Computers & Operations Research 63 (2015) 23–31

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

A non-parametric approach to demand forecasting in revenue management

Sh. Sharif Azadeh [a,*], P. Marcotte [b], G. Savard [a]

[a] Department of Mathematics and Industrial Engineering, Polytechnique Montréal, Montréal, Québec, Canada H3C 3A7
[b] Department of Computer Science and Operation Research, University of Montréal, Montréal, Québec, Canada H3T 1J4

ARTICLE INFO

Available online 2 May 2015

Keywords:
Revenue management
Forecasting
Integer programming
Branch-and-bound
Heuristics

ABSTRACT

In revenue management, the profitability of the inventory and pricing decisions rests on the accuracy of demand forecasts. However, whenever a product is no longer available, true demand may differ from registered bookings, thus inducing a negative bias in the estimation figures, as well as an artificial increase in demand for substitute products. In order to address these issues, we propose an original Mixed Integer Nonlinear Program (MINLP) to estimate product utilities as well as capturing seasonal effects. This behavioral model solely rests on daily registered bookings and product availabilities. Its outputs are the product utilities and daily potential demands, together with the expected demand of each product within any given time interval. Those are obtained via a tailored algorithm that outperforms two well-known generic software for global optimization.

1. Introduction

According to Cross [7], Revenue Management (RM) is the research area that focuses on the study of disciplined tactics for making product availability and pricing decisions, with the aim of maximizing revenue growth. In the service industry, this goal can only be achieved through accurate demand forecasting, which must take into account the volatility of product availabilities over the booking horizon. Clearly, registered bookings alone are not sufficient to depict the true demand. Indeed, as soon as a product reaches its capacity (*booking limit*), true demand is constrained (*censored*) and cannot be observed. Upcoming customers can then either switch to a higher fare product (*buy-up*), switch to a lower fare product (*buy-down*), or renege (*spill*). According to Weatherford and Belobaba [30], ignoring the data censorship phenomenon can lead to demand underestimation ranging from 12.5% to 25%, and negatively affect revenue by 1–3%, a significant amount for major rail or airline operators.

Although unconstraining techniques may have a big impact on the success of revenue management systems, this topic has not been paid much attention in the literature. In general, two frameworks have been considered to deal with the issue: statistics and optimization. In this paper, we tackle the problem of demand modeling by the use of optimization techniques. However, first, in this section, we present the challenges and rewards of using statistical methods for uncensoring demand.

Based on the literature, statistical techniques such as time series, exponential smoothing, or linear regression have mostly been considered in this context. All of these are able to include seasonal effects within their demand forecasts. Zeni [31] and Queenan et al. [21] have provided a comprehensive study of these methods, and have compared their respective impact on revenue. Their main drawback is that they cannot respond to sudden changes in customer behavior when a product becomes unavailable (see [24,15,20,29,14]). On the other hand, there are many researchers who have addressed the problem from optimization point of view [9,13,18,28]. Actually, authors such as van Ryzin [23] have claimed that revenue management systems should focus on customer behavior and choice probabilities, rather than blindly estimating demand from historical booking data.

Choice-based models were introduced by Andersson [1], and analyzed by Talluri and van Ryzin [26] and Vulcano et al. [27] within the framework of discrete choice theory. In the latter two works, the parameters of the model have been estimated by maximum-likelihood techniques. In another research, Ratliff et al. [22] have integrated historical demand data within a multi-flight heuristic procedure. Also, Vulcano et al. [28] have applied customer choice models to the estimation of product primary demand (first-choice demand). In all the above-mentioned optimization models, a parametric method of estimation (Expectation–Maximization, or EM in short) is used to estimate the parameters of the choice model, under demand independence assumptions. Although the approaches have

* Corresponding author.
E-mail addresses: shadi.sharifazadeh@polymtl.ca (Sh. Sharif Azadeh), marcotte@iro.umontreal.ca (P. Marcotte), gilles.Savard@poltml.ca (G. Savard).

http://dx.doi.org/10.1016/j.cor.2015.03.015
0305-0548/

been used for many years with some success, several issues still need to be addressed:

- Demand across fare products is not independent. Dealing with dependency yields a complex parameter estimation process that has been considered and tested by Stefanescu [25] on small instances.
- As the proportion of censored demand in historical data grows, the accuracy of the standard estimation methods decreases (see [26,28,10]).
- Several statistical methods fail to accurately capture seasonal effects.
- Choice probabilities should enter the optimization process as variables, not as parameters to be estimated. Indeed, these probabilities depend on the set of products available within each time period.

