OM Practice

Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization

Gustavo Vulcano

Leonard N. Stern School of Business, New York University, New York, New York 10012,
gvulcano@stern.nyu.edu

Garrett van Ryzin

Graduate School of Business, Columbia University, New York, New York 10027,
gjv1@columbia.edu

Wassim Chaar

Research Group, Sabre Holdings, Southlake, Texas 76092, chaar@sabre-holdings.com

Discrete choice models are appealing for airline revenue management (RM) because they offer a means to profitably exploit preferences for attributes such as time of day, routing, brand, and price. They are also good at modeling demand for unrestricted fare class structures, which are widespread throughout the industry. However, there is little empirical research on the practicality and effectiveness of choice-based RM models. Toward this end, we report the results of a study of choice-based RM conducted with a major U.S. airline. Our study had two main objectives: (1) to assess the extent to which choice models can be estimated well using readily available airline data, and (2) to gauge the potential impact that choice-based RM could have on a sample of test markets.

We developed a maximum likelihood estimation algorithm that uses a variation of the expectation-maximization method to account for unobservable data. The procedure was applied to data for a test market from New York City to a destination in Florida. The outputs are promising in terms of the quality of the computed estimates, although a large number of departure instances may be necessary to achieve highly accurate results. These choice model estimates were then used in a simulation study to assess the revenue performance of the EMSR-b (expected marginal seat revenue, version b) capacity control policies and the current controls used by the airline relative to controls optimized to account for choice behavior. Our simulation results show 1%–5% average revenue improvements using choice-based RM. Although such simulated results must be taken with caution, overall our study suggests that choice-based revenue management is both feasible to execute and economically significant in real-world airline environments.

Key words: choice behavior; multinomial logit; EM method; maximum likelihood; capacity control
History: Received: February 2, 2008; accepted: July 8, 2009. Published online in *Articles in Advance* November 13, 2009.

1. Introduction

Quantity-based revenue management (RM) involves optimally allocating capacity of resources to demand by controlling the availability of fare products. This allocation is done dynamically as demand materializes, and with considerable uncertainty about the quantity or composition of future requests.

Traditionally, both researchers and practitioners have made the assumption that demand for each fare class is an independent stochastic process that is not influenced by the firm's availability controls.

This so-called *independent demand model* does not endogenize customer behavior, such as choice behavior and purchase-timing behavior (see Talluri and van Ryzin 2004b for further discussion of the independent demand model). Clearly, this is a somewhat unrealistic assumption. For example, the probability of selling a full-fare ticket may very well depend on whether a discount fare is available at the same time, the probability that a customer buys at all may depend on the lowest available fare, etc. When a customer buys a higher fare when a discount class is closed it is called

a *buy-up*; when she chooses another flight from the same carrier when a discount class is closed, it is called *diversion*. When a customer willing to pay a high fare instead buys a low-fare ticket because it is available, it is called a *buy-down*. These behaviors have important revenue consequences.

Such choice behavior is well recognized in RM practice, and several ad hoc corrections to the independent demand model have been proposed to capture some of these effects (e.g., see Belobaba 1987a, b, Belobaba 1989; Belobaba and Weatherford 1996). The success of low-cost airlines offering simplified, undifferentiated fare structures has rekindled interest in customer choice models, because with minimal rules and restrictions separating fare classes (i.e., in so-called "fenceless" fare environments), customers have considerable latitude to buy up, buy down, or divert. Moreover, point-to-point operators often have many flight departures per day in a given origin-destination (O-D) market, so customers have a variety of departures at different price levels from which to choose. RM decisions can potentially be improved by properly accounting for this environment of flexibility and choice.

Choice-based RM has been an active research area of late. Belobaba and Hopperstad (1999) conducted simulation studies to understand the impact that passenger choice behavior has on traditional RM methods. Talluri and van Ryzin (2004a) provided an exact analysis of the optimal control policy for a single-leg model of RM under a general discrete choice model of demand. Zhang and Cooper (2006) analyzed choice among parallel flights in the same market (e.g., different departure times between the same O-D pair). Their model assumes that the customer chooses within the same fare class among different flights but not between fare classes. They developed bounds and approximations to the resulting dynamic program. Boyd and Kallesen (2004) illustrated the effect of considering demand models for price-sensitive customers in a single-leg setting, where customers are price sensitive and not perfectly segmented, and therefore may buy down. Gallego et al. (2004) and Liu and van Ryzin (2008) studied a deterministic network RM problem using a customer-choice-based linear programming model. The recent dynamic programming (DP) approximation proposal of Zhang and Adelman (2009), the column generation

algorithm of Miranda Bront et al. (2009), and the DP decomposition scheme of Kunnumkal and Topaloglu (2009) also belong to this line of research.

However, there is little empirical understanding of how choice behavior impacts airline RM. That is, how significant is choice behavior in real airline markets? Can it be estimated well using available data? And what do real-world estimates of choice behavior have to say about the potential revenue improvements from using choice-based RM? Exceptions are the work of Andersson (1989, 1998) and Algers and Besser (2001), who reported development efforts at SAS to apply logit choice models to estimate buy-up and recapture factors[1] at one of their hubs. Work close to ours is that of Ratliff et al. (2008), who proposed a heuristic methodology for estimating recapture and demand untruncation for parallel flights in the same market (e.g., different departure times between the same O-D pair).

Here, we report the results of a research study conducted in collaboration with a major U.S. commercial airline. The focus of the study was on O-D markets between the New York metropolitan area and various airports in Florida. The study had two main objectives: the first objective was to assess the feasibility of estimating choice behavior from readily available airline data; the second objective was to use these estimates of choice behavior to assess the potential improvement of using choice-based optimization methods in the targeted test markets.

Toward these ends, in a first phase we developed a maximum likelihood estimation (MLE) algorithm for inferring customer choice behavior from an airline's available operational data—namely, capacity availability, revenue accounting, and flight schedule data.

The first challenge here is constructing relevant choice models. Given a booking instance, one must infer which alternatives the customer was considering when making their purchase decision and the relevant attributes that influenced that customer's decision. This requires determining the consideration set of flight alternatives, the relevant attributes of each alternative, and how these attributes are to be parameterized.

---

[1] *Recapture* is the amount of demand that is retained by the firm's substitute products when a fare class is closed down or the cabin is sold out.

The second problem of estimating customer choice behavior is the available data. Airlines only observe bookings and not shopping behavior; i.e., they only record the outcomes of customers who decide to buy.[2] This means it is impossible to distinguish a small time period without an arrival from a period in which there was an arrival but the arriving customer did not purchase from our airline (i.e., the customer purchased from a competitor or did not purchase at all). One must therefore infer the real, uncensored volume of potential customers that the airline was facing in each market using only purchase data. Ignoring this censoring can cause a severe bias in estimation. Moreover, with this incompleteness in the data, the standard MLE procedure is difficult because of the complexity of the log-likelihood function. To circumvent this problem and account for unobservable no-purchase data, we used a variation of the expectation-maximization (EM) method (see Dempster et al. 1977, or the textbook by McLachlan and Krishnan 1996). The EM procedure we developed is an implementation of the generic framework described in Talluri and van Ryzin (2004a, §5).[3]

To test our estimation method, we first applied it to a set of simulated data. This step was taken to assess the quality of the estimation procedure in an environment where the volume of data and true underlying parameters of the choice behavior are controllable and known. We then applied the estimation procedure to our airline data set. We computed specific estimates of customer preferences for price, arrival time, and departure day. The accuracy of our estimation was validated through the calculation of asymptotic standard errors and χ^2 goodness-of-fit tests between the observed number of bookings and the expected number of bookings predicted by our model. The results of the tests showed the method was quite accurate for the simulated data cases when we generated large volumes of booking records. The quality of the estimates was acceptable, though of lower accuracy, in

the real airline markets that we analyzed. This was due mainly to the more limited number of departure instances in the real data set. Overall, the results indicate that with sufficient data, choice behavior can be reasonably accurately estimated from readily available airline data.

The second phase of the project involved conducting a simulation study based on the estimated choice models. Here, we compared the expected revenue obtainable from the RM controls used by the airline and assessed the potential revenue improvements that could be achieved by optimizing to account for choice behavior effects. To measure this impact, we simulated demand under the estimated choice behavior parameters and then processed that demand under the RM capacity controls used by the airline, EMSR-b (expected marginal seat revenue, version b) controls that we calculated based on the independent demand assumption, and an alternative set of controls computed using our previous algorithmic work on simulation-based optimization, which explicitly accounts for choice behavior (see van Ryzin and Vulcano 2008). Our simulation study showed that the benefits obtained from optimizing RM controls to account for choice behavior are significant: the average revenue gains ranged from 1.4% to 5.3% in the markets we tested. Although in some cases the confidence interval (CI) for the revenue improvements did include zero, the range was always strongly on the positive side. We also tested the robustness of these results by perturbing the parameters to mimic varying degrees of estimation error. The results show that the gains in revenue can be affected by estimation accuracy, particularly in cases where the value added is relatively low (e.g., around 1%). For bigger revenue gap cases, the revenue benefits are preserved even when parameter estimates have errors on the order of 25%. Although these revenue improvements and robustness studies are only simulation estimates, they are based on models fit to real-world data and hence give some sense of the potential performance of choice-based RM.

In summary, our analysis shows that choice behavior and customer preferences for price, flight time, and departure date can be estimated relatively well from readily available airline data, although the quality of these estimates may vary across different markets. The simulated benefit of using choice-based RM

[2] Although the airline's Web data provide more complete shopping information, direct sales through the company's own website only represent a small fraction of total sales. For this reason, we focused exclusively on bookings data.

[3] The EM method has also been used in other RM contexts, in particular by McGill (1995), to estimate multivariate normal demand data with censoring.

methods relative to incumbent, independent-demand RM methods appears quite significant. Together, these results suggest choice-based RM is both a practically feasible and economically significant improvement over current airline RM practice.

The remainder of this paper is organized as follows: For completeness, we begin in §2 by reviewing some general background on choice models and the multinomial logit (MNL) model used in our study. In §3 we specialize the choice model for an airline O-D market and explain our construction of choice sets and how we parameterized the relevant flight attributes. The EM estimation algorithm is described in §4. In §5 we describe the simulation study to assess the RM consequences of embedding choice behavior in capacity control decisions. Finally, we present our concluding remarks in §6.

2. Choice Modeling and Estimation

We first provide an overview of discrete-choice models and then discuss in detail the parametric choice model used in our study.

