Research Article

Estimating unconstrained demand rate functions using customer choice sets

Received (in revised form): 20th February 2010

Alwin Haensel* **and Ger Koole**

Department of Mathematics, VU University Amsterdam, The Netherlands.

Alwin Haensel is a doctoral candidate at the VU University Amsterdam. He received his Master's in Mathematics from the Humboldt University of Berlin with a focus on Stochastic Programming. During his studies he worked in the revenue management department of Avis Europe Plc. His research focuses on revenue management and demand forecasting.

Ger Koole is Full Professor at the VU University Amsterdam. He graduated from Leiden University with a thesis on the control of queuing systems. Since then he has held post-doctoral positions at CWI Amsterdam and INRIA Sophia Antipolis. His current research is centered around service operations, especially call centers, health care and, more recently, revenue management. He has supervised several PhD students and published over 60 papers in the international literature.

***Correspondence:** Alwin Haensel, Department of Mathematics, VU University Amsterdam, 1081 HV Amsterdam, The Netherlands.
E-mails: ahaensel@few.vu.nl, koole@few.vu.nl

ABSTRACT A good demand forecast should be at the heart of every revenue management model. Yet most demand models focus on product demand and do not incorporate customer choice behavior under offered alternatives. We use the ideas of customer choice sets to model the customer's buying behavior. A customer choice set is a set of product classes representing the buying preferences and choice decisions of a certain customer group. In this article we present a demand estimation method for these choice sets. The procedure is based on the maximum likelihood method, and to overcome the problem of incomplete data or information we additionally apply the expectation maximization method. Using this demand information per choice sets, the revenue manager obtains a clear view of the underlying demand. In doing so, the sales consequences from different booking control actions can be compared and the overall revenue maximized.
Journal of Revenue and Pricing Management, advance online publication, 30 May 2010; doi:10.1057/rpm.2010.1

Keywords: revenue management; demand forecasting; data unconstraining; customer choice; choice sets; EM algorithm

INTRODUCTION

Revenue management (RM) is concerned with maximizing profits of selling a limited number of a perishable product by controlling the price and availability for sale. Since the 1980s this topic has been extensively investigated by the airline industry owing to the deregulation of the market. Over the last decade there has also been a huge growth in literature focusing on RM problems, mostly set in an airline context, but also increasingly in the specific contexts of the hotel, car rental and logistics industries. Most of this work is reviewed in Talluri and Van Ryzin (2004b).

A majority of scientific papers concentrate on the optimization model of the RM problem. Generally they start the analysis by assuming that the demands for different products are independent of one another. This is a nice property for the model, but a very strong assumption. For example, let us consider two flights with the same itinerary, both departing with a time leg of 30 min and both having different prices. The independent product demand assumption would mean that if we vary the price of one flight this would have no impact in the demand for the other flight. Observing one's own buying behavior brings us easily to the conclusion that the customer choice depends crucially on the alternatives with which the customer is confronted. Cooper *et al* (2006) examine the 'spiral-down' effect in the dynamic behavior of RM models, for example neglecting the fact that availability of low-priced tickets can reduce the sales of high-priced tickets. In most cases the observed booking data do not reflect the real demand. Usually demand is not recorded after a flight or a hotel has been sold out, or, more specifically, for a price class if a booking limit has been reached. Therefore, sales observations are usually denoted as censored demand data. Some research papers are devoted to the topic of unconstraining censored data (see McGill, 1995; Lui *et al*, 2002; Weatherford and Poelt, 2002). A comparison of all commonly known unconstraining methods is given in Queenan *et al* (2007). In all these papers the number of the overall demand is unconstrained per product. These give some interesting insight into demand in general. However, the interesting questions for practitioners at a certain point in the booking horizon are 'What control actions should be taken?' and 'What effects will these actions have?' To answer these questions one needs information not only about the general demand arrivals, but also about the customer's decision process. This allows us to predict sales consequences for certain control actions. The above approaches do not consider the customer choice behavior; as Van Ryzin (2005, p. 206) formulated: what is needed in Revenue Management research is a change from product demand models to models of customer behavior. There are some research papers that deal with the problem of customer choice with regard to the RM issue: Andersson (1998) and Algers and Beser (2001) describe a pilot study at Scandinavian Airlines. They formulate a model for optimal seat allocations that takes into account that customers might buy up to higher fare classes or be recaptured on a different flight. The customer choice is modeled by a Multi Nomial Logit (MNL) model, and its parameters are estimated from interviews and historical sales data. Ratliff *et al* (2008) propose a multi-flight heuristic generalizing the earlier approach. For their model, in addition to historical sales and fare class availability data, they require information about the airline's market share and values for the relative customer attractiveness of flights and fare class alternatives. Talluri and Van Ryzin (2004a) study customer behavior using a discrete choice model, and state an optimal policy for fare class availability control. They suggest using maximum likelihood estimation for the attributes weight vector in an MNL to compute sale probabilities. The customer arrival rate is assumed to be constant over time. Vulcano *et al* (2009) focus on estimating the primary demand (customer's first-choice demand if all alternatives available) from historical sales and fare class availability data. Demand is modeled by a Poisson process over multiple time periods. Customers are assumed to choose fare classes among alternatives according to an MNL model.

From our data analysis we conclude that customer arrivals follow a time inhomogeneous process, where the intensity is strictly monotone increasing to the usage date of the product (end of booking horizon). In this article we are interested in establishing estimates for demand rate functions reflecting customer choice behavior from historical sales data. In contrast to previous studies, we do not assume a general customer arrival rate, with sale probabilities resulting from offered alternatives. Rather, we assume the demand to be made up of different customer groups representing different buying behavior and characteristics. Demand from

Table 1: Booking classes in airline example with groups high, medium and low

Booking class	*Miles earned (%)*	*Changes*	*Cancellations*	*Price*
H1	100	Charge 50	Charge 100	520
H2	100	Charge 50	Charge 100	460
M1	50	Charge 50	No	370
M2	50	Charge 50	No	320
L1	25	No	No	250
L2	25	No	No	230

each group can result in sales of different products depending on the sellers and competitors' offer of alternatives. We call these groups customer choice sets. Using the demand information on customer choice set level enables the RM practitioner to compute the optimal booking control action. In the airline context, a booking control action consists of enabling or disabling certain flight and fare class combinations for future bookings. With the right decisions, one is able to force certain customers to buy up to their upper budget limit.

The article continues in the next section with our customer choice model, followed by a section on some analysis on demand behavior. In the section after that an optimal allotment control based on dynamic programming is described. The parameter estimation method for the choice set demand is explained in the subsequent section. The numerical results are given in the penultimate section and the final section presents our conclusions.

CUSTOMER CHOICE SETS

In this section we explain the idea of a discontinuous choice set as earlier described in Ben-Akiva and Lerman (1985). The abovementioned idea wherein probabilities are attached to choice alternatives is now altered into a preference ordering of choice alternatives.

