033	0.061	0.012
42	34.53	36	0.5387	-0.015	-0.019	-0.011	-0.009	-0.010	0.032
48	44.65	42	0.3612	-0.043	-0.076	0.004	0.009	-0.035	0.010

Model for variable FIT_23

Period(s) of Differencing 364

No mean term in this model.

Autoregressive Factors

Factor 1: 1 - 1.86845 B**(1) + 1.04364 B**(2) - 0.16905 B**(3)

Factor 2: 1 - 0.09361 B**(7)

Moving Average Factors

Factor 1: 1 - 0.92053 B**(1)

Factor 2: 1 - 0.46835 B**(364)

APPENDIX D – LINEAR REGRESSION

Linear Regression One Bedroom Model

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	37	21268760	574831	125.94	<.0001
Error	1408	6426681	4564.40381		
Corrected Total	1445	27695441			

Root MSE	67.56037	R-Square	0.7680
Dependent Mean	174.41286	Adj R-Sq	0.7619
Coeff Var	38.73589		

Parameter Estimates
Dependent Variable: FIT_1

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
General Intercept						
Intercept	Intercept	1	93.20633	6.03273	15.45	<.0001
Yearly Trend Parameter						
SN		1	29.67332	1.61654	18.36	<.0001
Seasonal Period Parameters (Summer)						
S1	S1	1	-182.87644	9.57005	-19.11	<.0001
S1_DG		1	60.75948	11.36755	5.34	<.0001
S2	S2	1	-109.12375	8.96432	-12.17	<.0001
S2_D		1	33.89286	14.74288	2.30	0.0217
S2_EF		1	109.59592	11.71884	9.35	<.0001
S2_G		1	163.57477	31.24706	5.23	<.0001
S4	S4	1	105.84829	9.19576	11.51	<.0001
S4_BC		1	40.20238	11.65527	3.45	0.0006
S4_F		1	-58.54762	14.74288	-3.97	<.0001
S4_G		1	-114.27018	16.88808	-6.77	<.0001
S5_BE		1	-71.35990	7.98835	-8.93	<.0001
S5_G		1	-103.52768	24.34367	-4.25	<.0001
S6	S6	1	-153.78780	7.55306	-20.36	<.0001
Seasonal Period Parameters (Winter)						
W1	W1	1	-121.68495	9.39803	-12.95	<.0001
W1_D		1	39.26190	14.74288	2.66	0.0078
W1_E		1	115.62976	18.65431	6.20	<.0001
W2	W2	1	211.12617	13.56494	15.56	<.0001
W2_B		1	53.75000	18.05627	2.98	0.0030
W2_C		1	-79.89079	27.10031	-2.95	0.0033
W3_FG		1	72.83232	14.15107	5.15	<.0001
W4	W4	1	229.36735	15.08652	15.20	<.0001
W5	W5	1	101.50453	10.75644	9.44	<.0001
W5_C		1	-122.18418	31.83587	-3.84	0.0001
W6	W6	1	240.56038	15.75840	15.27	<.0001
W7_D		1	-81.75047	13.66522	-5.98	<.0001
W7_EF		1	-127.13583	13.79685	-9.21	<.0001
Day of Week Parameter						
SAT	SAT	1	51.88946	6.49948	7.98	<.0001

Linear Regression One Bedroom Model (contd.)

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Weekend Period Interaction Parameters						
S_12_WD		1	53.37471	9.11061	5.86	<.0001
S3_WD		1	178.04065	24.52147	7.26	<.0001
S4_WD		1	59.01425	11.57597	5.10	<.0001
S_56_WD		1	100.60750	9.00418	11.17	<.0001
W1_WD		1	63.54727	13.43712	4.73	<.0001
W3_WD		1	55.18529	11.46217	4.81	<.0001
W4_WD		1	-120.82670	22.43503	-5.39	<.0001
W6_WD		1	-134.58223	22.89227	-5.88	<.0001
W7_WD		1	76.31825	12.37653	6.17	<.0001

Linear Regression Two Bedroom Model

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	36	504306	14009	97.42	<.0001
Error	1409	202603	143.79194		
Corrected Total	1445	706909			

Root MSE	11.99133	R-Square	0.7134
Dependent Mean	27.12379	Adj R-Sq	0.7061
Coeff Var	44.20963		

Parameter Estimates
Dependent Variable: FIT_2

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
General Intercept Parameter						
Intercept	Intercept	1	-7.18376	1.13703	-6.32	<.0001
General Intercept Parameter						
SN		1	3.85587	0.28504	13.53	<.0001
Seasonal Period Parameters (Summer)						
S1_DE		1	4.79338	1.85556	2.58	0.0099
S1_FG		1	6.94904	2.74099	2.54	0.0113
S2	S2	1	7.91945	1.38184	5.73	<.0001
S2_E		1	16.59821	2.53363	6.55	<.0001
S2_F		1	20.40404	2.57096	7.94	<.0001
S2_G		1	24.86609	5.49774	4.52	<.0001
S3	S3	1	41.35331	3.56366	11.60	<.0001
S4	S4	1	41.86588	1.38184	30.30	<.0001
S4_EF		1	-4.86607	1.96254	-2.48	0.0133
S4_G		1	-12.80107	2.91469	-4.39	<.0001
S5	S5	1	11.89670	1.64566	7.23	<.0001
S5_BD		1	-4.64286	1.85030	-2.51	0.0122
S5_G		1	-8.79774	4.43843	-1.98	0.0477
Seasonal Period Parameters (Winter)						
W1_BC		1	4.48195	1.78699	2.51	0.0122
W1_D		1	11.30338	2.40022	4.71	<.0001
W1_E		1	26.81794	3.10680	8.63	<.0001
W2	W2	1	53.52199	1.69482	31.58	<.0001
W3	W3	1	19.18433	1.52886	12.55	<.0001
W3_D		1	5.76190	2.61672	2.20	0.0278
W3_E		1	13.15476	2.61672	5.03	<.0001
W3_FG		1	13.80060	2.73335	5.05	<.0001
W4	W4	1	53.16052	2.40022	22.15	<.0001
W4_B		1	-13.53906	4.43765	-3.05	0.0023
W5	W5	1	42.87725	1.90484	22.51	<.0001
W5_C		1	-20.57188	5.65149	-3.64	0.0003
W6	W6	1	49.20994	2.15325	22.85	<.0001
W7	W7	1	44.03552	1.78699	24.64	<.0001
W7_C		1	-15.62500	2.77545	-5.63	<.0001
W7_D		1	-28.51786	2.77545	-10.28	<.0001
W7_E		1	-35.81509	3.01970	-11.86	<.0001
W7_F		1	-45.71576	5.61428	-8.14	<.0001

Linear Regression Two Bedroom Model (contd.)