2.1. General Choice Models

Choice behavior can be modeled by assuming that customers are utility maximizers and individual customer utilities for alternatives are random variables. Specifically, consider a set of alternatives offered by a firm, denoted C. Each customer n has a choice (or consideration) set $C_n \subset C$. We denote by "0" the no-purchase alternative, which is normalized to have utility of zero. Let U_{in} be the utility of customer n for alternative $i \in C_n$. Without loss of generality, we can decompose this utility into two parts, a representative component v_{in} that is deterministic (also called the nominal or expected utility) and a random component ε_{in}. This leads to a utility function

$$U_{in} = v_{in} + \varepsilon_{in}. \qquad (1)$$

The representative component v_{in} is often modeled as a linear-in-parameters combination of observable attributes,

$$v_{in} = \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{in}, \qquad (2)$$

where $\boldsymbol{\beta}$ is an unknown vector of parameters (to be estimated), and $\mathbf{x}_{in}$ is a vector of attributes (explanatory deterministic values) of customer n for alternative flight i such as price paid, arrival time, departure day, etc.

The probability that an individual n selects alternative i from the set $C_n \cup \{0\}$ is given by

$$P_n(i) = \mathbb{P}(U_{in} \geq U_{jn}, \ \forall j \in C_n \cup \{0\}). \qquad (3)$$

2.2. The MNL Model

The most common random utility model, widely used in the economics and marketing, is the MNL model (see Ben-Akiva and Lerman 1994, Chapter 5). It is derived by assuming that the ε_{in}s in the utility functions are independent and identically distributed random variables with a Gumbel (or double-exponential) distribution having cumulative distribution function

$$F(x) = \mathbb{P}(\varepsilon_{in} \leq x) = \exp(-\exp(-\mu(x - \eta))),$$

where η is the location parameter (mode) and μ is a positive scale parameter. The mean and variance of ε_{in}, respectively, are

$$E[\varepsilon_{in}] = \eta + \gamma/\mu \quad \text{and} \quad \text{Var}[\varepsilon_{in}] = \frac{\pi^2}{6\mu^2}, \qquad (4)$$

where $\pi \approx 3.1416$ and γ is the Euler constant (i.e., $\gamma \approx 0.5772$). The Gumbel distribution is used because it is closed under maximization and hence can be applied to (3) (see Gumbel 1958).

For the MNL model, the probability that customer n chooses alternative $i \in C_n$ is given by

$$P_n(i) = \frac{e^{\mu v_{in}}}{\sum_{j \in C_n} e^{\mu v_{jn}} + 1},$$

where the one in the denominator stands for the zero no-purchase utility (i.e., $v_{0n} = 0$) leading to $e^{\mu v_{0n}} = 1$.

Note that in the case of linear-in-parameters utilities (Equation (2)), μ cannot be distinguished from the overall scale of $\boldsymbol{\beta}$. Thus, for convenience, we generally make an arbitrary assumption that $\mu = 1$. Although this is operationally necessary, from (4) we have to keep in mind that we are implicitly assuming equal variance random utilities. For future reference, we set

$$P_n(i) = \frac{e^{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{in}}}{\sum_{j \in C_n} e^{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{jn}} + 1}, \quad i \in C_n. \qquad (5)$$

We let zero denote the set of null alternatives, and let $P_n(0)$ be the probability that customer n does not purchase from our airline; i.e.,

$$P_n(0) = \frac{1}{\sum_{j \in C_n} e^{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_{jn}} + 1}. \qquad (6)$$

Section A1 in the online appendix provides a brief description of a standard approach for estimating parametric choice models: maximum likelihood estimation, which is indeed the approach we followed here.

3. A Choice Model for Airline Markets

We first describe the data available to airlines for estimating customer choice behavior. Then, we describe the design decisions we took in constructing our discrete choice model.

3.1. Airline Data

At our sponsor firm, there were three readily available sources of data to use in estimating customer choice behavior: flight schedule data, revenue accounting data, and availability data. The flight schedule database lists the flights offered by the carrier: flight number, origin/destination, day and time of departure and arrival, aircraft type, etc. The revenue accounting database has one record per ticket issued, i.e., a *customer booking record* or *passenger name record* (PNR) in the airline industry. The relevant fields for our study were ticket number, issue date, coupon number[4] (this field should be the sequence number of the coupon in the ticket), prorate (portion of the total ticket value assigned to a particular coupon, which henceforth will be called simply the *fare*), coupon origin and destination, and flight number. We complete the description of the purchase with information about the arrival time of the flight, which is taken from the flight schedule database.

The availability data are taken from the RM system. It describes the number of available seats for each of the demand classes (or *buckets*) for every flight offered by the airline. There is also a field for the average revenue value corresponding to each bucket.[5] There is one record per flight bucket for different *snapshot*

dates during the booking horizon. Typically, there is daily snapshot data, or one snapshot every two days in cases where the snapshot date is further than one week from the departure date of the flight.

A fourth source of data available for our study was *screen scrape* data. These are data obtained from third parties that automatically sample information about alternatives and fares offered by competitors at different points in time during the booking horizon. This can be obtained by Web crawler programs. Similar data can be obtained via global distributions systems (GDSs). GDSs communicate with the host reservation system of each airline to periodically obtain availability information at the bucket level and hence are in a position to track market-level prices and availability. However, airlines must pay to acquire these GDS data, their request is typically delayed by a month or more, and even with GDS data the problem of estimating customer choice behavior is still a challenge.

Although we obtained screen scrape data providing samples of competitor fares for our study, these data turned out to be quite incomplete and "dirty." In our first attempts at model building and estimation, we incorporated competitor alternatives and attributes in our choice model and estimation procedure to account for competitive effects. Our expectation was that this would improve the accuracy of the choice model. However, our preliminary results yielded very poor quality estimates. A single generic outside alternative produced more stable estimates with lower standard errors and better predictive performance. Hence, we decided to omit these competitive data. Nevertheless, we strongly believe that with more complete and clean data, explicitly incorporating competitive data could add precision to the quality of estimates. A short survey with pointers to different applications of discrete choice models under competitive considerations is provided in Dubé et al. (2002).

3.2. Choice Model Design

We specialized the general choice model described above to our airline setting. One major design decision was the definition of the choice sets C_ns. In our context, C_n includes all the alternatives from our airline that customer n evaluates at the moment of making her purchase decision. Without shopping data, we

---

[4] A *coupon* represents a single flight of a multiflight itinerary. For example, a round-trip ticket based on two nonstop flights has two coupons.

[5] Airline RM systems book reservations in buckets (or fare classes). Each compartment (first, business, and coach) has a number of fare classes—typically 15 or so for coach, 1 or 2 for business, and 1 or 2 for first. Each bucket is used to book tickets sold under different *fare-basis codes*. Each of these fare-basis codes has specific fares associated with them and involves somewhat different requirements, like the number of days of advance purchase.

could not observe C_n directly and therefore had to make a somewhat arbitrary, albeit plausible, definition of the set of relevant alternatives. For our study, and based on our conversations with managers at the sponsor airline, we tried different definitions of C_n, including all flights offered by our airline on a given departure day, all flights offered on a given set of consecutive departure days, and/or all flights offered from different subsets of arriving or departing airports (e.g., all New York metro airports versus JFK airport alone). Because of the sparsity of the airline real data, we decided on the simpler definition of C_n as being all flights on a given day between a specific pair of airports. This choice yielded the highest quality and most intuitive estimates. We used both the schedule and the availability files to build this set, assuming that for some k, the k lowest available buckets (i.e., the k cheapest fares) on each flight are the alternatives considered by an arriving customer.

Although the availability database gives the average revenue per bucket, in general there was a mismatch between these revenue numbers and the fares listed in the revenue accounting database (the sequence of PNRs). Moreover, it is not necessarily true that the fare in the PNR corresponds to the lowest open bucket available at the time of booking.[6] To guarantee consistency within the MNL model, we therefore replaced the true fare paid (taken from the revenue accounting file) by the average revenue reported in the availability file.[7] In this way, we ensured that the fares for the alternatives that are purchased and those that are not purchased are consistent, and hence the choice probabilities in (5) and (6) are well defined and add up to one. Still, this ad hoc fixing of the data introduces some noise that could hurt the estimation procedure.

Another major design decision relates to the time unit of analysis. In the extreme case where the time unit is very small—say, at the level of a second—for a given O-D market, most periods in a day will have

no-purchase outcomes and just a few will have bookings occurring. From our numerical experience, this situation leads to bad behavior in the estimation procedure. We found a good compromise was to partition the day into $T = 140$ small time periods (i.e., each time period corresponds to a 10-minute interval).

For the purpose of representing the arrival time of a flight, a day is split into four time slots: morning (between 5 A.M. and 11 A.M.), noon (between 9 A.M. and 3 P.M.), afternoon (between 1 P.M. and 7 P.M.), and evening (between 5 P.M. and midnight). These intervals overlap, and hence convex weights were assigned to different times slots to represent arrival times that fall in multiple slots. For example, the morning and noon slots overlap from 9 A.M. to 11 A.M. Therefore, a flight arriving at 7 A.M. is a morning flight with weight 1, a flight arriving at 11:30 A.M. is a noon flight with weight 1, a flight arriving at 9:30 A.M. is considered a morning flight with weight 0.75 and a noon flight with weight 0.25, a flight arriving at 10:30 A.M. is considered a morning flight with weight 0.25 and a noon flight with weight 0.75.

We acknowledge that this use of overlapping time slots and "fuzzy" indicator variables is nonstandard from a modeling point. Traditional discrete choice formulation approaches would typically define nonoverlapping slots, leading to four dummy variables, one of which would be eliminated by setting it as a reference alternative (e.g., see Coldren et al. 2003 and Coldren and Koppelman 2005 for characterizations of time of day preferences in airline passenger choice models using dummy variables).[8] However, the overlapping time windows were suggested by the sponsor airline because these are the time frames used to define "morning," "noon," "afternoon," and "evening" flights in the airline's website search engine. We felt, therefore, that they best represented the time differences as presented to customers. We also used the attribute of arrival time for our set of outbound flights from New York to Florida rather than the alternative of departure time. This was also based on discussions with the airline sponsor, who felt time spent in Florida was the primary concern of travelers in these markets.

---

[6] The most consistent approach would be to use the fare-basis code to map a booking request to an open bucket. Unfortunately, according to our experience, this was not straightforward even for the airline research staff itself to unravel this mapping because of the diversity of sales channels, travel agency agreements, etc.

[7] We also tried the reverse substitution, but the results were of worse quality.

[8] Yet an alternative approach is used by Carrier (2008), who defines a time-based cyclical utility for airline passenger choice based on a sine/cosine formulation.

Departure day may be considered as a choice along with arrival time of day, in which case the choice set C_n includes the flights of two or more consecutive days. This is the formulation we used for our simulated data example. The day of a flight is represented by dummy variables so that if the choice set C_n contains days $d_1, d_2, \ldots, d_{C_n}$, then we introduce $d_{C_n} - 1$ dummy variables, setting one reference alternative as the null alternative. Our data covered four days during the peak spring break vacation period, March 23 to March 26, 2005, with d_1 representing March 23 and d_4 representing March 26.

The other attribute included in the model is the base fare paid divided by 1,000 (because of numerical scaling issues). All flights considered in these markets are nonstop flights, and so we did not add an indicator to account for this attribute.