We have a product to sell, for example seats on an airplane or hotel rooms. The product offer is divided into several price classes. We can even assume the market to be segmented in such a way that different conditions are attached to the product itself, for example minimum length of stay, cancellation costs or membership credits. Each of these groups is again differentiated into different price categories. All of the resulting offers are called classes or subclasses (see Table 1 for an example segmentation of an airline). At each point in time all subclasses may or may not be available to buy. As the conditions on the product within a group are equal, except price, naturally a booking within a group will be in the lowest available subclass of the group. The customer's utility of a product decreases when the price increases while the conditions are unchanged. We assume that a customer always tries to maximize his utility within his decision process (compare to Ben-Akiva and Lerman, 1985). Thus, at each time a group of subclasses with the same product conditions has one active price, which is equal to the lowest available subclass in the group. In our model we assume that each customer has a set of subclasses that represent his willingness to buy (regarding price and conditions). We denote this set choice set and it may consist of any coherent sequence of subclasses, for example {L2, L1, M2}. In our context it makes no sense to consider choice sets that are not coherent. From the customer point of view this would mean that a customer is interested in buying low- or high-priced tickets, but no medium-priced tickets. It is obvious that we can neglect these choice sets. With a similar reflection of marketing and sales ideas we can also delete very large choice sets from our investigation. People will not consider too many different subclasses in there choice set because the products are very strongly segmented, that is, a customer who considers buying an upper-high-fare ticket is mostly not seen to be also interested in extremely low-fare tickets. As the price is still the major factor in the decision-making process, we say that the customer will always choose the lowest priced available subclass within his choice set. We display the subclasses in the choice sets in hierarchical order increasing from left to right. If none of the subclasses in which the customer is interested in is available, he will not buy at all.

Thus far we presented the idea of choice sets in the context of a single resource and a monopoly seller market. However, this can be easily extended to multiple resources, for example multiple daily flights with one itinerary, and competitor offers can be easily embedded in the choice sets structure. We only need to have a strict preference ordering, so that the customer is never indecisive between two offers. For example, consider two similar hotels with different distances to the sea. In general a customer prefers to be closer to the sea, but for a certain saving he will make a booking in the hotel further from the sea. Consequently, a choice set may consist of subclasses from different comparable resources and sellers.

DEMAND FUNCTIONS

For our analysis we were able to work with real sales data of a major airline and a European hotel reservation agency. Both problems are very similar. We have a perishable good: seats on a certain flight or hotel rooms for a certain date. For both data sets we observed that the demand increases as we approach the usage date of our product. We cannot usually observe the true demand. The company may sell all its capacity or influence the booking process, when making certain price offers available or unavailable for booking. We say a price class is open when it is available for bookings and closed otherwise. Figure 1 shows the airline sales behavior for a fixed itinerary and departure day of week combination. It contains all bookings with no differentiation of fare classes aggregated over some weeks. An example of the hotel sales behavior with aggregated sales data over multiple regions, as well as a single region, is shown in Figure 2. The data

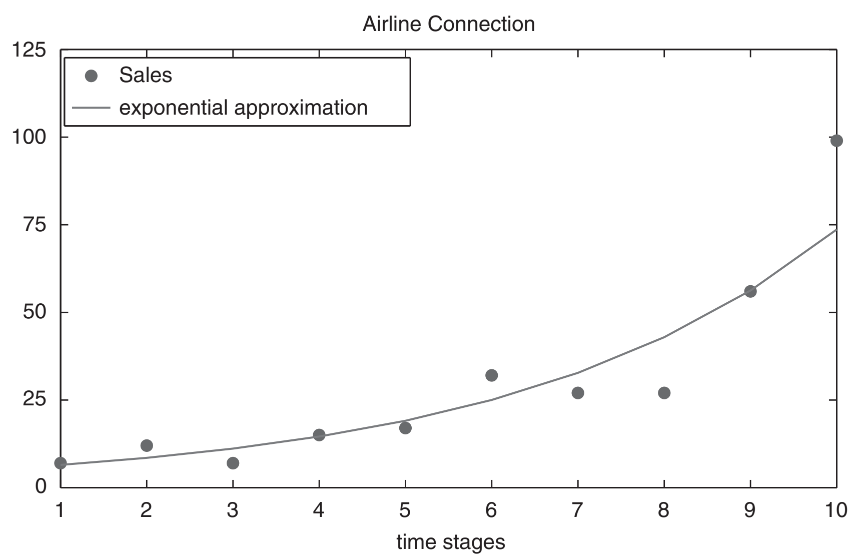

Figure 1: Airlane sales behavior.

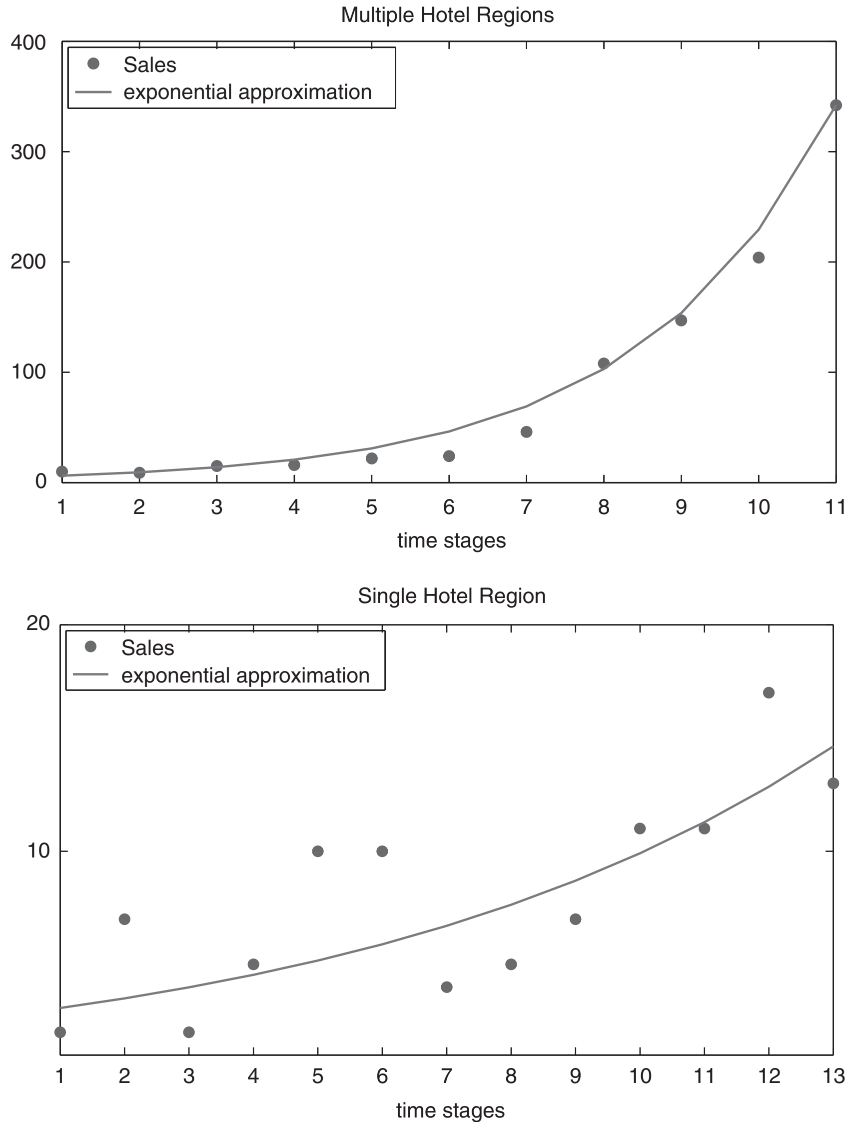

Figure 2: Hotel sales behavior.