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Weekend Period Interaction Parameter						
S5_WD		1	18.50671	2.14253	8.64	<.0001
Day of Week Parameters						
FRI	FRI	1	5.82153	0.97760	5.95	<.0001
SAT	SAT	1	11.50077	0.97770	11.76	<.0001
SUN	SUN	1	2.11260	0.93696	2.25	0.0243

APPENDIX E – POISSON REGRESSION

Poisson Regression Three Bedroom Model

Model Information

Data Set	MONTH.ALL_TR_3
Distribution	Poisson
Link Function	Log
Dependent Variable	FIT_3
Observations Used	1446

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1417	2105.3556	1.4858
Scaled Deviance	1417	2105.3556	1.4858
Pearson Chi-Square	1417	2165.7431	1.5284
Scaled Pearson X2	1417	2165.7431	1.5284
Log Likelihood		1641.2989	

The GENMOD Procedure
Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Seasonal Period Parameters (Summer)							
S1_BC	1	0.2380	0.1331	-0.0230	0.4989	3.19	0.0739
S1_DF	1	0.6223	0.0941	0.4379	0.8067	43.77	<.0001
S2	1	0.7532	0.0754	0.6054	0.9009	99.81	<.0001
S2_B	1	-0.3886	0.1591	-0.7004	-0.0767	5.96	0.0146
S2_EF	1	0.5260	0.0934	0.3429	0.7091	31.71	<.0001
S_34	1	1.4741	0.0367	1.4021	1.5460	1611.04	<.0001
S4_E	1	0.1866	0.0890	0.0122	0.3610	4.40	0.0360
S5	1	0.9180	0.0694	0.7820	1.0540	175.08	<.0001
S5_BD	1	-0.2357	0.0860	-0.4042	-0.0671	7.51	0.0061
S6	1	0.7819	0.0687	0.6473	0.9165	129.60	<.0001
Seasonal Period Parameters (Winter)							
W1	1	0.3747	0.0884	0.2015	0.5479	17.97	<.0001
W2	1	1.5032	0.0626	1.3804	1.6260	575.79	<.0001
W2_C	1	0.3294	0.1547	0.0262	0.6325	4.53	0.0332
W3	1	1.0786	0.0558	0.9692	1.1879	373.82	<.0001
W3_E	1	0.4361	0.0960	0.2479	0.6243	20.63	<.0001
W4	1	1.9394	0.0808	1.7809	2.0978	575.47	<.0001
W5	1	1.5509	0.0629	1.4275	1.6742	607.60	<.0001
W6	1	2.0281	0.0811	1.8692	2.1871	625.23	<.0001
W7	1	1.6372	0.0586	1.5223	1.7521	779.61	<.0001
W7_CF	1	-0.7557	0.0909	-0.9339	-0.5775	69.07	<.0001
Day of Week Parameter							
SAT	1	0.1595	0.0425	0.0762	0.2429	14.07	0.0002

Poisson Regression Three Bedroom Model (contd.)

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Weekend Period Interaction Parameters							
WKD_P1	1	0.4396	0.1247	0.1951	0.6840	12.42	0.0004
WKD_P2	1	0.4280	0.0938	0.2442	0.6117	20.84	<.0001
WKD_P5	1	0.7953	0.0880	0.6229	0.9678	81.74	<.0001
WKD_P6	1	0.5263	0.1112	0.3084	0.7443	22.40	<.0001
WKD_P7	1	0.7438	0.1259	0.4970	0.9906	34.89	<.0001
WKD_P9	1	0.2750	0.0892	0.1002	0.4498	9.51	0.0020
WKD_P10	1	-0.4642	0.1422	-0.7429	-0.1855	10.66	0.0011
WKD_P12	1	-0.5932	0.1442	-0.8759	-0.3105	16.92	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Lagrange Multiplier Statistics

Parameter	Chi-Square	Pr > ChiSq
Intercept	1.7128	0.1906

APPENDIX F – NONLINEAR REGRESSION

Nonlinear Regression One Bedroom Model

Dependent Variable A_TERM
Method: Gauss-Newton
OTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	7
R	5.287E-6
PPC(D7)	0.000042
RPC(D7)	0.000121
Object	2.68E-10
Objective	4971593
Observations Read	1446
Observations Used	1446
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	44	66710946	1516158	427.56	<.0001
Residual	1402	4971593	3546.1		
Uncorrected Total	1446	71682539			
Corrected Total	1445	27695441			

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Yearly Trend Parameters				
TREND1	235.0	9.2855	216.8	253.2
TREND2	75.9307	3.2913	69.4741	82.3872
Seasonal Period Parameters (Summer)				
P1	-3.2618	0.2502	-3.7527	-2.7709
P1_DG	1.1976	0.2663	0.6752	1.7199
P2	-1.8741	0.1180	-2.1056	-1.6427
P2_D	0.4532	0.1711	0.1176	0.7888
P2_EF	1.3201	0.1320	1.0612	1.5789
P2_G	2.0503	0.3356	1.3920	2.7086
P4	0.5275	0.0983	0.3346	0.7204
P4_BC	0.4574	0.1318	0.1987	0.7160
P4_F	-0.6225	0.1358	-0.8889	-0.3560
P4_G	-1.1311	0.1505	-1.4265	-0.8358
P5	-0.7504	0.0979	-0.9426	-0.5583
P5_BE	-0.6508	0.1129	-0.8723	-0.4293
P5_G	-1.0100	0.3068	-1.6119	-0.4081
P6	-2.2529	0.1976	-2.6405	-1.8652
P6_BE	-0.7657	0.2328	-1.2223	-0.3091

Nonlinear Regression One Bedroom Model (contd.)