To illustrate this parametrization, consider a market between New York and Florida,[9] with possible departure days d_1 or d_2. The complete specification of the model attributes is provided in Table 1.

For example, consider a flight from New York departing March 23 with arrival time 9:15 A.M. and fare $180. This alternative is represented as a tuple: $\mathbf{x}_{in} = (0.180, 0.875, 0.125, 0, 0, 1)$.

Note that in our discrete choice model we do not include *alternative specific constants*. The reason we omitted alternative specific constants is that the attributes that we consider are sufficient to distinguish a flight, because all flights in a set are operated by the same carrier between the same city pair, and they differ only in terms of arrival time, price, and departure date. Furthermore, when we tried using alternative specific constants, we got very poor quality estimates, most likely because of the increased number of parameters and the relatively low volume of data we had available.

We emphasize here that our choice model assumes a homogeneous market; i.e., customers' preferences are described by a single set of parameters $\boldsymbol{\beta}$ for all customers. According to the airline staff's assessment, this was a reasonable assumption; the New York–Florida market was believed to be relatively homogeneous, at

---

9 In our terminology, *market* refers to particular origin-destination airport combinations, like origin in LaGuardia (LGA) and destination in Palm Beach (PBI).

Table 1   **Choice Behavior: Attribute Definitions**

Attribute	Description	Value
x_{n1}	Base fare	Base fare divided by 1,000
x_{n2}	Morning flight (before 11 A.M.)	Weight as morning arrival
x_{n3}	Noon flight (9 A.M.–3 P.M.)	Weight as noon arrival
x_{n4}	Afternoon flight (1 A.M.–7 P.M.)	Weight as afternoon arrival
x_{n5}	Evening flight (5 P.M.–midnight)	Weight as evening arrival
x_{n6}	Indicator for flying on day d_1	1 if flight departs on day d_1, 0 if it departs on d_2

least with respect to price sensitivity, consisting overwhelmingly of vacation travelers. Still, within these markets the airline could potentially face different customer segments for whom the relative weights of different attributes may vary. In our model-building phase, we in fact tried a multisegment (or *latent class*) MNL model, assuming that customers belong to discrete segments $l = 1, \ldots, L$, for which we estimated both arrival rates λ_l and preference parameters $\boldsymbol{\beta}_l$ (e.g., see Train 2003, Chapter 6). However, as with the competitor data and alternative specific constants, the volume of data proved too limited, and we were not able to obtain good estimates even with $L = 2$ latent segments (note that this doubles the number of model parameters).

Even though our final single-segment MNL model yielded good results (as will be shown in the following sections), it also raises the usual concerns of the MNL model, most noticeably the property of independence from irrelevant alternatives, which implies proportional substitution across alternatives. Briefly, this property establishes that the ratio of probabilities for two alternatives is constant regardless of the consideration set containing them. Other choice models are more flexible with respect to the variety of substitution patterns that they exhibit (e.g., see Train 2003, Chapter 4). Among them, the nested logit model has been widely used in the marketing literature. In our case, we could think of a nested structure where "buy/no buy" is at the top of the hierarchy and the different flight options are under the "buy" branch. Yet the nested logit model can be more challenging to estimate than the basic MNL model because the log-likelihood function is not globally concave.

In summary, exploring alternative choice model specifications within the airline industry context is

certainly worthy of further study. Nevertheless, we believe that the single-segment MNL model used in our project provides a good first-cut sense of the possibilities and benefits of implementing choice-based RM.

4. Estimation

Having defined the consideration sets and the parametrization of the attributes, we can compute the MLE estimates for our model. We first describe the log-likelihood function with incomplete data for our model, where we recall that the incompleteness comes from the fact that only purchase transaction data are available and no-purchase outcomes are unobservable. To overcome this problem, we implement a variation of the EM method that estimates parameters for the utility function (2) and for the arrival rate of customers jointly. Then, we present a preliminary simulated data example to illustrate the quality of the estimates produced by the procedure. Finally, we present two examples based on real airline data.

4.1. The Incomplete Data Log-Likelihood Function

Consider a collection of booking histories, $h = 1, \ldots, H$, representing statistical replicas of a given booking horizon. Each booking horizon consists of B days, and bookings occur for flights departing on any day $1, \ldots, D$. For instance, we could consider the last $D = 5$ days of a month, a booking horizon of $B = 30$ days prior to those D days, and have data for each month during a calendar year, giving us $H = 12$ histories.

We assume a (discretized) Poisson arrival process of customers: Each booking day b is broken up into T small time intervals. In each small time period of day b, an arrival occurs with probability λ. Note that because of the Poisson assumption, and assuming intervals of time are small, there is at most one arrival per period with probability 1. Each arrival selects among the products available on day b (or does not purchase at all) according to the MNL model described by (5) and (6). Let $\mathscr{P}_{b,h}$ denote the set of periods in which customers purchase on day b of booking history h, and let $\overline{\mathscr{P}}_{b,h}$ denote the set of periods in which there are no purchase transactions on day b of booking history h (i.e., $|\mathscr{P}_{b,h}| + |\overline{\mathscr{P}}_{b,h}| = T$).

Consider a linear-in-parameters mean utility of the form $v_{in} = \beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_{in}$. Our objective is to estimate the parameters $\boldsymbol{\theta} = (\beta_0, \boldsymbol{\beta}, \lambda)$ from purchase data, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$. As discussed above, given only purchase data it is impossible to distinguish a period without an arrival from a period with an arrival but where the arriving customer chooses the no-purchase alternative 0 (i.e., she bought from the competitor or did not buy at all). The incomplete data log-likelihood function for our model is given by

$$\log \mathscr{L}(\mathbf{x}, \boldsymbol{\theta})$$
$$= \sum_{h=1}^{H}\sum_{b=1}^{B}\sum_{i \in \mathscr{P}_{b,h}} \left[\log\lambda + \beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i - \log\left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_j} + 1\right)\right]$$
$$+ \sum_{b=1}^{B}\left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}|\right)\log\left(\lambda\left(\frac{1}{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_j} + 1}\right) + (1-\lambda)\right).$$
$$\tag{7}$$

Unfortunately, this function is difficult to maximize directly because of the complexity of the second term above. To overcome this problem, we used the EM method of Dempster et al. (1977).

4.2. The EM Method

The EM method operates on the complete data log-likelihood function, which has a simpler form than (7) and is constructed assuming that all the arrivals, purchases, and nonpurchases can be observed. Specifically, for the periods in $\overline{\mathscr{P}}_{b,h}$, let $a_b = 1$ if there is an arrival in an observation period of day b, and $a_b = 0$ otherwise. Note that $a_b = 1$ accounts for an arrival that either buys from our competitors or decides not to buy at all.[10]

We can now write the complete data log-likelihood function:

$$\log \mathscr{L}(\mathbf{x}, \boldsymbol{\theta})$$
$$= \sum_{h=1}^{H}\sum_{b=1}^{B}\sum_{i \in \mathscr{P}_{b,h}} \left[\log\lambda + \beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_i - \log\left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T}\mathbf{x}_j} + 1\right)\right]$$

---

[10] For notational simplicity, we omit the fact that each booking day b is itself divided into many smaller time periods, so in fact there is an indicator a_{bt} for each interval t of day b. Because this extra set of indices does not change the form of the likelihood function, to minimize the complexity of the displayed equations, we omit them.

$$+ \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) a_b \left(\log \lambda - \log \left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_j} + 1 \right) \right)$$

$$+ \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) (1 - a_b) \log(1 - \lambda). \qquad (8)$$

The first line of (8) in accounts for the observed bookings, and there is one term for every record in the revenue accounting database for each booking history. The second line accounts for customers that either buy from competitors or do not purchase, and the third line refers to periods with no arrivals. These last two are unobserved data.

The broad strategy of the EM method consists of starting with arbitrary initial estimates $\hat{\boldsymbol{\theta}} \equiv (\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})$ of the parameters, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_K)$. Then we use these estimates to compute the conditional expected value of $\log \mathscr{L}(\mathbf{x}, \boldsymbol{\theta})$, $E[\log \mathscr{L}(\mathbf{x}, \boldsymbol{\theta}) \mid \mathbf{x}, \hat{\boldsymbol{\theta}}]$ (the expectation step, or E-step). Effectively, this replaces the missing data (the indicators a_b for the no-purchase time periods) by their expected values conditioned on the current estimates. We then maximize the resulting log-likelihood function (which has the same form as the complete log-likelihood function) to generate new estimates $\hat{\boldsymbol{\theta}}$ (the maximization step, or M-step). These steps are repeated until the procedure converges. We next describe these steps in complete detail.

4.2.1. The E-Step. In the E-step of each iteration, the unknown data are the values a_b in the second and third lines of (8), for all b, corresponding to periods with nonpurchase from our airline. However, given current estimates $\hat{\boldsymbol{\theta}} = (\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})$, we can determine the expected values of these indicators. Specifically, let $P_b(i)$ be the probability that a customer arriving in a small period of day b chooses alternative i (irrespective of the booking history). By the Bayesian formula, we get for $b = 1, \ldots, B$, irrespective of the booking horizon h:

$$\hat{a}_b := E[a_b \mid t \in \overline{\mathscr{P}}_b, \hat{\boldsymbol{\theta}}]$$

$$= \mathbb{P}(a_b = 1 \mid t \in \overline{\mathscr{P}}_b, \hat{\boldsymbol{\theta}})$$

$$= \frac{\mathbb{P}(t \in \overline{\mathscr{P}}_b \mid a_b = 1, \hat{\boldsymbol{\theta}}) \mathbb{P}(a_b = 1 \mid \hat{\boldsymbol{\theta}})}{\mathbb{P}(t \in \overline{\mathscr{P}}_b \mid \hat{\boldsymbol{\theta}})}$$

$$= \frac{P_b(0 \mid \hat{\beta}_0, \hat{\boldsymbol{\beta}}) \hat{\lambda}}{\hat{\lambda} P_b(0 \mid \hat{\beta}_0, \hat{\boldsymbol{\beta}}) + (1 - \hat{\lambda})},$$

where, from (6),

$$P_b(0 \mid \hat{\boldsymbol{\beta}}) = \frac{1}{\sum_{j \in C_b} e^{\hat{\beta}_0 + \hat{\boldsymbol{\beta}}^\mathsf{T} \mathbf{x}_j} + 1}.$$

Observe that the probability $P_b(0 \mid \hat{\boldsymbol{\beta}})$ for the nonobservable periods does not depend on a particular booking history h, $h = 1, \ldots, H$, because the choice sets for a given booking day b are the same for all h.