correspond to a fixed arrival date. As the hotel data set contains reservation data for multiple comparable hotels, we very rarely observe total area sell-outs, that is, a case in which a whole price class within an area is unavailable for bookings. Thus, we are able to observe the untruncated booking behavior with a kind of bird's eye view. In addition to the monotonic increasing demand behavior, we observe that the quantiles on days in advance (distance between booking-creation date and usage date of the product) in our hotel data set are consistent over time. In addition, from a correlation analysis on both data sets we see a high correlation between early and late bookings. This confirms our idea that demand rate follows a non-decreasing curve over time. From least squares analysis on an aggregated level, we learned that the demand rate per choice set c follows an approximate exponential curve:

$$\lambda_c(t) = \beta_c \cdot e^{\alpha_c t} \tag{1}$$

for time stage t and parameters α_c and β_c, which are supposed to be positive. Further, we tested on the airline data whether the demand per time stage is Poisson distributed. We used the likelihood ratio, the conditional chi-squared and the Neyman–Scott statistic as statistical tests, and found that the hypothesis cannot be rejected. Consequently, the demand of choice set c is assumed to follow an inhomogeneous Poisson Process with rate function $\lambda_c(t)$.

With the given demand rate functions for all choice set, we were able to compute the following interesting parameter. The arrival rate for a customer of any type at time stage t is given by $\lambda(t)=\sum_{c\in C}\lambda_c(t)$, the sum of arrival rates over C – the set containing all choice sets. Let us suppose that we fixed a time stage t: We must make a decision on which price classes/products to offer, that is, choosing a set $O\subseteq N$, where N denotes the set of all price classes. The probability that an arriving customer buys product j when O is offered is given by $P_j(O,t)$, where $P_j(O,t)=0$ if $j\notin O$. We can compute this probability by

$$P_j(O,t)=\frac{\sum_{c\in C}\lambda_c(t)\cdot \mathbb{I}_{\{U(c,O)=j\}}}{\lambda(t)} \quad (2)$$

where $\mathbb{I}$ denotes the indicator function and $U(c, O)$ is a function returning the offered price class with the highest utility in choice set c, that is, the lowest available class of c or zero if $c\cap O=\emptyset$. Further, we can compute the no-sales probability $P_0(O,t)$, that is, a customer arrives but does not decide to buy one of the offered products in O at time t. The probability is obtained by

$$P_0(O,t)=\frac{\sum_{c\in C}\lambda_c(t)\cdot \mathbb{I}_{\{U(c,O)=0\}}}{\lambda(t)} \quad (3)$$

The primary demand (first-choice demand), studied by Vulcano *et al* (2009), can also easily be computed. The estimated primary demand of product j at each time t is the sum of all demand rates $\lambda_c(t)$ over all choice sets c, where j is the first-choice price class, that is, that with the highest preference.

OPTIMAL ALLOCATION CONTROL

An optimal control policy for a single resource problem including customer choice behavior is described by Talluri and Van Ryzin (2004a, b). In the following we give a brief description of the dynamic programming model and how it is applied in our context. For detailed proofs the reader is referred to Talluri and Van Ryzin (2004a). The original model considers a time horizon divided into intervals $i=1,\ldots,I$ such that we have at most one customer arrival per interval. The probability of arrival is denoted by λ and assumed to be constant in time.

In our case we assume an inhomogeneous Poisson process, in which the arrival rate is supposed to be constant within each time stage $t=1,\ldots,T$. Therefore, each time stage is divided into such intervals I_t, such that we have at most one arrival per interval. The probability of any arrival at time stage t in all intervals i is given by $\Lambda(t)=\lambda(t)/I_t$. Let us suppose that we fixed a time stage t:

At each interval i we can make a decision on which fare classes to offer, that is, choosing a set $O\subseteq N$, where N denotes the set of all price classes. The probability that an arriving customer buys product j when O is offered is given by $P_j(O,t)$, where $P_j(O,t)=0$ if $j\notin O$. We can compute this probability as in the previous section by

$$P_j(O,t)=\frac{\sum_{c\in C}\lambda_c(t)\cdot \mathbb{I}_{\{U(c,O)=j\}}}{\lambda(t)} \quad (4)$$

Thus, the probability of a sale of class j in interval i is given by $\Lambda(t)\ P_j(O,t)$, given that price classes in O are offered. Similarly, we can compute the no-sales probability $P_0(O,t)$ for each arrived customer by

$$P_0(O,t)=\frac{\sum_{c\in C}\lambda_c(t)\cdot \mathbb{I}_{\{U(c,O)=0\}}}{\lambda(t)} \quad (5)$$

The no-sale probability at each interval i is then $\Lambda(t)\cdot P_0(O,t)$. The value function is defined as the maximal obtainable future revenue from period i in time stage t and inventory level x. The Bellman equation is then given by

$$V_i(x,t)=\Lambda(t)\cdot\max_{O\subseteq N}\{R(O,t)-Q(O,t)\times\Delta V_{i+1}(x,t)\}+V_{i+1}(x,t) \quad (6)$$

where $\Delta V_{i+1}(x,t)=V_{i+1}(x,t)-V_{i+1}(x-1,t)$. The purchase probability $Q(O,t)$ and the expected revenue $R(O,t)$ from offering set O are computed by

$$Q(O,t)=\sum_{j\in O}P_j(O,t)$$

$$R(O,t)=\sum_{j\in O}r_j\cdot P_j(O,t) \quad (7)$$

On the boundaries the value function is defined as

$$V_i(0,t)=0 \quad \forall i=1,\ldots,I$$

$$V_{I+1}(x,t)=\begin{cases}V_1(x,t+1), & \text{if } t<T\\ 0, & \text{else}\end{cases} \quad \forall x$$

Talluri and Van Ryzin show in their paper that we only have to consider the subclass of efficient sets over all possible subsets of N, when making the decision as to which set of price classes to offer.

Definition 1 A set $S\subseteq N$ is efficient if there exist no non-negative weights $\alpha(M)$ on all $M\subseteq N\cup\{\emptyset\}$ with $\sum_{M\subseteq N\cup\{\emptyset\}}\alpha(M)=1$ such that

$$\frac{R(S,t)}{Q(S,t)}\geqslant\frac{\sum_{M\subseteq N\{\emptyset\}}\alpha(M)R(M,t)}{\sum_{M\subseteq N\{\emptyset\}}\alpha(M)Q(M,t)}$$

We assume that the collection of efficient sets E is indexed in increasing revenue and probability order. This is possible because for an indexing of E such that $Q(S_1,\ t)\leqslant Q(S_2,\ t)\leqslant\cdots\leqslant\ Q(S_m,t)$ it follows that the expected revenues are ordered in the same way. The problem of maximizing (6) is simplified by choosing among all sets in E:

$$V_i(x,t)=\Lambda(t)\cdot\max_{k=1,\ldots,m}\{R(S_k,t)-Q(S_k,t)\times\Delta V_{i+1}(x,t)\}+V_{i+1}(x,t) \quad (8)$$

Theorem 1 Consider k^* such that S_{k^*} maximizes (8). For a fixed i, the largest optimal index k^* increases in the remaining capacity, and for any fixed x, k^* decreases in the remaining intervals.