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Seasonal Period Parameters (Winter)				
P7	-2.1008	0.1465	-2.3883	-1.8134
P7_D	0.5077	0.1852	0.1444	0.8711
P7_E	1.3932	0.1811	1.0379	1.7485
P8	1.9622	0.3228	1.3290	2.5954
P8_B	176.9	.	.	.
P8_C	-1.0090	0.3714	-1.7377	-0.2804
P9	-0.7077	0.0746	-0.8541	-0.5612
P9_E	0.3210	0.1238	0.0782	0.5638
P9_FG	0.9432	0.1334	0.6815	1.2050
P10	2.9136	0.7672	1.4087	4.4185
P11	0.5959	0.1120	0.3762	0.8156
P11_C	-1.3452	0.4011	-2.1320	-0.5585
P12	3.1819	1.0118	1.1970	5.1667
P13	0.1974	0.0996	0.00199	0.3927
P13_C	-0.7718	0.1385	-1.0434	-0.5001
P13_D	-1.4838	0.1623	-1.8021	-1.1654
P13_EF	-2.1677	0.2090	-2.5777	-1.7577
Day of Week Parameters (1=Monday)				
D6	0.6524	0.0736	0.5081	0.7968
D7	0.1638	0.0618	0.0426	0.2849
Weekend Period Interaction Parameters				
WDS1S2	0.8070	0.1131	0.5851	1.0289
WDS3	1.8641	0.5462	0.7927	2.9355
WDS4	0.9408	0.1629	0.6213	1.2603
WDS5S6	1.3709	0.1111	1.1530	1.5888
WDW1	0.8586	0.1573	0.5500	1.1673
WDW3	0.6771	0.1120	0.4575	0.8967
WDW4	-2.1914	0.7458	-3.6544	-0.7284
WDW6	-2.5790	0.9879	-4.5169	-0.6410
WDW7	0.5690	0.1322	0.3096	0.8284

Nonlinear Regression Two Bedroom Model

Dependent Variable FIT_2
Method: Gauss-Newton
NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	7
R	2.647E-6
PPC(D7)	0.000012
RPC(WB13)	0.000044
Object	1.27E-10
Objective	152506.8
Observations Read	1446
Observations Used	1446
Observations Missing	0

The NLIN Procedure
NOTE: An intercept was not specified for this model.

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Approx Pr > F**
Regression	40	1618224	40455.6	372.97	<.0001
Residual	1406	152507	108.5		
Uncorrected Total	1446	1770731			
Corrected Total	1445	706909			

Parameter	**Estimate**	**Approx Std Error**	**Approximate 95% Confidence Limits**	
Yearly Trend Parameters				
TREND1	31.5383	1.2706	29.0458	34.0307
TREND2	10.4943	0.4590	9.5940	11.3946
Seasonal Period Parameters (Summer)				
P1_DE	1.0387	0.2442	0.5598	1.5177
P1_FG	1.2220	0.3468	0.5418	1.9023
P2	1.3089	0.2102	0.8966	1.7212
P2_E	1.4409	0.1780	1.0918	1.7901
P2_F	1.6528	0.1840	1.2919	2.0137
P2_G	2.4922	0.4711	1.5680	3.4164
P3	3.7753	0.4029	2.9850	4.5655
P4	4.1942	0.2335	3.7362	4.6521
P4_EF	-0.5138	0.1635	-0.8345	-0.1931
P4_G	-1.1505	0.1959	-1.5349	-0.7662
P5	1.8613	0.2241	1.4217	2.3009
P5_BD	-0.4661	0.1516	-0.7635	-0.1687
P5_G	-0.7547	0.4000	-1.5394	0.0301
P6_CD	-0.3947	0.3349	-1.0517	0.2623

Nonlinear Regression Two Bedroom Model (contd.)

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Seasonal Period Parameters (Winter)				
P7_BC	0.7919	0.2568	0.2882	1.2957
P7_D	1.4522	0.2632	0.9358	1.9686
P7_E	2.6357	0.2626	2.1206	3.1508
P8	6.4129	1.2468	3.9671	8.8588
P8_B	812.0	.	.	.
P9	2.2680	0.2078	1.8604	2.6757
P9_D	0.5027	0.1679	0.1734	0.8320
P9_E	0.9797	0.1751	0.6362	1.3233
P9_FG	1.1693	0.1834	0.8096	1.5291
P10	61.7079	0.5122	60.7031	62.7128
P10_B	-57.7100	.	.	.
P11	4.6024	0.2989	4.0161	5.1887
P11_C	-1.7863	0.4989	-2.7649	-0.8077
P12	368.8	.	.	.
P13	4.9684	0.3442	4.2932	5.6435
P13_C	-1.8392	0.3021	-2.4317	-1.2466
P13_D	-2.7805	0.3142	-3.3968	-2.1642
P13_E	-3.4934	0.3516	-4.1832	-2.8037
P13_F	-5.0951	1.6321	-8.2969	-1.8934
Day of Week Parameters (1=Monday)				
D1	-3.0880	0.1958	-3.4722	-2.7038
D4	0.2193	0.0912	0.0404	0.3982
D5	1.0322	0.1034	0.8294	1.2350
D6	1.6507	0.1115	1.4320	1.8694
D7	0.3696	0.0891	0.1949	0.5443
Weekend Period Interaction Parameters				
WB5	0.7393	0.1716	0.4026	1.0760
WB6	0.6485	0.2787	0.1019	1.1951
WB13	-0.5984	0.2116	-1.0135	-0.1833

APPENDIX G – BASELINE REGRESSION

Baseline Regression (nonlinear) One Bedroom Model

Dependent Variable FIT_1
Method: Gauss-Newton
NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	20
Subiterations	2
Average Subiterations	0.1
R	5.78E-6
PPC(WC5)	0.000467
RPC(WC7)	0.049505
Object	9.346E-9
Objective	1.5702E8
Observations Read	131586
Observations Used	131586
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	DF	Sum of Squares	Mean Square	F Value	Approx Pr > F
Regression	112	1.4115E9	12602761	10552.5	<.0001
Residual	131474	1.5702E8	1194.3		
Uncorrected Total	131586	1.5685E9			
Corrected Total	131585	9.8471E8			