Substituting $\hat{a}_b$ into (8), we obtain the expected, complete data log-likelihood function

$$Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}) := E[\log \mathscr{L}(\mathbf{x}, \boldsymbol{\theta}) \mid \mathbf{x}, \hat{\boldsymbol{\theta}}]$$

$$= \sum_{h=1}^{H} \sum_{b=1}^{B} \sum_{i \in \mathscr{P}_{b,h}} \left[\beta_0 + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_i - \log \left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_j} + 1 \right) \right]$$

$$+ \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\mathscr{P}_{b,h}| \right) \log \lambda + \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right)$$

$$\cdot \hat{a}_b \left(\log \lambda - \log \left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_j} + 1 \right) \right)$$

$$+ \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) (1 - \hat{a}_b) \log(1 - \lambda). \qquad (9)$$

This function is relatively easy to maximize, as shown next.

4.2.2. The M-Step. We next determine a maximizer, $\hat{\boldsymbol{\theta}}^*$, of the expected log-likelihood function (9). Note that the function (9) is separable in λ and $(\beta_0, \boldsymbol{\beta})$. Define the function $F(\lambda)$ as

$$F(\lambda) := \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\mathscr{P}_{b,h}| \right) \log \lambda + \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) \hat{a}_b \log \lambda$$

$$+ \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) (1 - \hat{a}_b) \log(1 - \lambda).$$

Taking the derivative with respect to λ_b, we get

$$\frac{\partial F(\lambda)}{\partial \lambda} = \sum_{b=1}^{B} \sum_{h=1}^{H} |\mathscr{P}_{b,h}| \frac{1}{\lambda} + \sum_{b=1}^{B} \sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \hat{a}_b \frac{1}{\lambda}$$

$$- \sum_{b=1}^{B} \sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| (1 - \hat{a}_b) \frac{1}{1 - \lambda}.$$

Setting this derivative equal to zero, we obtain the updated estimate

$$\hat{\lambda}^* = \frac{\sum_{b=1}^{B} \sum_{h=1}^{H} |\mathscr{P}_{b,h}| + \sum_{b=1}^{B} \sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \hat{a}_b}{B \times H \times T}.$$

This says that our current best estimate of λ is the number of observed purchases for flights booked on any day b across the H booking histories, $\sum_{b=1}^{B} \sum_{h=1}^{H} |\mathscr{P}_{b,h}|$, plus the estimated number of arrivals from unobservable periods on any day b across all the booking histories, $\sum_{b=1}^{B} \sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \hat{a}_b$, divided by the total number of periods under consideration, $B \times H \times T$. The function $F(\lambda)$ is concave in λ, and hence the critical point $\hat{\lambda}^*$ is a well-defined global maximizer of F.

Next, we maximize the terms in (9) containing $(\beta_0, \boldsymbol{\beta})$ to obtain the updated estimates $(\hat{\beta}_0^*, \hat{\boldsymbol{\beta}}^*)$. Define the function $G(\beta_0, \boldsymbol{\beta})$ as

$$
\begin{aligned}
G(\beta_0, \boldsymbol{\beta}) \\
:= \sum_{h=1}^{H} \sum_{b=1}^{B} \sum_{i \in \mathscr{P}_{b,h}} \left[\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_i - \log\left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1 \right) \right] \\
- \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) \hat{a}_b \log\left(\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1 \right). \quad (10)
\end{aligned}
$$

To find a critical point $(\hat{\beta}_0^*, \hat{\boldsymbol{\beta}}^*)$ of $G(\beta_0, \boldsymbol{\beta})$, we solve this nonlinear optimization problem by using the conjugate gradient method. The partial derivative of $G(\beta_0, \boldsymbol{\beta})$ with respect to β_0 is

$$
\begin{aligned}
\frac{\partial G(\beta_0, \boldsymbol{\beta})}{\partial \beta_0} = \sum_{h=1}^{H} \sum_{b=1}^{B} \sum_{i \in \mathscr{P}_{b,h}} \left(1 - \frac{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j}}{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1} \right) \\
- \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) \hat{a}_b \frac{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j}}{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1}. \quad (11)
\end{aligned}
$$

For all $k = 1, \dots, K$, we compute the partial derivatives of $G(\beta_0, \boldsymbol{\beta})$:

$$
\begin{aligned}
\frac{\partial G(\beta_0, \boldsymbol{\beta})}{\partial \beta_k} = \sum_{h=1}^{H} \sum_{b=1}^{B} \sum_{i \in \mathscr{P}_{b,h}} \left(x_{i,k} - \frac{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} x_{j,k}}{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1} \right) \\
- \sum_{b=1}^{B} \left(\sum_{h=1}^{H} |\overline{\mathscr{P}}_{b,h}| \right) \hat{a}_b \frac{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} x_{j,k}}{\sum_{j \in C_b} e^{\beta_0 + \boldsymbol{\beta}^{\mathsf{T}} \mathbf{x}_j} + 1}. \quad (12)
\end{aligned}
$$

Although the function $G(\beta_0, \boldsymbol{\beta})$ is data dependent, it is jointly concave for linear-in-parameters utilities under weak regularity conditions on the data (see Train 2003, p. 65; McFadden 1974).

4.2.3. Convergence and Implementation. To check for convergence, in each iteration of the algorithm we compare the value of the expected log-likelihood function (9) before and after maximizing with respect to λ and $(\beta_0, \boldsymbol{\beta})$. We also check for the norm of the difference between two consecutive sets of these estimates. Thus, if $\boldsymbol{\theta}_{(i)} = (\beta_0^{(i)}, \boldsymbol{\beta}^{(i)}, \lambda^{(i)})$ is the optimal set of parameters from iteration i, we check if $\|Q(\boldsymbol{\theta}_{(i+1)} \mid \hat{\boldsymbol{\theta}}_{(i)}) - Q(\boldsymbol{\theta}_{(i)} \mid \hat{\boldsymbol{\theta}}_{(i)})\| < \epsilon$, or if $\|\boldsymbol{\theta}_{(i+1)} - \boldsymbol{\theta}_{(i)}\| < \epsilon$, for some small $\epsilon > 0$. We also set a maximum of 300 iterations for the EM method.

Because the expected log-likelihood function (9) is continuous in both $(\beta_0, \boldsymbol{\beta}, \lambda)$ and $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})$,[11] then Theorem 2 in Wu (1983) guarantees that if the sequence of estimates converges, the resulting value will be a stationary point of the incomplete data log-likelihood function (7). Whether the sequence diverges or converges to something other than the global maximum is more difficult to determine. In practice, the EM method has proved to be a robust and efficient way to compute maximum likelihood estimates for incomplete data problems (e.g., see McLachlan and Krishnan 1996, Chapter 3, for further discussions on convergence properties of the EM method).

Although the standard implementation of the EM method is computationally intensive because it makes a pass through all of the available data in every iteration, in our experience it is computationally feasible with the volume of data that a real airline may handle during a booking period of few weeks for a single O-D pair and for a small number of departure days.

Our implementation of the estimation algorithm was developed under Windows XP and coded in C++ using the Microsoft Visual C++ 6.0 compiler and the STL (Standard Template Library) as a source of basic classes and containers. To optimize the function (10), we selected from the library Numerical Recipes in C++ the Broyden-Fletcher-Goldfarb-Shanno variant of the Davidon-Fletcher-Powell minimization algorithm, a quasi-Newton-type method that requires computing (11) and (12).

4.3. Preliminary Estimation Case: Example 0
We first applied our EM method over a set of simulated data. We did this as an initial test to get a

---

[11] Note that the expected log-likelihood function (9) is also a function of $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})$ through $\hat{a}_b$.

sense of how many data are necessary to get good estimates and how closely those estimates match the (known) parameters that generated the data. Given a known underlying MNL choice model (i.e., given values for λ and $\boldsymbol{\beta}$) for a fixed O-D market, we used Monte Carlo simulation to generate $H = 100$ streams of data for a booking horizon of $B = 10$ days for flights departing on any day d_1, d_2, or d_3. We set $\lambda = 0.3$ and used the MNL probabilities (5) and (6) to determine the product chosen by each arrival. We have around 42,000 arrivals throughout the generated streams, out of which around 40,300 booked a flight. The data about the flights offered were taken from both the schedule and the availability files for the market under consideration. In this preliminary example, we assume that the customer considers the lowest available bucket of each flight in her consideration set (i.e., following the notation in §3.2, we set $k = 1$).

The parameters $\boldsymbol{\beta}$ used for this base case are given in Table 2 and were based on estimates obtained from the real airline data set. Descriptive statistics of the data used for this example can be found in Table A1 in the online appendix.

To determine the mean utility (2) for a given flight, we need one parameter β_j for every attribute value $\mathbf{x}_{in}$. However, we modified this parameter definition. Observe that both the set of attributes for the time slots and the set for the departure days add up to one, leading to a degree of freedom in the $\boldsymbol{\beta}$ parameters and hence to the identifiability problem (i.e., the inability to fully recover the original parameters from the estimated ones). In symbols, for customer n, given $\mathbf{x}_{in}$ and $\boldsymbol{\beta}$, and noting that

$$x_{in2} + x_{in3} + x_{in4} + x_{in5} = 1 \quad \text{and} \quad x_{in6} + x_{in7} + x_{in8} = 1,$$

we would have

$$
\begin{aligned}
v_{in} &= \beta_1 x_{in1} + \beta_2 x_{in2} + \cdots + \beta_8 x_{in8} \\
&= \beta_1 x_{in1} + (\beta_2 - \beta_5) x_{in2} + (\beta_3 - \beta_5) x_{in3} + (\beta_4 - \beta_5) x_{in4} \\
&\quad + \beta_5 + (\beta_6 - \beta_8) x_{in6} + (\beta_7 - \beta_8) x_{in7} + \beta_8.
\end{aligned} \tag{13}
$$

By defining $\beta_0 = \beta_5 + \beta_8$ and relabeling the parameters to have a consecutive numbering, we can equivalently consider a mean utility of the form $v_{in} = \beta_0 + \boldsymbol{\beta}^\mathsf{T} \mathbf{x}_{in}$ for the different purchase options (recall that the no-purchase option has been normalized to have zero mean utility). Here, we are choosing as reference variables both *evening flight* and *departure day* d_3. In general, one of the criteria for picking the reference variables is that there is at least one alternative in every choice set that matches it (e.g., in our instance this would mean that in every booking day there is an evening flight offered with departure day d_3). This is indeed the case in our data. We refer to β_0 as the *base utility*.

We applied the EM method starting from estimates $\hat{\beta}_0 = 1$ and $\hat{\boldsymbol{\beta}}$, $\hat{\lambda}$ set equal to zero. The output after reindexing the $\boldsymbol{\beta}$ parameters is included in Table 3.