The problem of identifying the efficient sets can be solved in an inductive way. Initially we set $S_0=\emptyset$. Having the rth efficient set S_r we find the $(r+1)$th by

$$\text{argmax}_{S\subseteq N}\frac{R(S,t)-R(S_r,t)}{Q(S,t)-Q(S_r,t)} \quad (9)$$

Definition 2 A control policy is called nested if there exists an increasing family of subsets $S_1\subseteq S_2\ \subseteq\cdots\subseteq\ S_m$, with $S_j\subseteq N$, $j=1,\ldots,m$. The set index $k_i(x)$, increasing in x, is chosen at interval i when the remaining capacity is x.

Classes are seen as 'higher' if they appear earlier in the sequence. In addition, a policy is in fare order if the nesting order coincides with the fare order. In the case of a nested optimal policy we can compute the protection levels by

$$p_k^*(i)=\max\ x \quad \text{s.t.}\quad \Delta V_{i+1}(x,t)>\frac{R(S_{k+1},t)-R(S_k,t)}{Q(S_{k+1},t)-Q(S_k,t)} \quad (10)$$

At interval i only the classes contained in S_k are open if the remaining capacity is less than $p_k^*(t)$. Nested booking limits are equivalently defined by $b_k^*(i)=\text{capacity}-p_{k-1}^*(i)$.

In most cases we have this nested policy, and usually the optimal policy is even nested by fare order. Nevertheless, the nested structure is not certain and the hypothesis has to be checked to apply (10) for the computation of optimal protection levels. In our test cases the policy are in fare order.

PARAMETER ESTIMATION

The demand is assumed to follow an inhomogeneous Poisson process. As described in the 'Demand Functions' section, we assume the demand rate per choice set to be approximately as follows:

$$\lambda_c(t) = \beta_c \cdot e^{\alpha_c t} \tag{11}$$

for time stage t in the booking horizon $t = 1, \ldots, T$ and parameters α and β depending on choice set c, which are supposed to be positive. In our analysis we only work with observable sales data. Our goal is to unconstrain these data to obtain a good approximation of the real demand per choice set. The given data sets contain information as to whether a subclass was open for bookings at a certain time and if yes the observed number of bookings or sales. For the parameter estimation we use the maximum likelihood method. The likelihood function for choice set c is

$$L_c(\alpha_c, \beta_c) \prod_{t=1}^{T} P(S_c(t)|\lambda_c(t)) \tag{12}$$

where $S_c(t)$ denotes the number of sales corresponding to choice set c at time stage t. Remember that a sale corresponding to a certain choice set will always materialize in its lowest available subclass. We denote with 'c open at t' the fact that at least one subclass in c was open for bookings at time t. Thus, we obtain an expression for probabilities by

$$P(S_c(t)|\lambda_c(t)) = \begin{cases} \dfrac{e^{-\lambda_c(t)} \cdot \lambda_c(t)^{S_c(t)}}{(S_c(t))!}, & \text{if } c \text{ open at } t \\ 1, & \text{else} \end{cases} \tag{13}$$

As we are interested in the maximum of the likelihood function, we examine the loglikelihood function

$$\begin{aligned} \mathcal{L}_c(\alpha_c, \beta_c) &= \log(L_c(\alpha_c, \beta_c)) \\ &= \sum_{t=1}^{T} \log(P(S_c(t)|\lambda_c(t))) \\ &= \sum_{t=1}^{T} S_c(t)(\log(\beta_c) + \alpha_c t) \\ &\quad - \beta_c \exp(\alpha_c t) - \log(S_c(t)!) \end{aligned} \tag{14}$$

for which the gradient and Hessian are given by

$$\nabla \mathcal{L}_c = \sum_{t=1}^{T} \begin{pmatrix} S_c(t)t - \beta_c t \exp(\alpha_c t) \\ \frac{S_c(t)}{\beta_c} - \exp(\alpha_c t) \end{pmatrix} \tag{15}$$

$$H(\mathcal{L}_c) = \sum_{t=1}^{T} \begin{pmatrix} -\beta_c t^2 \exp(\alpha_c t) & -t \exp(\alpha_c t) \\ -t \exp(\alpha_c t) & -\frac{S_c(t)}{\beta_c^2} \end{pmatrix} \tag{16}$$

The following theorem will help us in the maximization of our log-likelihood function.

Theorem 2 The negative log-likelihood function is unimodal and thus has a unique minimizer in $\mathbb{R}^2_+$.

Proof: We use Theorem 50 from Demidenko Criterion of Unimodality. Let $F(u)$ be a twice differentiable function of $u \in \mathbb{R}^n$ such that $||u|| \to \infty$ implies that $F(u) \to \infty$. If at each point where the gradient is zero the Hessian is positive definite, then F has a unique minimum on $\mathbb{R}^n$, that is, the function is unimodal.

We define $-\mathcal{L}_c(\alpha_c, \beta_c) = \infty$ for α_c or β_c being non-positive. Further, we have that if $||(\alpha_c, \beta_c)|| \to \infty$ the likelihood (12) $L_c(\alpha_c, \beta_c) \to 0$ and this implies that $-\mathcal{L}_c(\alpha_c, \beta_c) \to \infty$. It remains to be shown that the Hessian of the negative log-likelihood function is positive definite at all stationary points, that is, where $\nabla(-\mathcal{L}_c) = 0$. To check the positive definiteness of a matrix, we must verify that all the determinants of its leading principal minors are positive. By differentiating $-\mathcal{L}_c(\alpha, \beta)$ (see (14)), we easily conclude that the gradient is given by $\nabla(-\mathcal{L}_c) = -\nabla\mathcal{L}_c$ and from this we can compute the Hessian by $H(-\mathcal{L}_c) = -H(\mathcal{L}_c)$ and obtain

$$H(-\mathcal{L}_c) = \sum_{t=1}^{T} \begin{pmatrix} \beta_c t^2 \exp(\alpha_c t) & t\exp(\alpha_c t) \\ t\exp(\alpha_c t) & \frac{S_c(t)}{\beta_c^2} \end{pmatrix}$$

The determinant of the first leading principal minor is $\sum_{t=1}^{T} \beta_c t^2 \exp(\alpha_c t) > 0$, as $\beta_c > 0$. For the second leading principal minor, which coincides with the Hessian, we need the condition that the gradient is zero.

$$\sum_{t=1}^{T} S_c(t)t - \beta_c t\exp(\alpha_c t) = 0 \qquad (17)$$

$$\sum_{t=1}^{T} \frac{S_c(t)}{\beta_c} - \exp(\alpha_c t) = 0 \qquad (18)$$

The determinant of the second leading principal minor is given by

$$\left(\sum_{t=1}^{T} t^2 \exp(\alpha_c t)\right)\left(\sum_{t=1}^{T} \frac{S_c(t)}{\beta_c}\right) - \left(\sum_{t=1}^{T} t\exp(\alpha_c t)\right)^2 \qquad (19)$$