Parameter	Approx Estimate	Std Error	Approximate 95% Confidence Limits	
Yearly Trend Parameters				
TREND1	299.6	4.6484	290.5	308.7
TREND2	118.6	1.8370	115.0	122.2
Seasonal Period Parameters (Summer)				
P1_B	0.4745	0.1527	0.1752	0.7738
P1_C	0.3188	0.1699	-0.0143	0.6519
P1_D	1.4219	0.1129	1.2006	1.6433
P1_E	1.8412	0.1077	1.6302	2.0523
P1_F	1.4939	0.1185	1.2616	1.7262
P1_G	2.1718	0.1583	1.8616	2.4821
P2	-0.0652	0.1552	-0.3693	0.2390
P2_B	0.1250	0.0347	0.0570	0.1930
P2_C	0.1460	0.0344	0.0785	0.2135
P2_D	0.3472	0.0323	0.2838	0.4105
P2_E	1.0071	0.0287	0.9509	1.0632
P2_F	1.1594	0.0284	1.1038	1.2151
P2_G	1.4427	0.0414	1.3615	1.5239
P4	2.1906	0.1058	1.9833	2.3979
P4_B	0.1718	0.0100	0.1521	0.1915
P4_C	0.0943	0.0102	0.0742	0.1144
P4_E	-0.0805	0.0109	-0.1017	-0.0592
P4_F	-0.3306	0.0121	-0.3543	-0.3069
P4_G	-0.7391	0.0166	-0.7717	-0.7065

Baseline Regression (nonlinear) One Bedroom Model (contd.)

Parameter	Approx Estimate	Approximate Std Error	95% Confidence	Limits
Seasonal Period Parameters (Summer) contd.				
P5	0.9046	0.1368	0.6364	1.1729
P5_B	-0.4706	0.0198	-0.5094	-0.4319
P5_C	-0.4202	0.0193	-0.4580	-0.3825
P5_D	-0.3880	0.0190	-0.4251	-0.3508
P5_E	-0.1042	0.0166	-0.1368	-0.0716
P5_G	-0.7148	0.0412	-0.7956	-0.6340
P6	-0.5575	0.4620	-1.4630	0.3479
P6_B	-0.5323	0.0549	-0.6399	-0.4247
P6_C	-0.9356	0.0710	-1.0749	-0.7964
P6_D	-0.9741	0.0730	-1.1172	-0.8310
P6_E	-0.9276	0.1244	-1.1714	-0.6838
Seasonal Period Parameters (Winter)				
P7_B	0.1739	0.0373	0.1007	0.2470
P7_C	-0.0360	0.0405	-0.1154	0.0435
P7_D	0.6539	0.0329	0.5895	0.7183
P7_E	1.4913	0.0308	1.4310	1.5516
P8	2.6874	0.1124	2.4671	2.9077
P8_B	0.4065	0.00919	0.3884	0.4245
P8_C	-0.7174	0.0157	-0.7482	-0.6866
P9	2.1928	0.1084	1.9804	2.4053
P9_B	-0.0492	0.0133	-0.0753	-0.0231
P9_C	-0.3038	0.0146	-0.3325	-0.2751
P9_D	-0.2197	0.0141	-0.2474	-0.1920
P9_E	0.1691	0.0125	0.1446	0.1936
P9_F	0.6026	0.0121	0.5789	0.6263
P9_G	0.9630	0.0275	0.9091	1.0169
P10	4.9812	0.1052	4.7749	5.1875
P10_B	-0.8181	0.0151	-0.8476	-0.7886
P11	2.7607	0.1087	2.5477	2.9738
P11_B	0.0435	0.0118	0.0203	0.0667
P11_C	-0.7283	0.0386	-0.8039	-0.6527
P12	3.9043	0.1057	3.6971	4.1116
P12_B	0.0700	0.0201	0.0306	0.1093
P13	2.5969	0.1107	2.3799	2.8140
P13_B	-0.1867	0.0105	-0.2074	-0.1661
P13_C	-0.7043	0.0131	-0.7300	-0.6786
P13_D	-1.5075	0.0209	-1.5485	-1.4665
P13_E	-2.1900	0.0368	-2.2621	-2.1180
P13_F	-3.3720	0.2521	-3.8661	-2.8778
Day of Week Parameters (1 = Monday)				
D1	-5.2222	0.1042	-5.4264	-5.0180
D4	0.0493	0.00458	0.0403	0.0583
D5	0.1632	0.0154	0.1331	0.1934
D6	0.2966	0.0154	0.2665	0.3268
D7	0.0176	0.00465	0.00850	0.0267
Seasonal Period Lead Time Interaction Parameters				
B	0.3563	0.0126	0.3316	0.3810
B2	2.7764	0.1108	2.5593	2.9936
B3	4.0208	0.1074	3.8102	4.2313
B4	2.3603	0.0241	2.3130	2.4075
B5	2.4883	0.0922	2.3076	2.6690
B6	2.5387	0.4687	1.6200	3.4573
B7	2.3528	0.1037	2.1495	2.5562
B8	2.8170	0.0422	2.7343	2.8997
B9	1.6793	0.0324	1.6158	1.7427
B10	0.8467	0.0232	0.8012	0.8922

Baseline Regression (nonlinear) One Bedroom Model (contd.)

Parameter	Approx Estimate	Std Error	Approximate 95% Confidence	Limits
Seasonal Period Lead Time Interaction Parameters (contd.)				
B11	1.9923	0.0329	1.9279	2.0567
B12	1.6666	0.0277	1.6123	1.7208
B13	2.0104	0.0403	1.9314	2.0895
Weekend Period Lead Time Interaction Parameters				
WB	0.5295	0.0175	0.4952	0.5638
WB3	0.7336	0.0289	0.6769	0.7902
WB4	-0.3546	0.0157	-0.3854	-0.3237
WB5	0.7073	0.0172	0.6737	0.7409
WB6	1.0736	0.0779	0.9209	1.2263
WB7	0.2568	0.0237	0.2105	0.3032
WB8	-0.4050	0.0170	-0.4383	-0.3717
WB10	-1.1678	0.0186	-1.2043	-1.1313
WB11	-0.3596	0.0187	-0.3963	-0.3230
WB12	-1.4620	0.0234	-1.5078	-1.4162
WB13	-0.2831	0.0157	-0.3139	-0.2523
Exponent Parameters for Seasonal Period Lead Time Interactions (e.g. $B2*T^{C2}$ where T=lead time between 0 and 1)				
C	16.0914	0.8948	14.3376	17.8452
C2	0.8596	0.0560	0.7498	0.9693
C3	0.3351	0.0153	0.3051	0.3650
C4	1.4256	0.0308	1.3653	1.4860
C5	1.2074	0.0642	1.0816	1.3332
C6	1.5767	0.3297	0.9305	2.2228
C7	0.7442	0.0524	0.6416	0.8468
C8	0.6351	0.0199	0.5960	0.6742
C9	0.8465	0.0342	0.7794	0.9135
C10	2.2505	0.1032	2.0483	2.4528
C11	1.2605	0.0459	1.1706	1.3504
C12	1.3545	0.0454	1.2655	1.4435
C13	0.9487	0.0368	0.8766	1.0209
Exponent Parameters for Weekend Seasonal Period Lead Time Interactions (e.g. $WB3*WEEKEND*T^{C3}$ where T=lead time between 0 and 1, WEEKEND=1 if a Friday or Saturday night)				
WC	2.7774	0.1658	2.4525	3.1023
WC3	-0.2562	0.0167	-0.2890	-0.2234
WC4	1.0027	0.1612	0.6868	1.3186
WC5	-0.0232	0.0468	-0.1149	0.0685
WC6	0.1010	0.2468	-0.3829	0.5848
WC7	0.1942	0.2051	-0.2078	0.5963
WC8	0.3478	0.0695	0.2116	0.4841
WC10	-0.1091	0.0103	-0.1292	-0.0889
WC11	-0.1494	0.0583	-0.2637	-0.0351
WC12	-0.0289	0.0143	-0.0568	-0.00094
WC13	0.1968	0.1010	-0.00116	0.3949