The third column in this table reports the true parameter values, following the adjustment described in (13). The fourth column reports the estimates computed by the EM method. The fifth column reports the percentage bias between the estimated and true values,[12] followed by the asymptotic standard error (ASE) of the corresponding estimate (see §A5 in the online appendix for further details). Note that for all the coefficients except the indicator for morning flight, we can reject the null hypothesis that the true value is zero at the 0.01 significance level.[13] However, the true value of the indicator for morning flight is indeed zero. The number of iterations performed in the EM method in this case was 36, taking 15 minutes of computing time. Given the identifiability problem inherent to the data, it is not possible

Table 2 Input β Values for Example 0

Attribute	Description	Value
β_1	Base fare	−1.0
β_2	Morning flight (before 11 A.M.)	0.5
β_3	Noon flight (9 A.M.–3 P.M.)	0.7
β_4	Afternoon flight (1 A.M.–7 P.M.)	0.3
β_5	Evening flight (5 P.M.–midnight)	0.5
β_6	Indicator for flying on day d_1	0.6
β_7	Indicator for flying on day d_2	0.3
β_8	Indicator for flying on day d_3	0.9

[12] Note the results suggest an apparent bias in the estimates, which is not unexpected because the MLE is only asymptotically unbiased.

[13] The quasi-*t* statistic is computed as the ratio between the estimated value of the parameter and the ASE. Recall that for a two-tailed test, the critical values of this statistic are ±1.65, ±1.96, and ±2.58 for the 0.10, 0.05, and 0.01 significance levels, respectively.

Table 3   **Output Parameters for Example 0**

Parameter	Description	True value	Est. value	Bias (%)	ASE	*t*-stat
$\hat{\beta}_0$	Base utility for any purchase option	1.4	1.5881	13.43	0.0131	121.34
$\hat{\beta}_1$	Base fare	−1.0	−1.2265	22.65	0.0644	−19.06
$\hat{\beta}_2$	Morning flight (before 11 A.M.)	0.0	0.0169		0.0273	0.62
$\hat{\beta}_3$	Noon flight (9 A.M.–3 P.M.)	0.2	0.2376	18.80	0.0128	18.60
$\hat{\beta}_4$	Afternoon flight (1 P.M.–7 P.M.)	−0.2	−0.2178	8.89	0.0132	−16.39
$\hat{\beta}_5$	Indicator for flying on day d_1	−0.3	−0.2715	−9.51	0.0155	−17.51
$\hat{\beta}_6$	Indicator for flying on day d_2	−0.6	−0.5756	−4.06	0.0126	−45.62
$\hat{\lambda}$	Arrival rate	0.3	0.2988	−0.41	2.3E−05	13,229.40

to distinguish how much of the value $\hat{\beta}_0$ is driven by each reference (*evening slot* versus *day* d_3). However, the quality of the estimates is quite good, as indicated by the last three columns of Table 3. In the online appendix (§A3), we include another simulated example that does not suffer from this identifiability problem, although the bias of the estimates is higher.

Figure 1 shows a measure of the goodness of fit of our estimates for Example 0 for the first five days of the booking horizon under consideration. For a given pair (b, d) (i.e., booking day b, departure day d), and recalling that $P_n(i)$ is the probability that customer n

chooses flight i, we have that $\lambda \times P_n(i)$ is the Poisson rate of arrivals per unit of time on booking day b that choose flight i departing on day d. Therefore,

$$E[\text{number of bookings on day } b \text{ for flight } i$$
$$\text{departing on day } d] = T \times \lambda \times P_n(i),$$

where $T = 140$ is the number of booking periods per day. Based on the true parameters of Table 2, we can compute the true expected number of bookings and utilities to be observed from a population of customers that choose according to them. Based on the estimated parameters of Table 3, we can compute the

Figure 1   **Goodness of Fit for the Bookings per Combination** (b, d) **for Booking Days** $b = 1, \ldots, 5$ **in Example 0**
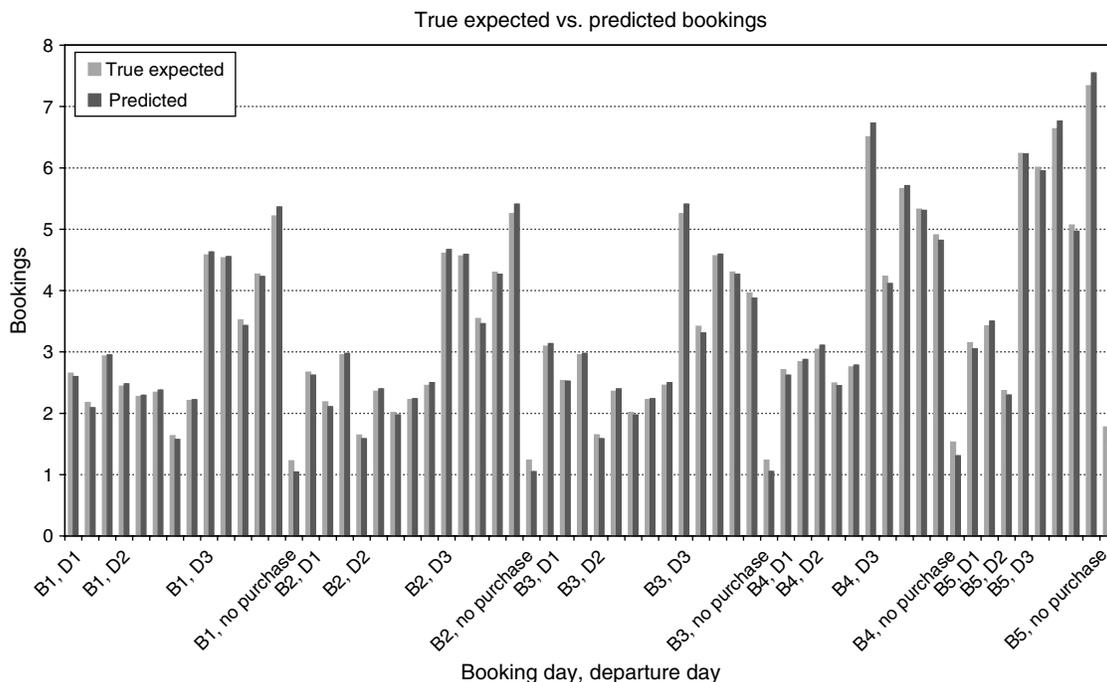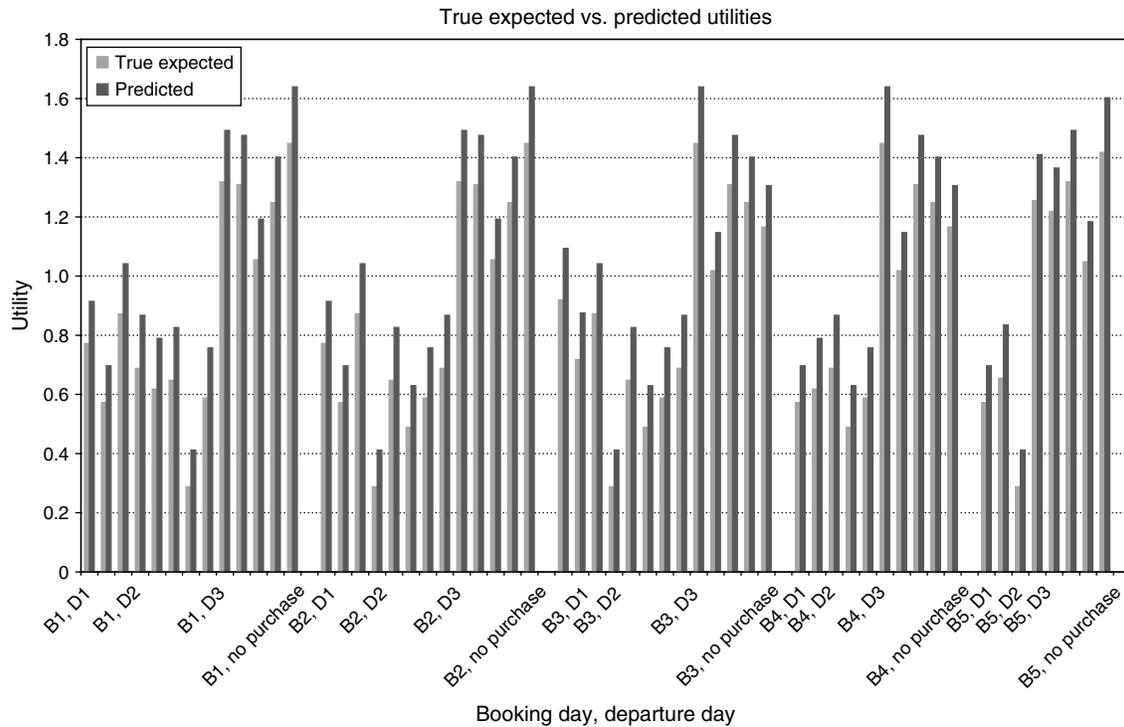
Figure 2   Goodness of Fit for the Utilities per Combination (b, d) **for Booking Days** $b = 1, \ldots, 5$ **in Example 0**

Booking day, departure day

number of bookings and utilities predicted by the output of the EM method. In Figure 1, for every possible combination (b, d) there are several tick marks, each one representing a flight. The graph shows an excellent quality of fit; clearly, the difference between the average number of observable bookings and the number predicted by the EM method is negligible.

We then aggregated the observed number of bookings per flight across the entire booking horizon and the expected number of bookings predicted by the EM method for the 14 flights contained in the three departure days under consideration (see Table A2 in the online appendix). A χ^2-test gives a p-value $= 1$, giving a very strong justification not to reject the null hypothesis that the observed bookings indeed follow the distribution estimated via the EM procedure.

Figure 2 shows another measure of the goodness of fit; in this case, we compared the true expected utilities and the mean utilities predicted by our EM procedure. Here, even though the estimated utilities are slightly lower than the original utilities, the shift preserves their relative order. Again, this reflects a bias in the component estimates, even though the resulting

total predicted sales are quite accurate. Moreover, we note that multiple values $(\hat{\beta}_0, \hat{\boldsymbol{\beta}}, \hat{\lambda})$ may produce the same probabilities of sale as observed by Talluri and van Ryzin (2004a, §5). In such cases, the EM method finds only one such pair. In the online appendix (Figure A2.1) we show the comparison between the true and predicted MNL choice probabilities for different alternatives (including the no-purchase), where we also found a very good quality of fit.

We next studied the sensitivity of the procedure to different starting points. For example, starting from $\hat{\beta}_0 = 1$, $\hat{\boldsymbol{\beta}} = 0$, and two different arrival rates (the true value of the arrival rate $\hat{\lambda} = 0.3$, and $\hat{\lambda} = 0.6$), we obtain high-quality results, similar to the case when the initial $\hat{\lambda} = 0$. Details are provided in Table A3 in the online appendix.