Using (18) we can eliminate $S_c(t)$ in (19)

$$\left(\sum_{t=1}^{T} t^2 \exp(\alpha_c t)\right)\left(\sum_{t=1}^{T} \exp(\alpha_c t)\right) - \left(\sum_{t=1}^{T} t\exp(\alpha_c t)\right)^2 \qquad (20)$$

which can further be simplified to

$$\sum_{t_1, t_2=1}^{T} (t_1^2 - t_1 t_2)\exp(\alpha_c(t_1 + t_2)) \qquad (21)$$

We add the term $(t_2^2 - t_1 t_2)$ exp $(\alpha_c(t_1 + t_2)$ into (21) and obtain

$$\sum_{t_1, t_2=1}^{T} \underbrace{\exp(\alpha_c(t_1 + t_2))}_{>0} \underbrace{t_1^2 - t_1 t_2 + t_2^2 - t_1 t_2}_{=(t_1 - t_2)^2 \geqslant 0} \qquad (22)$$

For $T > 1$ it follows that (22) is positive. From symmetry we have that

$$\sum_{t_1, t_2=1}^{T} (t_1^2 - t_1 t_2) = \sum_{t_1, t_2=1}^{T} (t_2^2 - t_1 t_2)$$

and so we can rewrite (22) into

$$\sum_{t_1, t_2=1}^{T} \exp(\alpha_c(t_1 + t_2)) \cdot 2(t_1^2 - t_1 t_2) > 0 \qquad (23)$$

Dividing (23) by two gives that (21) is positive and hence the second leading principal minor is positive. Thus, we have shown the unimodality of the negative log-likelihood function. □

We obtain the maximum likelihood estimates for the parameters α and β for choice set c by minimizing the negative log-likelihood function

$$(\hat{\alpha}_c, \hat{\beta}_c) = \arg\min_{\alpha, \beta > 0} \quad -\mathcal{L}_c(\alpha, \beta) \qquad (24)$$

For now, we did not take the interaction of choice sets into account and supposed that we know to which choice set a sale corresponds. Let us illustrate this with a small example. We have two overlapping choice sets: $Set_1 = \{L3, L2\}$ and $Set_2 = \{L3, L2, L1\}$. Say we observe at a certain day a sale in subclass $L3$; we do not know whether this is a realization corresponding to a customer of Set_1 or Set_2, as we only have information on whether a class was open, and if yes, the number of sales in it.

This leads to the following log-likelihood function:

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{f=1}^{F} \log P\left[X = S(t,f) | X = Poission\left(\sum_{c \in C} \mathbb{I}_{\{U(c,t)=f\}} \cdot \lambda_c(t)\right)\right] \quad (25)$$

where $\mathbb{I}$ denotes the indicator function, $U(c, t)$ returns the lowest available price class in choice set c at time t ($U(c, t) = 0$ means no available class in c at time t), F represents the number of price classes and $S(t,f)$ denotes the observed sales in price class f at time t. From simulation we learned that the negative log-likelihood function is generally not unimodal. As we are usually confronted with multiple overlapping choice sets we must consider $2 \cdot |C|$ variables in the resulting non-convex optimization problem.

To overcome this problem we suggest the expectation maximization (EM) method explained in Dempster *et al* (1977). Initially we estimate the parameters separately for all choice sets by ignoring the interaction of choice sets. Further, we use the fact that the random combination of Poisson processes again gives a Poisson process. Given the initial estimates of the corresponding rates $\lambda_c^0(t)$ for all times t and all choice sets c, we start the iterative procedure at $i = 1$ by computing

$$P_c^i(S(t,f)) = \begin{cases} P\left[X = \left\lceil \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} \cdot S(t,f) \right\rceil\right] & \text{if } f > 0 \\ \times \mathbb{I}_{\{U(c,t)=f\}}, & \\ 1, & \text{else} \end{cases}$$
$$\text{where } X \sim Poisson(\lambda_c^i(t)) \quad (26)$$

The $\lceil x \rceil$ operator returns the closest integer greater than or equal to x. $\lambda_{overlap}^j(c, t)$ denotes the sum of the estimated rates from iteration j over all choice sets for which the lowest available price classes coincide with choice set c. In the maximization step, the resulting negative log-likelihood function is then minimized separately for all choice sets c to obtain new estimates of $\lambda_c^i(t) = \beta_c^i \exp(\alpha_c^i t)$:

$$(\alpha_c^i, \beta_c^i) = \arg \min_{\alpha, \beta > 0} \; -\mathcal{L}_c(i) \quad (27)$$

where the log-likelihood function at iteration i for choice set c is given by

$$\mathcal{L}_c(i) = \sum_{t=1}^{T} \log P_c^i(S(t, U(c,t))) \quad (28)$$

With the new rate estimates for choice set c, we update the previous estimates in equation (26), which is the expectation step. The procedure is repeated until it converges. Although convergence is in general not certain, the EM method is known to be very robust, and has been satisfactorily employed by McGill (1995), Weatherford and Poelt, Talluri and Van Ryzin (2004a) and Vulcano *et al* (2009).

For the maximization step we are again interested in the shape of the negative log-likelihood function and the formulation of its gradient and Hessian. As the log-likelihood function is computed separately for each choice set, we can compute the gradient and Hessian at each iteration i as in the initial case for non-overlapping choice sets:

$$\nabla \mathcal{L}_c(i) = \sum_{t=1}^{T} \begin{pmatrix} S(c,t,i)t - \beta_c t \exp(\alpha_c t) \\ \frac{S(c,t,i)}{\beta_c} - \exp(\alpha_c t) \end{pmatrix} \mathbb{I}_{\{U(c,t)>0\}} \quad (29)$$

$$H(\mathcal{L}_c(i)) = \sum_{t=1}^{T} \begin{pmatrix} -\beta_c t^2 \exp(\alpha_c t) & -t\exp(\alpha_c t) \\ -t \exp(\alpha_c t) & -\frac{S(c,t,i)}{\beta_c^2} \end{pmatrix} \mathbb{I}_{\{U(c,t)>0\}} \quad (30)$$

If c is open the number of corresponding sales is given by

$$S(c,t,i) = \left\lceil \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} \cdot S(c,U(c,t)) \right\rceil \quad (31)$$

With the same arguments as in Theorem 2 we can show that the negative log-likelihood function (28) is again unimodal. This proves that the M-step is always well defined.

EM algorithm for unconstraining demand rate function per customer choice sets
Initialization: compute separately for all choice sets $c \in C$

$$(\alpha_c^0, \beta_c^0) = \arg \min_{\alpha,\beta>0} \quad -\sum_{t=1}^{T} \log \times \begin{cases} P(X_t = S(t,U(c,t))), & \text{if } U(c,t) > 0 \\ 1, & \text{else} \end{cases}$$

where $X_t \sim$ *Poisson* $(\lambda = \beta \exp(\alpha t))$.
Iteration Loop $i = 1, \ldots$

E-step:

For all $c \in C$

For all $t = 1, \ldots, T$

$$P_c^i(S(t,U(c,t))) = \begin{cases} P\left[X = \left\lceil \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} \cdot S(t,U(c,t)) \right\rceil\right], & \text{if } U(c,t) > 0 \\ 1, & \text{else} \end{cases}$$

$$\text{where } X \sim Poisson(\lambda_c^i(t) = \beta_c^i \exp(\alpha_c^i t))$$

end for

$$\mathcal{L}_c(i) = \sum_{t=1}^{T} \log P_c^i(S(t,U(c,t)))$$

end for

M-step:

For all $c \in C$

$$(\alpha_c^i, \beta_c^i) = \arg \min_{\alpha,\beta>0} \quad -\mathcal{L}_c(i)$$

end for

Until Stopping criteria reached.