Baseline Regression (nonlinear) Two Bedroom Model

Dependent Variable FIT_2
Method: Gauss-Newton
NOTE: Convergence criterion met.

Estimation Summary

Method	Gauss-Newton
Iterations	45
R	8.307E-6
PPC(C_1011)	0.000196
RPC(C_1011)	0.000254
Object	1.88E-10
Objective	7532086
Observations Read	131586
Observations Used	131586
Observations Missing	0

NOTE: An intercept was not specified for this model.

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Approx Pr > F**
Regression	85	69107115	813025	14194.4	<.0001
Residual	131501	7532086	57.2778		
Uncorrected Total	131586	76639201			
Corrected Total	131585	41937547			

Parameter	**Estimate**	**Approx Std Error**	**Approximate 95% Confidence Limits**	
Yearly Trend Parameters				
TREND1	35.7075	0.3258	35.0689	36.3461
TREND2	14.4171	0.1286	14.1650	14.6691
Seasonal Period Parameters (Summer)				
P1_DE	1.9699	0.1060	1.7621	2.1776
P1_FG	2.2014	0.1100	1.9858	2.4170
P2	1.4666	0.1485	1.1756	1.7577
P2_B	-0.3220	0.0322	-0.3851	-0.2589
P2_D	-0.1274	0.0290	-0.1843	-0.0705
P2_E	0.9098	0.0208	0.8691	0.9506
P2_F	1.0240	0.0207	0.9835	1.0646
P2_G	1.6122	0.0375	1.5387	1.6856
P3	3.8499	0.1075	3.6392	4.0605
P4	3.6717	0.1059	3.4642	3.8793
P4_B	0.0416	0.0121	0.0178	0.0653
P4_C	0.1129	0.0120	0.0894	0.1364
P4_E	-0.2848	0.0130	-0.3103	-0.2593
P4_F	-0.1618	0.0126	-0.1865	-0.1371
P4_G	-0.5726	0.0155	-0.6031	-0.5422
P5	1.9732	0.1230	1.7321	2.2143
P5_B	-0.3236	0.0231	-0.3689	-0.2782
P5_C	-0.5789	0.0261	-0.6301	-0.5277
P5_D	-0.3272	0.0232	-0.3727	-0.2818
P5_G	-1.0436	0.0640	-1.1690	-0.9183
P6	-1.1979	0.2356	-1.6596	-0.7362
P6_C	-0.5374	0.0928	-0.7193	-0.3555
P6_D	-0.7844	0.1114	-1.0029	-0.5660

Baseline Regression (nonlinear) Two Bedroom Model (contd.)

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Seasonal Period Parameters (Winter)				
P7_B	0.2209	0.0415	0.1396	0.3022
P7_C	-0.1911	0.0485	-0.2862	-0.0960
P7_D	0.4585	0.0390	0.3821	0.5348
P7_E	1.7226	0.0351	1.6539	1.7914
P8	5.2560	0.1057	5.0488	5.4632
P8_B	0.4398	0.0154	0.4098	0.4699
P8_C	-0.5051	0.0196	-0.5434	-0.4668
P9	3.4555	0.1041	3.2515	3.6594
P9_C	-0.1193	0.0149	-0.1485	-0.0900
P9_D	0.1611	0.0137	0.1342	0.1880
P9_E	0.6330	0.0127	0.6080	0.6579
P9_FG	0.8219	0.0130	0.7965	0.8474
P10	6.3793	0.1105	6.1627	6.5958
P10_B	-0.4505	0.0209	-0.4915	-0.4094
P11	4.6634	0.1049	4.4578	4.8691
P11_B	0.0391	0.0143	0.0111	0.0670
P11_C	-0.7306	0.0367	-0.8025	-0.6588
P12	5.8774	0.1075	5.6667	6.0882
P13	4.4725	0.3508	3.7850	5.1600
P13_B	-0.3237	0.0126	-0.3485	-0.2990
P13_C	-1.1948	0.0159	-1.2261	-1.1636
P13_D	-2.0204	0.0210	-2.0615	-1.9793
P13_E	-2.9338	0.0346	-3.0015	-2.8660
P13_F	-4.2937	0.2173	-4.7195	-3.8679
Day of Week Parameters (1 = Monday)				
D1	-5.5321	0.1047	-5.7373	-5.3269
D4	0.1060	0.00608	0.0940	0.1179
D5	0.5753	0.0101	0.5556	0.5950
D6	0.6332	0.0100	0.6136	0.6529
D7	0.0605	0.00614	0.0485	0.0726
Period Lead Time Interactions				
B	1.1450	0.0173	1.1111	1.1790
B2	1.2430	0.1010	1.0451	1.4409
B4	0.9535	0.0213	0.9118	0.9951
B5	0.8846	0.0714	0.7446	1.0246
B6	2.2000	0.2192	1.7703	2.6297
B7	2.0859	0.1086	1.8730	2.2989
B8	0.4572	0.0370	0.3846	0.5298
B_1011	0.3312	0.0279	0.2766	0.3858
B12	0.2207	0.0374	0.1474	0.2941
B13	0.8927	0.3290	0.2479	1.5374
Weekend Period Lead Time Interactions				
WB	0.3940	0.0159	0.3628	0.4252
WB3	0.5351	0.0331	0.4703	0.5999
WB4	-0.5179	0.0117	-0.5408	-0.4949
WB5	0.8392	0.0270	0.7863	0.8921
WB6	1.0675	0.1054	0.8609	1.2741
WB7	0.1479	0.0288	0.0915	0.2043
WB8	-0.9414	0.0260	-0.9922	-0.8905
WB10	-2.3222	0.0348	-2.3903	-2.2541
WB11	-0.4434	0.0193	-0.4814	-0.4055
WB12	-1.9353	0.0274	-1.9890	-1.8817
WB13	-0.7828	0.0138	-0.8097	-0.7558

Baseline Regression (nonlinear) Two Bedroom Model (contd.)