We then checked the behavior of the EM method when the model does not capture all parameters of the utility function (e.g., there are omitted variables in the specification of the model). Toward this end, taking again the data generated based on the description in Table 2 with $\lambda = 0.3$, we assume that customers choose based only on the fare and the departure days,

Table 4   **Estimated Parameters for Example 0 When Customers Choose Based on Fare and Departure Day Only**

Parameter	Description	True value	Est. value	Bias (%)	ASE	*t*-stat
$\hat{\beta}_0$	Base utility for any purchase option	0.9	1.6207	80.07	0.0133	122.27
$\hat{\beta}_1$	Base fare	−1.0	−0.6232	−37.68	0.0627	−9.94
$\hat{\beta}_5$	Indicator for flying on d_1	−0.3	−0.4283	42.76	0.0150	−28.50
$\hat{\beta}_6$	Indicator for flying on d_2	−0.6	−0.6823	13.72	0.0126	−54.25
$\hat{\lambda}$	Arrival rate	0.3	0.2977	−0.76	0.0000	13,182.51

and we omit the fact that the flight arrival time also factors into their total utility. Table 4 shows the results obtained from this misspecified model. Note that for all the coefficients we can reject the null hypothesis that the true value is zero at the 0.01 significance level. Another measure of the goodness of fit is given by comparing the values of the incomplete data log-likelihood function evaluated at the original parameter values and at the estimated parameters, which are −176,323 and −171,629, respectively. Running a χ^2-test of the observed bookings versus the expected number of bookings predicted by the estimated model for the 14 flights under consideration gives strong support for the estimated model with a p-value = 0.93.

Next, we assumed that customer utilities are based only on the fare and the flight arrival times (ignoring flight departure day) and estimated this misspecified model based on booking data for the $D = 3$ booking days under consideration. The results are reported in Table 5. Note that for all the coefficients (including the indicator for morning flight), we can reject the null hypothesis that the true value is zero at the 0.01 significance level. In this case, using (13), we can recover the four time coordinates of $\hat{\beta}$ from the value of $\hat{\beta}_0$. However, in this case, the quality of the estimates is rather poor, with very pronounced biases.

Running a χ^2-test of the observed bookings versus the number of bookings predicted by our model for the 14 flights under consideration, we get a small p-value of 0.08. This case shows the importance of defining the customers' consideration sets; accounting for different departure days is important.

We also tested how the volume of data affects the quality of estimates from our procedure. We did this by changing the value of the arrival rate λ. We could also do this by changing the number of histories, H, or the booking periods per day, T. However, varying λ gives a better sense of the impact of the relative mix of periods with observations and nonobservations in the sample. Fixing $H = 100$, $T = 140$, and $\boldsymbol{\beta}$ as in Example 0, and taking $\lambda = 0.01$, we obtain rather poor estimates with large biases and small t-statistics (see Table 6). This is we have many periods with no observations and only a few observable purchases. In this case, the total number of bookings across the 100 histories is around 1,400, and this does not seem sufficient to obtain good estimates.

When we increase the value of the arrival rate to $\lambda = 0.05$, which corresponds to around 7,000 recorded bookings, we start getting better estimation results, and for an even larger value $\lambda = 0.7$ (corresponding to around 98,000 arrivals), we obtain high-quality estimates (even better than the base case in Table 3 for $\lambda = 0.3$). Running a χ^2-test of the expected number

Table 5   **Estimated Parameters for Example 0 When Customers Choose Based on Fare and Flight Times**

Parameter	Description	True value	Est. value	Bias (%)	ASE	*t*-stat
$\hat{\beta}_0$	Base utility	0.5	1.6138	222.76	0.0131	123.43
$\hat{\beta}_1$	Base fare	−1.0	−2.3879	138.79	0.0607	−39.31
$\hat{\beta}_2$	Morning	0.0	0.1841		0.0274	6.72
$\hat{\beta}_3$	Noon	0.2	0.4030	101.48	0.0126	31.94
$\hat{\beta}_4$	Afternoon	−0.2	−0.1694	−15.30	0.0134	−12.68
$\hat{\lambda}$	Arrival rate	0.3	0.2982	−0.59	0.0000	13,205.42

Table 6   **Estimated Parameters for Example 0 Under Different Arrival Rates**

Param.	Description	True value	Arrival rate $\lambda = 0.01$			Arrival rate $\lambda = 0.70$		
			Estimated value	Bias (%)	t-stat	Estimated value	Bias (%)	t-stat
$\hat{\beta}_0$	Base utility	1.40	14.3635	925.96	0.00	1.5905	13.61	198.29
$\hat{\beta}_1$	Base fare	−1.00	−0.5423	−45.77	−1.13	−1.0761	7.61	−28.79
$\hat{\beta}_2$	Morning	0.00	−0.1608		−0.89	−0.0161		−0.98
$\hat{\beta}_3$	Noon	0.20	0.1549	−22.56	1.66	0.2194	9.71	29.32
$\hat{\beta}_4$	Afternoon	−0.20	−0.2357	17.83	−3.07	−0.2045	2.27	−24.09
$\hat{\beta}_5$	Indicator for d_1	−0.30	−0.3333	11.11	−3.58	−0.2965	−1.17	−30.55
$\hat{\beta}_6$	Indicator for d_2	−0.60	−0.5868	−2.20	−7.71	−0.6160	2.67	−79.24
$\hat{\lambda}$	Arrival rate	0.01	0.0098	−2.14	434.80	0.6915	−1.22	30,618.87

of bookings predicted by our model in the $\lambda = 0.7$ case versus the observed number of bookings for the 14 flights under consideration, we see there is very strong support for the estimated model, with p-value $= 1$. Overall, what matters for the quality of the estimates is the volume of data available and the definition of the base period. In the extreme case where we have infinitely many small periods T, the probability λ of having an arrival per period would be very small, and the quality of the estimates would be very poor, leading also to some potential numerical problems in the estimation process. Having longer base periods increases the probability of arrival, but one must be careful to keep the probability of having two or more arrivals per period small.

4.4. Estimation Results for Airline Data

We next present the results of the EM method applied to an O-D market in our airline data set for two different departure dates. We describe the estimation for each departure date in turn. Table A6 in the online appendix reports descriptive statistics for each market.

4.4.1. Example 1. In this market, we consider a booking horizon of $B = 39$ days, with data from February 6 to March 17, 2005 (i.e., we take just a single history $H = 1$), with 549 reservations. The market is defined by an origin airport in New York and a destination airport in Florida. We consider flights for one departure day: March 25, 2005. There is a total of five flights on this departure day in this market. When building the consideration set C_n, we assume that each customer takes into account the five lowest

available fares (i.e., buckets) of each flight. That is, for each flight in the consideration set, we take the five lowest available fares from the availability file (i.e., we set $k = 5$ in terms of the notation introduced at the beginning of §3.2). Recall that a flight's time is defined as its arrival time in Florida.

The initial set of estimates for the EM method was $\hat{\beta}_0 = \hat{\boldsymbol{\beta}} = 0$ and $\hat{\lambda} = 0.1$. The EM method reached the maximum of 300 iterations for this case in 208 seconds of computational time.[14] The norm difference between the last two sets of estimates was $\|\boldsymbol{\theta}_{(300)} - \boldsymbol{\theta}_{(299)}\| = 0.002$.

In Table 7, the three columns under Example 1 give a summary of the final set of estimates. Note that we cannot reject the null hypothesis that the true value of the morning flight indicator is zero. The χ^2-test of the total number of observed bookings versus the expected number of bookings per flight shows a reasonably high p-value $= 0.41$, providing good support for the estimated distribution of sales.

A few aspects of the model summarized in Table 7 are also worth noting. First, in terms of the linear-in-parameters utility described in (2), the price coefficient has the right (negative) sign. Second, it is interesting to interpret the relative value of the coefficients as indicators of the sensitivity of the choices. For instance, according to the attributes described in Table A6 in the online appendix, Flight 3 is a noon flight and Flight 4 is a 0.092 morning/0.908

---

[14] The computational times reported here were obtained with an Intel Pentium IV, 2.4 GHz, and 512 MB of RAM. Although undoubtedly more sophisticated nonlinear solvers and platforms could be used, this setup proved to be sufficient for our study.

Table 7   Estimated Parameters for Examples 1 and 2

Parameter	Description	Example 1			Example 2		
		Value	ASE	*t*-stat	Value	ASE	*t*-stat
$\hat{\beta}_0$	Base utility	−2.0566	0.1185	−17.3619	−2.3468	0.0888	−26.4284
$\hat{\beta}_1$	Base fare	−1.7348	0.2148	−8.0745	−1.9175	0.1601	−11.9746
$\hat{\beta}_2$	Morning flight (5 A.M.–11 A.M.)	1.0861	1.2384	0.8770	3.3114	1.3590	2.4366
$\hat{\beta}_3$	Noon flight (9 A.M.–3 P.M.)	1.3056	0.1465	8.9133	0.2851	0.1284	2.2203
$\hat{\beta}_4$	Afternoon flight (1 P.M.–7 P.M.)	0.7751	0.1538	5.0406	−0.3186	0.1387	−2.2976
$\hat{\lambda}$	Arrival rate	0.1927	0.0011	177.2174	0.3485	0.0011	322.7402

noon flight. The mean utilities for this flights are, respectively,

$$v_{3n} = -2.0566 - 1.7348p_3 + 1.3056 \quad \text{and}$$

$$v_{4n} = -2.0566 - 1.7348p_4 + 1.0861 \times 0.092$$
$$+ 1.3056 \times 0.908$$
$$= -2.0566 - 1.7348p_4 + 1.2854,$$

where p_i is the price charged for Flight i, $i = 3, 4$. Then, $v_{4n} \geq v_{3n}$ when $p_4 \leq p_3 - 0.012$. Recalling that the scale of the prices is in thousands, a discount of \$12 over the current price p_3 will make Flight 4 a more attractive alternative (on average), indicating a rather similar willingness to pay for both flights. Similarly, and just as another example, we can compare the mean utilities of Flights 4 (a late morning flight) and 5 (an early evening flight):

$$v_{4n} = -2.0566 - 1.7348p_4 + 1.0861 \times 0.092$$
$$+ 1.3056 \times 0.908$$
$$= -0.7712 - 1.7348p_4,$$

$$v_{5n} = -2.0566 - 1.7348p_5 + 0.7751 \times 0.675$$
$$= -1.5334 - 1.7348p_5.$$

Then, $v_{5n} \geq v_{4n}$ when $p_5 \leq p_4 - 0.439$. Thus, a discount of \$439 over the current price p_4 will make Flight 5 an alternative more attractive than Flight 4, on average. This type of analysis could be useful for the airline to determine price strategies, or even flight schedule strategies, to influence the revenue performances and load factors of the different flights.

Columns for Example 1 in Table 8 provide the calculation of the market shares based on the estimated parameters and the linear-in-parameters utilities (2)

for the five flights under consideration, assuming that the fare is set at the intermediate point of the fare range open during the booking horizon.[15] Although this decision is arbitrary, the numbers obtained give a sense of the explanatory power of the MNL parameters when compared with the relative frequencies of the observed bookings.

4.4.2.   Example 2.   For the same market as above, we next considered a booking horizon of $B = 39$ days (between February 7 and March 18, 2005). Here, we considered flights for March 26, 2005. There is a total of five flights in this market on this day with a total of 506 recorded bookings. Like in Example 1, when building the consideration set C_n, we assume that each customer takes into account the five lowest available fares of each flight. Again, a flight's time is defined as its arrival time in Florida.