The stopping criteria could be either a maximum number of iterations or some kind of numeric convergence bound on changes in α and β between iterations. The demand rate function for each choice set c at time t is estimated as long as at least one class contained in c is offered at t. A no-sale observation at t is regarded as a realization of the stochastic arrival process with no arrivals from all choice sets that intersect with the set of offered price classes at time t. The estimation is extrapolated over periods when no price class contained in choice set c is offered, that is, no customer arrivals corresponding to c are observable. Notice that the EM method is independent of the demand rate function form. The exponential curve is a result from our data analysis of the airline and hotel data sets. Only the unimodality of the negative log-likelihood function needs to be checked for different demand functions; it can be easily shown that it also holds for constant or linear demand functions.

NUMERICAL TESTS

In this section we present the numerical results of our estimation method. We evaluate the estimation results on two criteria: first on how well the real rate function is approximated and second on the expected revenue gain by using the estimated values compared to the real values in the optimal control policy (section 'Optimal Allocation Control'). These are evaluated on 100 independent demand scenarios, and the obtained revenues are averaged. The computation is carried out in Matlab 7.4.0; for optimization we used the function 'fmincon' from the optimization toolbox. The starting parameters for 'fmincon' are computed using a small Monte Carlo simulation.

Our input sales simulations are generated by independent demand realization from the known demand functions per choice sets. Here the booking process is controlled via given initial booking limits. Using this procedure we generate 1000 sales observations from which we estimate the parameters α and β for each choice set. The resulting estimates are averaged. In all examples the booking horizon is set to 10 time stages.

The EM method is stopped if the maximal change of the minimized value of the negative log-likelihood function over all choice sets is less than 10^{-6} per cent compared to the optimum from the previous iterations. In all the following cases this convergence bound is reached within less than 30 steps.

Two overlapping choice sets

First, we analyze the case of two overlapping choice sets, considering two products, A and B, (prices are 50 and 100) with the corresponding choice sets $Set_1=\{A\}$ and $Set_2=\{A,B\}$ (parameters are shown in Table 2). The initial booking limits are 10 and 10 for products A and B, respectively (see Figure 3 to compare the estimated demand function with the original).

After time stage 6, the estimated demand function for Set_1 is a very poor estimate of the real underlying demand function. The booking limit for product A is reached between time stages 5 and 7. This means that the parameters for Set_1 are only fitted to sales data up to this moment and extrapolated into the future. In contrast, we observe that the estimated demand function for Set_2 approximates the real function closer to the end of the booking horizon. When A is closed we know for sure that we observe a realization of Set_2 and thus the estimate is more precise. The estimated α and β parameters for both choice sets, and the generated mean revenue using the estimates and real parameter in the optimal booking control, are shown in Table 2. With perfect information about the demand functions for both choice sets, we generate in average 12.9 sales of product A and 6.06 of product B giving an average generated revenue of 1251. With our estimated parameters we generate on average 11.6 and 5.26 sales for products A and B, respectively, which results in a mean revenue of 1106. This means that from only given sales data of products A and B (generated using non-optimal booking limits), we were able to approximate the demand rate functions such that using these estimates we can gain

Table 2: Case 1 – Parameter and generated revenue

	α_{Set_1}	β_{Set_1}	α_{Set_2}	β_{Set_2}	*Revenue*
Original	0.1	1	0.2	0.3	1251
Estimates	0.1646	0.9249	0.0736	0.9124	1106

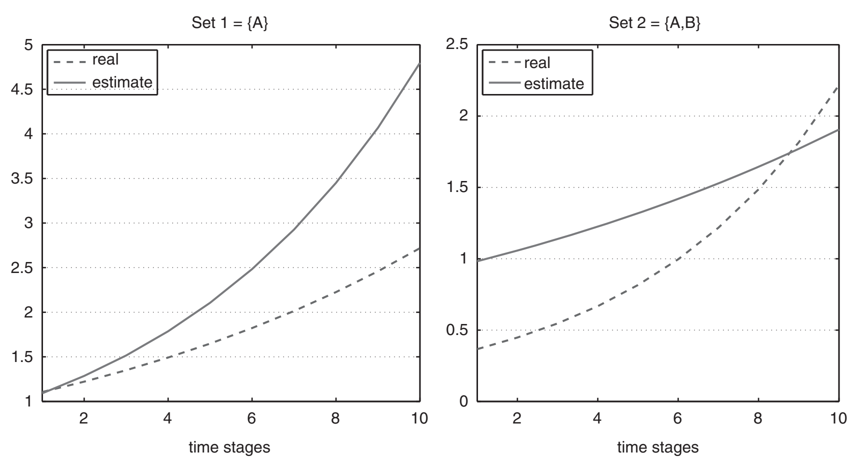

Figure 3: Case 1 – Parameter and generated revenue.

88.4 per cent of the expected revenue from having perfect demand rate information (*PI Rev*). For the moment we analyzed a case with small demand rate values. In case 2 we increase the α and β parameters for Set_1 and Set_2, to determine whether the precision of our estimation performs better for higher rate values. The new parameter and the new estimates are shown with their generated revenue in Table 3. The initial booking limits to generate the sales data are 50 and 100 for products A and B, respectively; the prices are the same as above (see Figure 4 for a comparison of the real and estimated demand functions). The demand rate function for Set_2 is now much better approximated. As in the previous case, we observe a better fit for time stages when A is closed. For Set_1 we now observe an underestimation of the demand rate, in contrast to the previous case in which the estimates overestimated the demand rate. This underestimation seems to be very large, but analyzing the obtained revenue from the estimated parameter we observe that this lack of accuracy does not imply a huge loss of revenue. In fact, with our estimates we are very close to the revenue obtained by perfect demand rate information. With this perfect information about the demand functions, we can generate on average 63.52 sales of product A and 72.78 sales of product B. With the estimated parameter we generated on average 26.15 sales of product A and 88.26 sales of product B. This means that by using estimated demand functions we can generate 97 per cent of the *PI Rev*.

Table 3: Case 2 – Parameter and generated revenue

	α_{Set_1}	β_{Set_1}	α_{Set_2}	β_{Set_2}	*Revenue*
Original	0.2	3	0.4	0.6	10 454
Estimates	0.1155	3.0593	0.3169	1.3101	1106

Using rounding instead the ceiling operator

Another question is that of how the estimation would perform if we replace the upper integer operator ($\lceil \cdot \rceil$) in equations (26) and (31) with the rounding operator ($\lfloor \cdot \rceil$). To analyze this case we use the same parameters as before (see Table 4 for a comparison of the estimates and original values). The resulting demand rate functions for both cases are shown in Figure 5.

As can be seen from the graphs, the change of operator does not have a big impact on problems on larger demand rates. But for smaller demand rates the rounding operator tends to overestimate the demand much more than the ceiling operator. This is a result of the generally ill-conditioned problem of fitting an exponential curve using maximum likelihood.