Parameter	Estimate	Approx Std Error	Approximate 95% Confidence Limits	
Exponent Parameters for Seasonal Period Lead Time Interactions (e.g. $B2*T^{C2}$ where T=lead time between 0 and 1)				
C	0.9418	0.0258	0.8912	0.9923
C2	0.6642	0.0926	0.4826	0.8458
C4	1.4767	0.0718	1.3360	1.6174
C5	1.4063	0.1725	1.0681	1.7444
C7	0.2152	0.0197	0.1767	0.2538
C_1011	4.1863	0.6267	2.9580	5.4145
C13	0.1883	0.0928	0.00636	0.3703
Exponent Parameters for Weekend Seasonal Period Lead Time Interactions (e.g. $WB5*WEEKEND*T^{WC5}$ where T=lead time between 0 and 1, WEEKEND=1 if a Friday or Saturday night, 0 otherwise)				
WC	6.2312	0.4447	5.3596	7.1028
WC5	0.3805	0.0744	0.2347	0.5263
WC8	0.1220	0.0207	0.0815	0.1625

Baseline Regression (Poisson) Three Bedroom Model

Model Information

Data Set	MONTH.ALL_TR_3
Distribution	Poisson
Link Function	Log
Dependent Variable	FIT_3
Observations Used	131586

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	13E4	154544.8740	1.1752
Scaled Deviance	13E4	154544.8740	1.1752
Pearson Chi-Square	13E4	147757.0998	1.1236
Scaled Pearson X2	13E4	147757.0998	1.1236
Log Likelihood		14550.0430	

The GENMOD Procedure
Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
General Intercept Parameter							
Intercept	1	-1.2267	0.0309	-1.2873	-1.1662	1577.68	<.0001
Seasonal Period Parameters (Summer)							
S1_B	1	0.2301	0.0350	0.1616	0.2987	43.27	<.0001
S1_C	1	0.6894	0.0317	0.6273	0.7516	472.35	<.0001
S1_D	1	0.9263	0.0298	0.8679	0.9846	968.54	<.0001
S1_E	1	0.6613	0.0302	0.6022	0.7205	480.27	<.0001
S1_F	1	0.7138	0.0324	0.6502	0.7774	484.16	<.0001
S2	1	0.6046	0.0373	0.5315	0.6776	263.08	<.0001
S2_B	1	-1.2675	0.0330	-1.3322	-1.2028	1473.06	<.0001
S2_CD	1	-0.3522	0.0197	-0.3909	-0.3136	318.44	<.0001
S2_E	1	0.3997	0.0195	0.3615	0.4379	420.80	<.0001
S2_F	1	0.6024	0.0188	0.5656	0.6393	1024.93	<.0001
S3	1	1.7301	0.0540	1.6243	1.8359	1026.51	<.0001
S4	1	1.4885	0.0331	1.4237	1.5534	2025.74	<.0001
S4_B	1	0.2730	0.0132	0.2472	0.2989	428.48	<.0001
S4_D	1	0.0595	0.0142	0.0316	0.0873	17.52	<.0001
S4_E	1	0.2020	0.0135	0.1756	0.2285	223.57	<.0001
S4_F	1	0.0926	0.0140	0.0651	0.1201	43.50	<.0001
S5	1	1.3419	0.0340	1.2753	1.4085	1558.97	<.0001
S5_B	1	-0.3028	0.0168	-0.3358	-0.2698	324.18	<.0001
S5_C	1	-0.2754	0.0166	-0.3080	-0.2428	273.77	<.0001
S5_D	1	-0.4254	0.0176	-0.4599	-0.3908	582.41	<.0001
S5_E	1	-0.1264	0.0158	-0.1573	-0.0955	64.35	<.0001
S6	1	1.0702	0.0352	1.0012	1.1392	924.57	<.0001
S6_C	1	-0.1081	0.0178	-0.1430	-0.0732	36.86	<.0001
Seasonal Period Parameters (Winter)							
W1	1	0.3386	0.0404	0.2595	0.4178	70.25	<.0001
W1_B	1	-0.1563	0.0255	-0.2063	-0.1063	37.52	<.0001
W1_C	1	-0.2723	0.0264	-0.3239	-0.2206	106.77	<.0001
W1_D	1	0.0777	0.0240	0.0305	0.1248	10.43	0.0012
W1_E	1	0.3322	0.0253	0.2826	0.3818	172.49	<.0001
W2	1	2.3765	0.0337	2.3105	2.4425	4979.85	<.0001
W2_C	1	0.1324	0.0187	0.0957	0.1691	50.04	<.0001
W3	1	1.6435	0.0347	1.5755	1.7115	2245.51	<.0001
W3_B	1	-0.3153	0.0186	-0.3518	-0.2787	286.07	<.0001
W3_C	1	-0.2188	0.0181	-0.2544	-0.1833	145.56	<.0001
W3_D	1	-0.1544	0.0178	-0.1893	-0.1194	75.03	<.0001