The initial set of estimates for the EM method was $\hat{\beta}_0 = \hat{\boldsymbol{\beta}} = 0$ and $\hat{\lambda} = 0.1$. The EM method also reached the maximum of 300 iterations for this case. The norm difference between the last two sets of estimates was $\|\boldsymbol{\theta}_{(300)} - \boldsymbol{\theta}_{(299)}\| = 0.0022$. It took 193 seconds of computational time to run the procedure.

In Table 7, the three columns under Example 2 give a summary of the final coefficient estimates. Note that we can reject the null hypothesis that the true value of the parameters is zero at the 0.05 confidence level for all estimates, implying that the attributes under consideration are relevant for the sake of capturing choice behavior. The χ^2-test of the total number of observed bookings versus the expected number of bookings per

---

[15] See descriptive statistics in Table A6 in the online appendix. For instance, for Flight 1, the range of fares for the 63 observed bookings is $[57, 212]$, so we assume that the fare is set at $(57 + 212)/2 = 134.50$.

Table 8   Estimated Market Shares for Examples 1 and 2

	Example 1		Example 2	
Flight	Utility-based market share	Booking-based market share	Utility-based market share	Booking-based market share
1	0.110	0.115	0.203	0.213
2	0.239	0.237	0.214	0.223
3	0.285	0.260	0.155	0.170
4	0.294	0.260	0.264	0.251
5	0.073	0.128	0.164	0.142

flight shows a good p-value $= 0.50$, confirming the quality of the estimated distribution.

In this example, the price coefficient again has the right (negative) sign. We could also perform sensitivity calculations as the ones presented for Example 1. Regarding the consistency of the predicted market share with the booking frequencies, we see in columns under Example 2 in Table 8 that, once again, the predicted shares based on the estimated parameters are remarkably close to the observed bookings.

5.   Assessing RM Improvements
As mentioned above, the second phase of our study used the choice model estimates fit from the test markets to assess the potential improvement in revenues from using choice-based RM optimization methods. Toward this end, we used the model and optimization algorithm developed in van Ryzin and Vulcano (2008). In this model, we assume the firm controls the availability of products using a nested, protection-level-based, capacity control strategy. Using a simulation-based optimization algorithm, we calculate sample path gradients of the network revenue function with respect to the protection levels. The gradients are then used sequentially to construct a stochastic steepest ascent method. Starting from an initial set of protection levels (possibly computed under the traditional independent demand model assumption), the algorithm is guaranteed to converge (in probability) to a stationary point of the expected revenue function under mild conditions.

The input data to the optimization algorithm are provided by a *choice model simulator* module that generates streams of demand based on customers' preferences described by the choice model estimated from our test markets. Every arriving customer is

Figure 3   An Integrative Estimation/Optimization Approach for Choice-Based RM
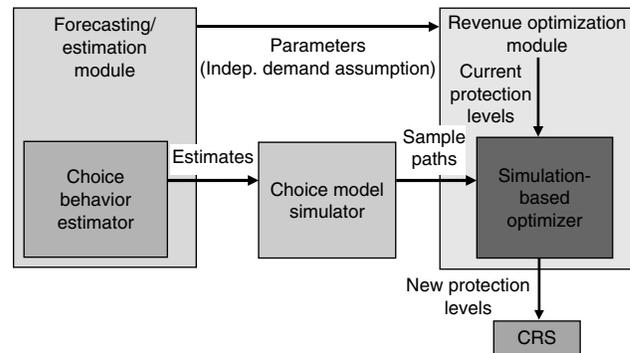
characterized by a preferred list of products, built randomly using the utilities based on the MNL parameters. Figure 3 sketches the integration of the whole estimation/optimization approach.

5.1.   Simulating Customer Choice Behavior
Once we have the estimates that characterize customer choice behavior from our EM method, we can use them to simulate the arrival of customer requests to the central reservation system (CRS) as follows: On booking day b, customers arrive in accordance with a Poisson process with rate λ per unit of time. Hence, we generate a Poisson random variable with mean $140 \times \lambda$ to have an instance of the number of arrivals in day b.

Next, for each generated arrival n, we build a preference list based on the utilities described in §2.1 as follows: First, we construct the set of alternatives C_n as we did in the estimation phase.[16] Then, for each alternative $i \in C_n$, we compute the utility U_{in} in (1) by taking the vector of attributes $\mathbf{x}_{in}$ corresponding to this alternative, and by simulating the random term ε_{in}. We use the standard inverse function method for the latter; that is, for a Unif(0,1) value u, we compute

$$\varepsilon_{in} = -\log(-\log u) - \gamma.$$

We also simulate the random noise ε_{0n} for the no-purchase alternative. Recalling that the mean no-purchase utility has been normalized to zero, we

---

[16] Recall that the consideration set C_n is built primarily upon the schedule and the availability files. The difference from the estimation phase is that when simulating we do not take information from the revenue accounting file.

rank our airline's alternatives with utility in excess of ε_{0n}. This ordered set of alternatives constitute customer n's preference list.

This is repeated for all days in the booking horizon under analysis, following the sequence $b = 1, \ldots, B$ to build one stream of arrivals. The sequence of arrivals is saved in a file. For simplicity, we assume single-unit demand for all customers.

5.2. Optimizing RM Controls

The last stage in our analysis is to assess the performance of RM controls based on the traditional independent demand assumption relative to those incorporating choice behavior effects. To this end, we ran simulation tests using Examples 1 and 2 in §4.4. We tested the performance of our procedure based on two sets of initial protections levels[17]: one based on the current controls implemented by the airline, and the other using single-leg, EMSR-b controls (e.g., see Talluri and van Ryzin 2004b, §2.2, for a definition and discussion of EMSR-b). For the former, we used the "availability" file provided by the airline and constructed an initial set of nested protection levels taking the highest values of the controls for each class observed in the file during the booking horizon under consideration, preserving the nested property of the protection levels. For the EMSR-b controls, we first computed statistics for the independent demands of the different products. To do so, we took the simulated streams of demand and retained the first element of each customer's preference list. In this way, we obtained the mean and variance for the uncensored demand of each product (i.e., of each flight-bucket combination). Then, following the usual practice, we assumed normal distribution of each product's demand and computed nested protection levels using EMSR-b. We compared these two sets of protection levels with the improved ones that we obtained when applying the choice-based stochastic gradient (SG) algorithm described in van Ryzin and Vulcano (2008). We emphasize here that the protection levels provided by the airline were calibrated in an ad hoc way, because of some dirtiness in the

---

[17] A protection level y_i is defined as the number of seats reserved for classes i and higher. The labeling of the classes assumes that class 1 is the highest (i.e., most expensive fare) one, followed by class 2, class 3, etc.

data and the fact that their value changed during the booking horizon. In this sense, the protection levels given by EMSR-b constitute a more sensible "null model."

For the examples below, we simulated 2,000 streams of customer arrivals. Every arrival is specified as a "customer type," characterized by a particular preference list (i.e., two customers that have the same preference list are considered the same type). In each example, there are five parallel flights and 80 products. According to the sponsor airline policy, there are 16 buckets per flight (i.e., 15 protection levels per leg).

After generating the new set of protection levels, we checked the revenue obtained with the original and new protection levels over 2,000 simulated streams of arrivals. We report expected revenues under both protection levels policies, 95% confidence intervals for the revenue gap and corresponding (*network*) *load factors*, defined as the average of the *leg load factors*, i.e., the ratio between average number of seats sold on leg i and its initial capacity c_i.

We ran the numerical experiments under Windows XP, with a CPU Intel Core Duo of 2.0 GHz, and 1 GB of RAM.

5.2.1. Example 1. This is the example described in § 4.4.1. Across the 2,000 streams of demand, there are, on average, 841 arrivals per stream that have one of our flights as the first choice, with a total of 592,605 customer types. The initial capacities available are still significant (between 76 and 155 seats).

First, we ran the stochastic gradient algorithm over the set of protection levels provided by the airline. Comparing the revenue obtained from the protection levels improved by the algorithm and the original ones, we observed an expected revenue gap of 3.04%, with a 95% CI of $(-0.12\%, 6.21\%)$. Even though this confidence interval technically includes zero, it is strongly on the positive side, and the improvement is significant at slightly less than the 95% level. We also observe a small increase in the load factor from 0.68 to 0.70. However, as we explained above, because these protection levels were built in an approximate way from the airline's availability file, we place greater significance on the values given by the EMSR-b heuristic.

Table 9 reports the original EMSR-b and updated sets of protection levels obtained by the stochastic

Table 9   **Protection Levels for Example 1: EMSR-b vs. SG**

Flight	Capacity c_i	Stage	Protection levels y_i										
			1	2	3	4	5	6	7	8	9	10	11
1	136	Original	3	8	9	10	12	16	20	39	62	83	105
		Updated	0	0	0	17	17	17	136	136	136	136	136
2	155	Original	9	26	56	91	126	155	155	155	155	155	155
		Updated	9	10	147	155	155	155	155	155	155	155	155
3	152	Original	49	96	119	134	145	152	152	152	152	152	152
		Updated	34	152	152	152	152	152	152	152	152	152	152
4	150	Original	48	86	102	112	113	115	128	138	149	150	150
		Updated	80	83	83	83	83	89	125	138	149	150	150
5	76	Original	25	66	76	76	76	76	76	76	76	76	76
		Updated	42	76	76	76	76	76	76	76	76	76	76

gradient algorithm. Even though according to the airline system there were 16 virtual classes in the original flights, we report the protection levels of the open classes (all the remaining classes were closed for the five flights, and remained closed after the application of the algorithm). Revenue-wise, the stochastic gradient algorithm led to a revenue improvement of 1.42%, with a 95% CI for the revenue gap of $(-2.47\%, 5.30\%)$, suggesting the improvement in this case is not very statistically significant. It took just 15 seconds to compute the new set of protection levels. As we described in our previous paper (see van Ryzin and Vulcano 2008, §4), the impact of the algorithm over different buckets is not unidirectional; that is, some classes become more protected and others become less. Some classes are collapsed (e.g., classes 1–4 in Flight 1), and others are closed (e.g., classes 8–12 in Flight 1). In this case, the load factor significantly decreased from 0.90 to 0.75, meaning that the new protection levels dramatically reduced the number of tickets sold but improved the mix of passengers by forcing several buy-ups (e.g., note the increment in y_1 for Flights 4 and 5) and rejecting several formerly accepted low-fare tickets.

The above test assumes the choice-based RM optimization uses the true parameter values of the utility function. However, in reality there will always be estimation error as well as specification error. We were interested in how such errors in estimation effect the revenue results. For example, one might posit that choice-based RM only "works" if one has highly accurate estimates of the parameters of the choice model.