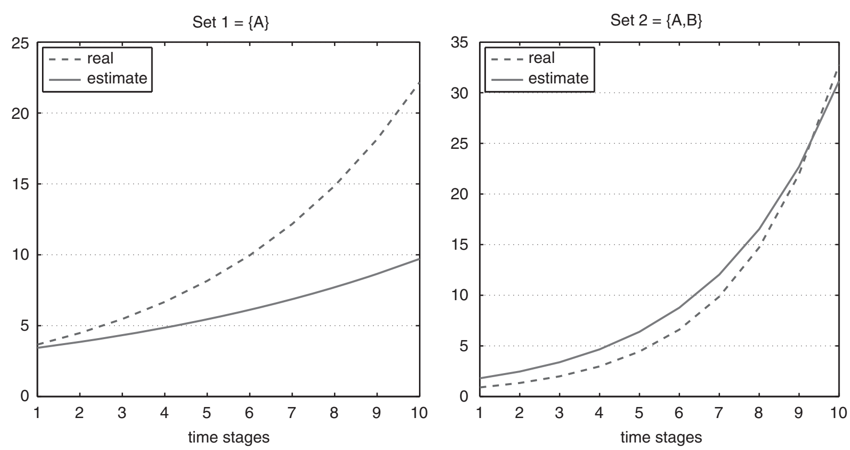

Figure 4: Case 2 – Demand function of real parameters and estimates.

Table 4: Parameter and generated revenue

	α_{Set_1}	β_{Set_1}	α_{Set_2}	β_{Set_2}	*Revenue*
Original	0.1	1	0.2	0.3	1251
Estimates $\lceil \cdot \rceil$	0.1646	0.9249	0.0736	0.9124	1132
Estimates $\lfloor \cdot \rceil$	0.2528	0.7687	0.1742	0.5293	1079
Original	0.2	3	0.4	0.6	10 454
Estimates $\lceil \cdot \rceil$	0.1155	3.0593	0.3169	1.3101	10 134
Estimates $\lfloor \cdot \rceil$	0.1221	2.8343	0.3495	1.0280	10 124

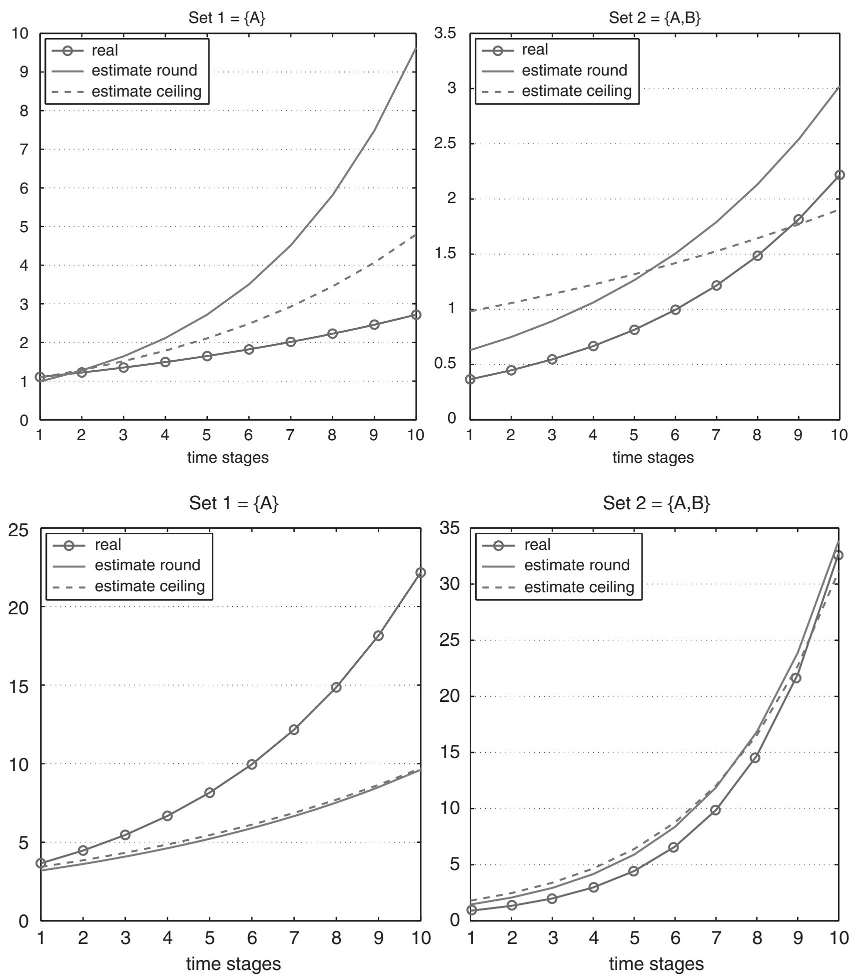

Figure 5: Demand function for smaller (top) and larger (bottom) rates.

The solver is between multiple parameter combinations. With the ceiling operator, small steps of rate changes result in larger steps of realization estimates, and therefore the parameter estimation is more conservative. We therefore suggest using the ceiling operator instead of the rounding operator as described in the previous section.

Table 5: Parameter and generated revenue

	α_{Set_1}	β_{Set_1}	α_{Set_2}	β_{Set_2}	*Revenue*	*% PI Rev*
Original	0.2	3	0.4	0.6	10 454	100
First estimates	0.1155	3.0593	0.3169	1.3101	10 134	97
Second estimates	0.2053	2.7838	0.3429	1.0500	10 442	99

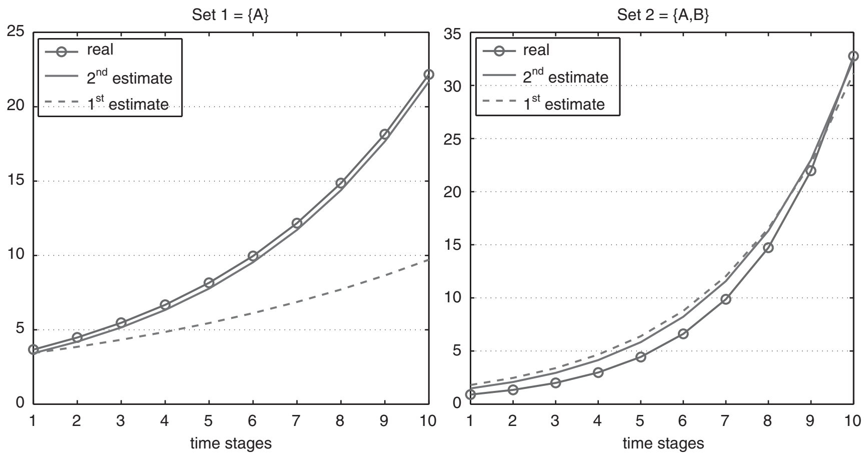

Figure 6: Demand function of real parameters and estimates.

Stepwise improvement

Further, we test whether the estimation can be improved in a stepwise fashion. We compare the 'first' estimates with the 'second'. The first estimates are generated as before, and the second estimates are obtained by the estimation from sales realization generated using the first estimates instead of fixed booking limits. The results are shown in Table 5 and the demand rate functions are shown in Figure 6. Comparing the demand rate functions for Set_2 we observe only a small increase in fitting accuracy. However, for Set_1 the results of the second estimates are almost a perfect fit, compared to the huge underestimation resulting from the first estimates. There is also no loss of accuracy for time stages when A is closed. Using the first estimates results already in 97 per cent of the *PI Rev*, with the second estimates we improve this value to 99 per cent. Thus, we conclude that the estimation method improves under more optimal booking controls, as the information contained in the sales data better reflects the real demand, and therefore does not suffer from the spiral-down effect as discussed in Cooper *et al* (2006), when used in dynamic booking controls.