W3_E	1	0.3329	0.0159	0.3018	0.3640	440.38	<.0001
W3_F	1	0.0382	0.0179	0.0031	0.0734	4.55	0.0329
W3_G	1	-0.3396	0.0564	-0.4501	-0.2291	36.29	<.0001
W4	1	2.8293	0.0349	2.7610	2.8976	6590.30	<.0001
W4_B	1	-0.1838	0.0212	-0.2254	-0.1423	75.28	<.0001
W5	1	2.2089	0.0357	2.1391	2.2788	3836.97	<.0001
W5_B	1	0.0583	0.0159	0.0271	0.0896	13.38	0.0003
W5_C	1	-0.3978	0.0323	-0.4612	-0.3345	151.44	<.0001
W6	1	2.9556	0.0347	2.8875	3.0237	7235.79	<.0001
W6_B	1	-0.0808	0.0275	-0.1347	-0.0269	8.64	0.0033
W7	1	2.5086	0.0334	2.4431	2.5741	5628.91	<.0001
W7_B	1	-0.1997	0.0132	-0.2256	-0.1737	227.40	<.0001
W7_C	1	-0.4802	0.0144	-0.5083	-0.4520	1116.46	<.0001
W7_D	1	-1.5010	0.0208	-1.5418	-1.4602	5202.78	<.0001
W7_E	1	-1.3569	0.0216	-1.3993	-1.3145	3932.15	<.0001
W7_F	1	-1.8073	0.0524	-1.9100	-1.7046	1189.47	<.0001
Day of Week Parameters							
TUE	1	0.0254	0.0067	0.0123	0.0385	14.45	0.0001
WED	1	0.0499	0.0066	0.0369	0.0629	56.33	<.0001
THR	1	0.0996	0.0066	0.0868	0.1125	230.23	<.0001
SAT	1	0.0635	0.0059	0.0518	0.0751	114.60	<.0001
SUN	1	0.1013	0.0066	0.0885	0.1142	238.64	<.0001
Lead Time Element Parameters							
T2	1	1.1060	0.0299	1.0475	1.1645	1372.37	<.0001
T2_P2	1	0.4936	0.0380	0.4192	0.5680	169.12	<.0001
T2_P3	1	-0.1624	0.0633	-0.2866	-0.0383	6.58	0.0103
T2_P4	1	-0.0914	0.0340	-0.1580	-0.0249	7.25	0.0071
T2_P5	1	-0.1722	0.0355	-0.2419	-0.1025	23.47	<.0001
T2_P6	1	-0.1601	0.0390	-0.2365	-0.0838	16.89	<.0001
T2_P7	1	0.4128	0.0420	0.3305	0.4951	96.65	<.0001
T2_P8	1	-0.7417	0.0374	-0.8150	-0.6684	393.03	<.0001
T2_P9	1	-0.2746	0.0348	-0.3428	-0.2063	62.19	<.0001
T2_P10	1	-0.7608	0.0396	-0.8385	-0.6832	368.68	<.0001
T2_P11	1	-0.6391	0.0394	-0.7164	-0.5619	263.22	<.0001
T2_P12	1	-0.8468	0.0396	-0.9245	-0.7691	456.21	<.0001
T2_P13	1	-0.5886	0.0350	-0.6571	-0.5200	283.20	<.0001
Weekend Seasonal Period Parameters							
WKD_P1	1	0.3230	0.0183	0.2872	0.3588	312.86	<.0001
WKD_P2	1	0.3852	0.0141	0.3575	0.4128	746.39	<.0001
WKD_P3	1	0.3355	0.0359	0.2651	0.4060	87.20	<.0001
WKD_P4	1	0.2380	0.0109	0.2166	0.2593	478.09	<.0001
WKD_P5	1	0.6723	0.0120	0.6488	0.6958	3140.29	<.0001
WKD_P6	1	0.4085	0.0156	0.3779	0.4391	682.80	<.0001
WKD_P7	1	0.6866	0.0171	0.6531	0.7201	1617.62	<.0001
WKD_P9	1	0.2296	0.0120	0.2061	0.2531	366.83	<.0001
WKD_P10	1	-0.4249	0.0188	-0.4619	-0.3880	508.50	<.0001
WKD_P11	1	0.3183	0.0186	0.2819	0.3546	294.12	<.0001
WKD_P12	1	-0.5634	0.0216	-0.6057	-0.5210	680.90	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

APPENDIX H – APPROACHES TO BOOKING CURVE ADJUSTMENT

Table 10: Booking curve estimate calculation approaches

Method	**Booking Curve Adjustment Calculation**	**Reduction in Squared Error from Baseline**
Direct Multiplicative	$[D_{LT=Y}+1]EB_{LT=0}$	44.98%
Mean Absolute Percentage Error (MAPE)	$\left[D_{LT=Y}\left(\frac{MAPE_{LT=0}}{MAPE_{LT=Y}}\right)+1\right]EB_{LT=0}$	44.96%
Geometric Mean Absolute Percentage Error (GMAPE)	$\left[D_{LT=Y}\left(\frac{GMAPE_{LT=0}}{GMAPE_{LT=Y}}\right)+1\right]EB_{LT=0}$	44.95%
Median Absolute Percentage Error (MdAPE)	$\left[D_{LT=Y}\left(\frac{MdAPE_{LT=0}}{MdAPE_{LT=Y}}\right)+1\right]EB_{LT=0}$	44.91%
Autoregressive integrated moving average (ARIMA)	See Appendix I	-

where:

LT	Lead time; number of days prior to a target date. Target date occurs when LT=0.
$AB_{LT=Y}$	Actual bookings (room nights reserved to be occupied on the target date) Y days prior to the target date.
$EB_{LT=Y}$	Expected bookings (room nights reserved to be occupied on the target date) Y days prior to the target date. Expected bookings taken from booking curve baseline model for a specific target date and lead time.
$D_{LT=Y}$ where: $D_{LT=Y}=\left(\frac{AB_{LT=Y}-EB_{LT=Y}}{EB_{LT=Y}}\right)$	Booking deviation Y days prior to target date
$APE_{LT=Y}$ where: $APE_{LT=Y}=\left\lvert\frac{EB_{LT=Y}-AB_{LT=Y}}{EB_{LT=Y}}\right\rvert$	Absolute percentage error Y days prior to a target date. Note: APE is generally calculated as a percentage of actual value (AB), whereas in this case it is calculated as a percentage of expected value (EB) so that it may be used as a multiplier.
$MAPE_{LT=Y}$ where: $MAPE_{LT=Y}=\left(\frac{1}{n}\right)\sum_{i=1}^{n}APE_{LT=Y,i}$	Mean absolute percentage error Y days prior to a target date. n is the number of days in the sample (n=1,446).
$GMAPE_{LT=Y}$ where: $GMAPE_{LT=Y}=\left(\prod_{i=1}^{n}APE_{LT=Y,i}\right)^{\left(\frac{1}{n}\right)}$	Geometric mean absolute percentage error Y days prior to a target date. n is the number of days in the sample (n=1,446).
$MdAPE_{LT=Y}$ where: $MdAPE_{LT=Y}=Median(APE_{LT=Y,i})\forall i\in N$	Median absolute percentage error Y days prior to a target date. N is the sample set of days Y days prior to a target date (N=1,446).