To assess the robustness of our combined estimation and optimization procedure with respect to estimation noise, we perturbed the original $\hat{\lambda}$ and $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$ estimates provided by our estimation algorithm, by random multiplicative factors in the range $\pm 10\%$, $\pm 25\%$, and $\pm 50\%$. Then, we simulated 2,000 streams of demand based on each perturbed set of estimates, and, starting from the original set of EMSR-b reported in Table 9, we computed the corresponding improved set of (*noisy*) protection levels using the stochastic gradient algorithm. Finally, to assess the impact of the random noise in the estimates, we processed the original streams of demand (i.e., the demand streams generated under the "true" $\hat{\lambda}$ and $(\hat{\beta}_0, \hat{\boldsymbol{\beta}})$) and compared the revenues obtained with the noisy protection levels to the revenue obtained using the original set of EMSR-b protection levels. Note that given the number of bookings for the different flights described in Table A6 in the online appendix (between 63 and 143), these perturbation of the estimated choice parameters can indeed have an impact on the distribution of passengers across different flights. The results are summarized in Table A7 in the online appendix, showing the outcome of five different perturbations in each of the ranges $\pm 10\%$, $\pm 25\%$, and $\pm 50\%$. Considering the relatively narrow, original expected revenue gap of 1.42%, we see that even with a 10% perturbation, the additional revenues become negative in two of the five cases.

5.2.2.   Example 2. This is the example described in §4.4.2. Each of the five parallel flights has 16 classes. Each of the 2,000 demand streams has an average of

1,097 arrivals that have one of our flights as the first choice, and produce a total of 351,654 customer types. The capacities available vary between 106 and 135.

As in Example 1, first we ran the stochastic gradient algorithm over the set of protection levels provided by the airline. Comparing the revenues obtained from both sets of protection levels, we observed a expected revenue gap of 2.49%, with a 95% CI of $(-0.52\%, 5.50\%)$. Again, even though the confidence interval includes the zero, it is strongly on the positive side and significant at slightly less than 95%. We also observe an increase in the load factor from 0.82 to 0.87. However, as explained above, because these protection levels were built in an approximate way, we also consider the protection levels given by the EMSR-b heuristic.

Table 10 reports the original EMSR-b and the corresponding updated sets of protection levels obtained by the stochastic gradient algorithm. Even though there were 16 buckets in the original flights, we report the protection levels of the open classes (all the remaining classes were closed for the five flights, and remained closed after the application of the algorithm). It took 70 seconds to compute the new set of protection levels. For this market, the stochastic gradient algorithm led to a revenue improvement of 5.30%, with a 95% CI of $(2.28\%, 8.32\%)$. The initial and improved sets of protection levels are shown in Table 10. In contrast to Example 1, here the load factor significantly increases from 0.82 to 0.89, suggesting that part of the increase in revenues is

explained by an increased sales volume. This is particularly noticeable in Flights 1 and 4, where protection levels y_1 to y_5 were significatively reduced. This phenomenon occurs jointly with a change of the passenger mix that is clear in Flight 3, which now focuses exclusively on passengers for Buckets 1 and 2, and in Flight 5, which closes Bucket 4 but opens more the highest three classes.

Even though the expected revenue gap of 5.30% might seem very high by RM standards, it is not surprising when compared to other results reported in the literature. For instance, Liu and van Ryzin (2008, §7.1) report revenue improvements between 0.10% and 7% in 18 of 20 parallel flight cases when accounting for choice behavior effects versus traditional methods. Also for a parallel flight case, Miranda Bront et al. (2009) report revenue improvements of approximately 10% for similar load factor scenarios. We also observed significant gains for simulated parallel flight networks in our previous paper (see Examples 2, 3, and 4 in van Ryzin and Vulcano 2008, §4.)

We again tested the robustness of our estimation/optimization procedure with respect to estimation noise. In this case, given the large initial expected revenue gap of 5.30%, there is more room for estimation error. As shown in Table A8 in the online appendix, the choice-based protection levels still produce significant revenue gains with up to 25% errors in the parameter estimates.

6. Conclusions

Our analysis shows that it is feasible to estimate choice behavior and customer preferences for price, arrival time, and departure dates from readily available airline data. The main caveat is that the quality of these estimates may vary across markets, and they improve significantly if one has access to a large number of historical booking records. Having customer-level shopping data would, in our view, provide for even more robust estimates of choice effects.

Our simulation study suggests that the benefits that could be obtained from optimizing RM controls to account for choice behavior are significant: the revenue gains were between 1.4% and 5.3% in the markets we tested. For the cases of relatively low revenue improvements, the accuracy of the choice model

Table 10 **Protection Levels for Example 2: EMSR-b vs. SG**

Flight	Capacity c_i	Stage	Protection levels y_i						
			1	2	3	4	5	6	7
1	118	Original	27	34	43	51	56	79	102
		Updated	6	9	10	36	40	118	118
2	122	Original	6	47	85	122	122	122	122
		Updated	1	4	122	122	122	122	122
3	106	Original	7	40	70	97	106	106	106
		Updated	0	106	106	106	106	106	106
4	135	Original	48	76	87	98	109	130	135
		Updated	15	65	65	65	68	130	135
5	121	Original	22	61	101	121	121	121	121
		Updated	5	37	121	121	121	121	121

estimates was important for providing these benefits. When gains are higher (i.e., around 5% revenue lift), the benefit of choice-based RM does not require highly accurate estimates of the underlying choice model parameters. Although these are only simulation estimates and rely on the fact that demand follows the model exactly (i.e., no model specification error), they give some sense of the potential for improvement using choice-based RM. More testing with industry-wide data of both the estimation and optimization procedures are needed to confirm these initial results. We have worked out the basic technical building blocks for doing so in this paper.

Several extensions of this work warrant attention. It would be desirable to further test a model with unobservable (or *latent*) segments within the population (e.g., business and leisure segments) so that customer choice behavior can be predicted more accurately. However, as we discussed in §3.2, more complex models appear to require more data to estimate accurately. Another extension that could prove useful when dealing with a large volume of data is the sampling-based implementation of the EM method proposed by Jank (2005), which consists of intelligently incrementing the number of observations needed to compute the log-likelihood function toward the end of the algorithm instead of using all the available data in each iteration of the EM method. Finally, our choice model assumes that customers' valuations for different attributes do not depend on the offer set. It thus ignores framing or reference effects. Accounting for such behavioral biases would be worth further study.

Electronic Companion

An electronic companion to this paper is available on the *Manufacturing & Service Operations Management* website (http://msom.pubs.informs.org/ecompanion.html).

Acknowledgments

References

Algers, S., M. Besser. 2001. Modeling choice of flight and booking class: A study using stated preference and revealed preference data. *Internat. J. Services Tech. Management* **2** 28–45.

Andersson, S.-E. 1989. Operational planning in airline business—Can science improve efficiency? Experiences from SAS. *Eur. J. Oper. Res.* **43**(1) 3–12.

Andersson, S.-E. 1998. Passenger choice analysis for seat capacity control: A pilot project in Scandinavian airlines. *Internat. Trans. Oper. Res* **5**(6) 471–486.

Belobaba, P. P. 1987a. Air travel demand and airline seat inventory management. Ph.D. thesis, Flight Transportation Laboratory, MIT, Cambridge, MA.

Belobaba, P. P. 1987b. Airline yield management: An overview of seat inventory control. *Transportation Sci.* **21**(2) 63–73.

Belobaba, P. P. 1989. Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* **37**(2) 183–197.

Belobaba, P. P., C. Hopperstad. 1999. Boeing/MIT simulation study: PODS results update. *Proc. 1999 AGIFORS Reservations and Yield Management Study Group Sympos., London.*

Belobaba, P. P., L. R. Weatherford. 1996. Comparing decision rules that incorporate customer diversion in perishable asset revenue management situations. *Decision Sci.* **27**(2) 343–363.

Ben-Akiva, M., S. Lerman. 1994. *Discrete Choice Analysis: Theory and Applications to Travel Demand,* 6th ed. The MIT Press, Cambridge, MA.

Boyd, E. A., R. Kallesen. 2004. Practice papers: The science of revenue management when passengers purchase the lowest available fare. *J. Revenue Pricing Management* **3**(2) 171–177.

Carrier, E. 2008. Modeling the choice of an airline itinerary and fare product using booking and seat availability data. Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge.

Coldren, G., F. Koppelman. 2005. Modeling the competition among air-travel itinerary shares: GEV model development. *Transportation Res. Part A* **39**(4) 345–365.

Coldren, G. M., F. S. Koppelman, K. Kasturirangan, A. Mukherjee. 2003. Modeling aggregate air travel itinerary shares: Logit model development at a major U.S. airline. *J. Air Transport Management* **9**(6) 361–369.

Dempster, A. P., N. M. Laird, D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B* **39**(1) 1–38.

Dubé, J.-P., P. Chintagunta, A. Petrin, B. Bronnenberg, P. B. Seetharaman, K. Sudhir, R. Thomadsen, Y. Zhao. 2002. Structural applications of the discrete choice model. *Marketing Lett.* **13**(3) 207–220.

Gallego, G., G. Iyengar, R. Phillips, A. Dubey. 2004. Managing flexible products on a network. Report TR-2004-01, Computational Optimization Research Center. Department of Industrial Engineering and Operations Research, Columbia University, New York.

Gumbel, E. J. 1958. *Statistics of extremes.* Columbia University Press, New York.

Jank, W. 2005. Fast and Eficient Model-Based Clustering with the Ascent-EM Algorithm. B. Golden, S. Raghavan, E. Wasil, eds. *The Next Wave in Computing, Optimization, and Decision Technologies.* Springer, New York, 201–212.

Kunnumkal, S., H. Topaloglu. 2009. A new dynamic programming decomposition method for the network revenue management problem with customer choice behavior. *Production Oper. Management*. Forthcoming.

Liu, Q., G. van Ryzin. 2008. On the choice-based linear programming model for network revenue management. *Manufacturing Service Oper. Management* **10**(2) 288–310.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. P. Zarembka, eds. *Frontiers in Econometrics*. Academic Press, New York, 105–142.

McGill, J. I. 1995. Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints. *Ann. Oper. Res.* **60**(1) 209–240.

McLachlan, G., T. Krishnan. 1996. *The EM Algorithm and Extensions*. John Wiley & Sons, New York.

Miranda Bront, J., I. Méndez-Díaz, G. Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Oper. Res.* **57**(3) 769–784.

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

Talluri, K. T., G. J., van Ryzin. 2004a. Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* **50**(1) 15–33.

Talluri, K., G. J. van Ryzin. 2004b. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, New York.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

van Ryzin, G. J., G. Vulcano. 2008. Computing virtual nesting controls for network revenue management under customer choice behavior. *Manufacturing Service Oper. Management* **10**(3) 448–467.

Wu, C. F. J. 1983. On the convergence properties of the EM algorithm. *Ann. Statist.* **11**(1) 95–103.

Zhang, D., D. Adelman. 2009. An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Sci.* **43**(3) 381–394.

Zhang, D., W. L. Cooper. 2006. Revenue management for parallel flights with customer-choice behavior. *Oper. Res.* **53**(3) 415–431.