Multiple choice sets

In reality we have of course more than two choice sets, and thus our final study case consists of three price classes: Economy, Business and Premium, with prices 100, 150 and 200, respectively. The resulting choice sets are $\{E\}$, $\{E, B\}$, $\{B\}$, $\{B, P\}$ and $\{P\}$; the demand parameters are shown in Table 6. Again, we consider a booking horizon of 10 time stages and the initial booking limits are set to 100, 100 and 50 for fare classes E, B and P, respectively. Figure 7 shows the resulting demand rate functions of the first and second estimates compared to the real values. Even the first estimates give a good approximation of the real demand functions. But there is still a big overestimation for set $\{P\}$, and the shape of set $\{E, B\}$ is very poorly approximated. The

Table 6: Real parameters with first (′) and second (″) estimates

Choice set	α	β	α'	β'	α''	β''
{E}	0.1	6	0.0989	4.6706	0.1191	3.8758
{E, B}	0.15	2	0.0334	5.1163	0.1611	2.4389
{B}	0.2	1.5	0.1801	1.8443	0.2089	1.6768
{B, P}	0.25	1	0.1506	2.0707	0.2098	1.6605
{P}	0.3	0.5	0.3784	0.3854	0.3064	0.5213

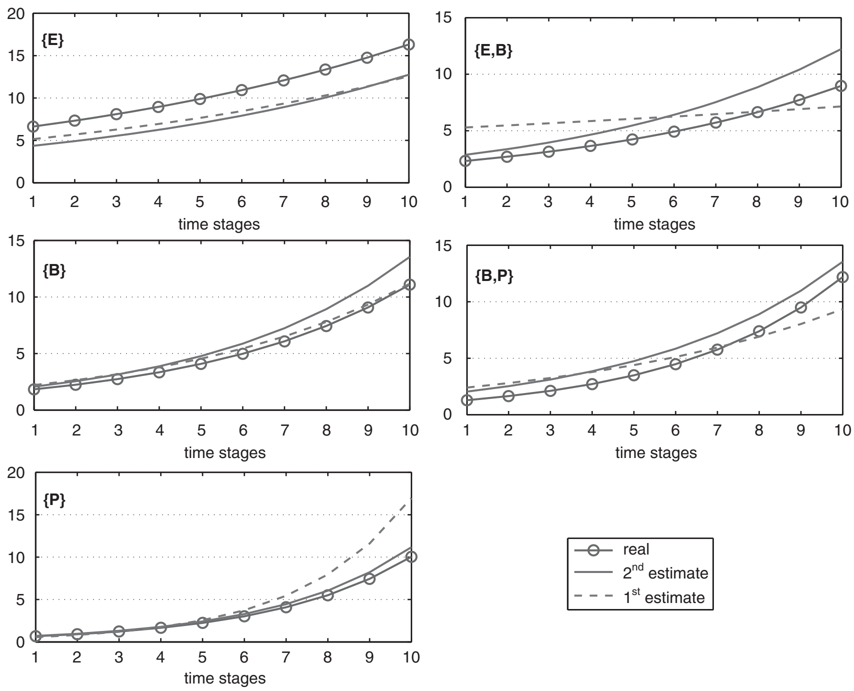

Figure 7: Demand function of real parameters and estimates.

Table 7: Generated revenue using real parameters and estimates

	Sales E	*Sales B*	*Sales P*	*Cap. util.*	*Revenue*	*% PI rev*
Original	92.3	119.5	36.7	248.5	34498	100
First estimates	89.8	119.7	36.7	246.2	34286	99.4
Second estimates	81.2	125	37.5	243.7	34360	99.6

approximation becomes much tighter with the second estimates. The sales and revenue results from using the estimated values in the booking control are shown in Table 7. With the first estimates we have already gained a per cent *PI Rev* of 99.4 per cent, which is increased to 99.6 per cent by using the second estimates. The lower expected revenues are a result of the underestimation of the $\{E\}$ demand and the overestimation of $\{B, P\}$ and $\{P\}$ demand. This also explains the lower capacity utilization (Cap. Util.) by using the estimates.

CONCLUSION

In this article we have stated and analyzed a general method to estimate customer choice set parameters from given sales observations. Previous research papers have shown that there is a huge need in revenue management for a move from product-based demand to customer choice demand. We are convinced that our method will be helpful in achieving this goal. Even though using maximum likelihood to approximate an exponential curve from data points is in general ill conditioned, the resulting demand curves fit the original ones reasonably well. In addition, using the excerpt information to control the future booking process increases the resulting revenue, as well as the accuracy of future estimations. We even observe that using the estimates in the future booking process, we were able to generate revenue results close to optimality.

A further case study to test the presented method on real sales data will follow. Another interesting idea is to develop a more usable optimal allotment control or heuristic based on customer choice behavior. Even for a small setting, the above-stated dynamic programming control is very time-intensive, and therefore not applicable for practical use.

ACKNOWLEDGEMENT

We thank both companies for providing the current sales data sets. Especially we thank the airline company for the starting idea of using choice sets to model the customer's choice behavior.

REFERENCES

Algers, S. and Beser, M. (2001) Modelling choice of flight and booking class – A study using stated preferences and revealed preference data. *International Journal Services and Technology Management* 2(1/2): 28–45.

Andersson, S.E. (1998) Passenger choice analysis for seat capacity control: A pilot study in Scandinavian airlines. *International Transactions in Operations Research* 5(6): 471–486.

Ben-Akiva, M. and Lerman, S.R. (1985) *Discrete Choice Analysis – Theory and Applications to Travel Demand*. Cambridge, MA: MIT Press.

Cooper, W.L., Homen-de-Mello, T. and Kleywegt, A.T. (2006) Models of the spiral-down effect in revenue management. *Operations Research* 54(5): 968–987.

Demidenko, E. (2004) *Mixed Models – Theory and Applications*, Wiley Series in Probability and Statistics, New York: Wiley.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39: 1–38.

Lui, P.H., Smith, S., Orkin, S. and Carey, G. (2002) Estimating unconstrained hotel demand based on censored booking data. *Journal of Revenue and Pricing Management* 1(2): 121–138.

McGill, J.I. (1995) Censored regression analysis of multiclass passenger demand subject to joint capacity constraints. *Annals of Operations Research* 60: 209–240.

Queenan, C.C., Ferguson, M., Kapoor, R. and Higbie, J. (2007) A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management* 16(6): 729–746.

Ratliff, R., Rao, B., Narayan, C. and Yellepeddi, K. (2008) A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management* 7(2): 153–171.

Talluri, K.T. and Van Ryzin, G.R. (2004a) Revenue management under a general discrete choice model of customer behavior. *Management Science* 50(1): 15–33.

Talluri, K.T. and Van Ryzin, G.R. (2004b) *The Theory and Practice of Revenue Management*. New York: Springer.

Van Ryzin, G.R. (2005) Future of revenue management – Models of demand. *Journal of Revenue and Pricing Management* 4(2): 204–210.

Vulcano, G., Van Ryzin, G.R. and Ratliff, R. (2009) Estimating Primary Demand for Substitutable products from Sales Transaction Data. Stern School of Business, New York University. Working Paper, http://pages.stern.nyu.edu/~gvulcano/PrimaryDemandEM-Rev1.pdf.

Weatherford, L.R. and Poelt, S. (2002) Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management* 1(3): 234–254.