APPENDIX I – ARIMA MODELLING OF BOOKING CURVE ADJUSTMENT

The short-term booking curve model was developed to provide immediate estimates of demand at the resort using an Excel spreadsheet without any technical assistance. In order to investigate whether a more thorough procedure could provide better estimates (although requiring a more sophisticated statistical application than MS Excel) an ARIMA approach was tested for the booking curve adjustment. Booking curve baseline error is defined as the difference between the booking curve baseline estimate of bookings to date and actual bookings to date for a specific target date and lead time. The test was to see whether an ARIMA adjusted projection would provide more accurate results than the direct multiplicative approach. The major difference between an ARIMA approach and the direct multiplicative (DM) approach is the number of actual booking to date terms used in the booking curve projection. The DM estimate uses a single booking to date term at the most recent lead time and multiplies that by a baseline derived multiple. The ARIMA approach estimates the pattern of baseline error terms (all error terms to date for a specific target date) and projects that pattern out to the target date. The projected ARIMA error estimate is then added to the original baseline demand estimate in order to achieve the booking curve projection estimate.

The procedure was to apply seven different ARIMA specifications to booking curve baseline errors prior to a target date, and then use the best fitting ARIMA specification based on Akaike Information Criterion (AIC) to forecast booking curve errors out to the target date. In order to have an adequate amount of data upon which to create an ARIMA forecast for each target date, a minimum of 40 data points (lead time days 90 to 50) was used to calibrate the ARIMA error forecast. As the target date approached, more data was used to calibrate the ARIMA forecast (for example, 10 days prior to a target date, 80 data points would be used to calibrate the ARIMA model; lead time days 90 to 10). In order to determine which ARIMA specifications were appropriate, a stratified sample of 45 data sets was modelled using standard ARIMA procedures (see Box & Jenkins (1976)) in SAS ETS (Econometric and Time Series) statistical software. The data sets represented an equal mix of occupancy (high, medium, and low days), bedrooms (one, two, three plus), and lead times (randomly chosen between lead time days 50 and 1). 30 of the 45 sample data sets were completely described (white noise achieved in model residuals) by 7 different ARIMA specifications, while the remaining 15 sample data sets while having more parameters could be approximated quite well by the 7 basic specifications (see Table 11 for 7 ARIMA specifications).

Table 11: ARIMA specifications for booking curve error forecasts

Label	ARIMA description	Deterministic drift	Notation
AR(1)	ARIMA(1,1,0)	No	$\nabla y_t = \frac{a_t}{1-\phi_1 B}$
AR(1) with drift	ARIMA(1,1,0)	Yes	$\nabla y_t = \mu + \frac{a_t}{1-\phi_1 B}$
MA(1)	ARIMA(0,1,1)	No	$\nabla y_t = a_t(1-\theta_1 B)$
MA(1) with drift	ARIMA(0,1,1)	Yes	$\nabla y_t = \mu + a_t(1-\theta_1 B)$
White noise	ARIMA(0,1,0)	No	$\nabla y_t = a_t$
Linear trend	ARIMA(0,1,0)	Yes	$\nabla y_t = \mu + a_t$
Quadratic trend	ARIMA(0,2,0)	Yes	$\nabla^2 y_t = \mu + a_t$

Where:

y_t Booking curve error estimate at time *t*

a_t Random component at time *t*

μ Deterministic drift

B Backward shift operator (e.g. $Ba_t = a_{t-1}$)

∇ Backward difference operator = $1-B$ (e.g. $\nabla y_t = (1-B)y_t = y_t - y_{t-1}$)

The algorithm written in SAS ETS modelled all 7 ARIMA specifications for a specific lead time and target date. The specification with the lowest AIC was then used to forecast ahead h periods (h = lead time) to provide an error estimate for the target date. This estimated error term was then added to the original booking curve baseline for a final booking curve projection estimate. Final ARIMA projections were then compared to DM booking curve projections. A sample of 101 days (with 50 lead time estimates for each day) spanning a sample period from May 15, 1999 to Jan. 20, 2002 was used to compare estimates. The results of the ARIMA methods were mixed. Based on the traditional MSE metric, the ARIMA estimates were 60% worse than the DM estimates. However, based on the more robust median absolute percentage error (MdAPE) metric (see Equation 2 and Equation 4), the ARIMA results provided a 14% improvement over the DM method. Due to the added computational cost (in both computer time and model configuration) and inability to implement ARIMA methods at the resort, the DM method was used for the booking curve adjustment in the booking curve short-term estimate.

REFERENCES

Armstrong, J. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting, 8, 69-80.

Baker, T. & Collier, D. (1999). A comparative revenue analysis of hotel yield management heuristics. Decision Sciences, 30(1), 239-263.

Box, G. & Jenkins, G. (1976). Time Series Analysis: Forecasting and Control, Revised Edition. San Francisco: Holden-Day.

Chatfield, C. (1989). The Analysis of Time Series: An Introduction, 4th edition. New York: Chapman and Hall.

Ghalia, M. & Wang P. (2000). Intelligent system to support judgmental business forecasting: The case of estimating hotel room demand. IEEE Transactions of Fuzzy Systems, 8(4), 380-397.

Harris, P. & Marucci, G. (1983). A short term forecasting model. AGIFORS Symposium Proc. 23. Memphis, USA.

L'Heureux, E. (1986). A new twist in forecasting short-term passenger pickup. AGIFORS Symposium Proc. 26. Bowness-On-Windermere, England.

Min, C. & Zellner, A. (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. Journal of Econometrics, 56, 89-118.

Rajopadhye, M., Ghalia, M., Wang, P., Baker, T., & Ester, C. (1999). Forecasting uncertain hotel room demand. Proc. American Control Conference, San Diego, CA, pp.1925-1999.

SAS Institute Inc. (1999). SAS ETS User's Guide, Version 8, [Electronic version]. North Carolina, USA.

Weatherford, L., Kimes, S., & Scott, D. (2001). Forecasting for hotel revenue management: Testing aggregation against disaggregation. Cornell Hotel and Restaurant Administration Quaterly, August, 53-64.

Woolridge, J. (2000). Introductory Econometrics: A Modern Approach. USA: South-Western College Publishing.