All these issues have motivated us to develop a non-parametric and distribution-free estimation procedure that, based upon historical bookings, takes explicitly into account the set of available products. The contribution of this work is twofold. First, we formulate a model for minimizing the difference between estimated and registered bookings. In order to obtain a realistic representation of customer behavior, cross-temporal utilities enter the model as variables, and seasonal effects are captured by classifying daily demand flows into a predefined number of clusters. Next, we formulate the problem as a MINLP (mixed-integer nonlinear program) , for which we develop a semi-global optimization algorithm.

We close this introductory section with an outline of the paper's structure. Following the description of the problem, together with its underlying assumptions and mathematical formulation (Section 2), we provide a detailed description of the solution algorithm, including the node selection strategy and the valid inequalities used for enhancing the branch-and-bound framework (Section 3). Computational results on synthetic data are analyzed in Section 4, while the concluding Section 5 opens avenues for future research.

2. Problem formulation

To illustrate demand censorship, let us consider the two-product example involving the data displayed in Table 1. As soon as demand for product A exceeds its booking limit 35, which it does since true demand is equal to 40, the data collection system stops counting the number of upcoming customers. As a result, the real demand for A is censored and may exceed 35. In the present case, one A-customer switched to B, while the other 4 reneged.

The main objective of our mathematical model is to minimize the difference between temporal registered bookings and their estimates. Let us introduce its main elements: a *product* i corresponds to a fare class offered at a given period,[1] and is endowed with a utility u_i. The set of products available at a given period j is the *choice set* S_j. A *cluster* c denotes the set of periods that share common features based on the demand flow, such as weekdays, weekends, and holidays. Each *daily potential demand* d_j is associated with a unique cluster. For given utilities u_i and choice sets S_j the *choice probability* p_{ij} of selecting product i on day j is computed according to the multinomial logit (MNL) formula [16]

$$p_{ij}(S_j, u_i) = \begin{cases} \exp(u_i) \Big/ \left(\sum_{k \in S_j} \exp(u_k) + \exp(u_0) \right) & \text{if } i \in S_j \\ 0 & \text{otherwise,} \end{cases} \qquad (1)$$

where u_0 represents the utility of the no-choice option.

[1] Throughout, the terms 'time period' and 'day' are used interchangeably.

Table 1
Demand censorship.

Product	A	B
Availability status	Yes	Yes
Observed demand	35	5
Booking limit	35	6
Real demand	40	4

Table 2
Summary of notations (MINLP).

Sets	
Product	$i \in I = \{1, \dots, \vert I\vert\}$
Day	$j \in J = \{1, \dots, \vert J\vert\}$
Cluster	$c \in C = \{1, \dots, \vert C\vert\}$
Choice set S_j	set of products available on day j
Parameters	
O_{ij}	Observed bookings for product $i \in I$ on day $j \in J$
A_{ij}	Availability status of product $i \in I$ on day $j \in J$
Variables	
d_j	Daily potential demand (integer)
p_{ij}	Probability of selecting product $i \in I$ on day $j \in J$
z_{jc}	Cluster membership variable (binary)
u_i	Utility of product i
δ_c	Potential demand of cluster c

Table 3
Additional summary of notations (RELAX).

Parameters	
R_c^U	Upper bound on potential demand for cluster $c \in C$
R_c^L	Lower bound on potential demand for cluster $c \in C$
D_j^U	Upper bound on potential demand on day $j \in J$
D_j^L	Lower bound on potential demand on day $j \in J$
P_{ij}^U	Upper bound on choice probability for product $i \in I$ on day $j \in J$
P_{ij}^L	Lower bound on choice probability of product $i \in I$ on day $j \in J$
Variables	
e_{ij}	Difference between estimated demand w_{ij} and observed bookings O_{ij}
w_{ij}	Expected demand for product $i \in I$ on day $j \in J$
d_{jc}^N	Normalized daily potential demand $\in [0, 1]$
r_c^N	Normalized potential demand for each cluster $\in [0, 1]$

For a given time horizon, $d_1, d_2, \dots, d_{|J|}$ and a set of products I, we wish to minimize the discrepancy e_{ij} between the expected bookings w_{ij} of each available product i at a given day j and its associated observed registered booking O_{ij}, thus simultaneously capturing seasonal effects and customer behavior. We will therefore have achieved the three following goals:

- external segmentation (classification of days within clusters);
- estimation of daily potential demand;
- estimation of product utilities.

A summary of the notation used in the model is displayed in Tables 2 and 3.

The objective of the model is to minimize the difference between estimated and observed reservations, through the estimation of potential demand, product utilities, and cluster membership. This is achieved by solving the following mathematical model:

$$\text{MINLP}: \quad \min_{\delta_c, u, z, d_j} \sum_{j \in J} \sum_{i \in S_j} (p_{ij}(S_j, u_i) d_j A_{ij} - O_{ij})^2 \qquad (2)$$