Coordinating Overbooking and Capacity Control Decisions on a Network

By Itir Karaesmen and Garrett van Ryzin

June 20th, 2004
Coordinating Overbooking and Capacity Control Decisions on a Network

Itir Karaesmen • Garrett van Ryzin

R.H. Smith School of Business, University of Maryland, College Park, MD 20742, USA
Graduate School of Business, Columbia University, New York, NY 10025, USA
itir@umd.edu • gjv1@columbia.edu

June 20, 2004

Overbooking and capacity control are two central problems in revenue management. Roughly, overbooking determines a “virtual capacity” on each flight – a capacity in excess of the physical capacity to hedge against cancellations and no-shows – while capacity control determines how this virtual capacity is allocated to various itineraries and fare classes. In practice, these problems are typically solved using separate models, despite the fact that the two sets of decisions are quite interrelated. While methods have been proposed for combining overbooking and seat allocation decisions on a single flight leg, few techniques have been proposed for networks. In this paper, we propose a generic method for coordinating these two sets of decisions in general networks. While the method applies to a wide range of model types, we illustrate it for the commonly used deterministic linear programming and economic overbooking models. We analyze structural properties of the method and show that a control policy derived from it is asymptotically optimal when sales volumes and capacities are large. We also provide a computationally efficient algorithm to determine the optimal policy parameters. Finally, we provide a numerical study of the method’s performance relative to other ad hoc approaches representative of those used in practice. While our method is not uniformly better than other approaches in all cases, the examples suggest that it yields more consistent and robust performance.

1. Introduction

Revenue management is viewed by many as among the most significant management science and operations research applications (Bell (1998), Talluri and van Ryzin (2004)). In a survey
paper, McGill and van Ryzin (1999) identified overbooking, capacity control, pricing and forecasting to be four main research areas in revenue management. Research in these areas progressed almost independently until very recently - exceptions include joint pricing and capacity control problems studied by Bitran and Caldentey (2003) and Côté et al. (2003).

Overbooking is the practice of selling more seats than an airline has physical capacity to provide. This is done as a hedge against the uncertainty that accepted reservations may cancel prior to departure of a flight or may become no-shows at the time of departure. Airlines set overbooking levels in order to balance the opportunity cost of empty seats and the costs of denied service at the time of departure (the overbooking costs). Beckmann (1958), Thompson (1961), Simon (1968), Rothstein (1971, 1974) are among the earlier papers addressing how to make such overbooking decisions. The amount by which an airline overbooks its flight is called the overbooking pad and the capacity plus the overbooking pad is called the virtual capacity of the flight. While overbooking is a somewhat mature area in terms of methodology, it is nevertheless regarded as among the most economically important functions of revenue management. For example, Smith et al. (1992) at American Airlines estimate that 15% of seats on sold-out flights would be lost if overbooking were not practiced and that the benefit of overbooking at American in 1990 exceeded $225 million.

Another critical problem in revenue management is capacity (or seat inventory) control, a practice which grew out of the deregulation of the U.S. airline industry in 1978. To compete against the low fares offered by new entrants, major airlines introduced a variety of discounted fares offered with advance-purchase, Saturday-night-stay, non-refundability and other restrictions. But to prevent potential revenue losses, airlines had to carefully control how many seats they allocated to these discounted fares. Thus, revenue management practice broadened in this post-deregulation period to incorporate capacity control methodology, which focuses on how to optimally allocate capacity to differentiated classes of demand. Many articles on capacity control were published in the operations research literature beginning in this time period, including Littlewood (1972), Belobaba (1987, 1989), Pfeifer (1989), Curry (1990), Brumelle and McGill (1993), Lee and Hersh (1993) and Robinson (1995) to name a few.

Collectively, capacity control and overbooking practices have produced dramatic improvements in revenues in the airline industry (see Weatherford and Bodily (1992), Alstrup et al. (1989) and Smith et al. (1992)). The hotel/hospitality industry (see Yeoman and Ingold (1997)) has widely adopted revenue management techniques as well, as have many
other industries. Indeed, while our focus is on airline applications, the problem and methods we investigate apply to a range of other network revenue management contexts, for example length-of-stay controls in hotels. (See Talluri and van Ryzin (2004) for a discussion of other industry applications of network models.)

Our focus in this paper is not overbooking or capacity control methods per se, but rather on the relationship between these two sets of decisions. The virtual capacities that are computed by overbooking models are critical inputs to capacity control models. Conversely, capacity control models provide estimates of the opportunity cost of an unsold seat, which is a critical input to overbooking models. Thus, the overall performance of a revenue management system is fundamentally affected by the interaction of its overbooking and capacity-control models.

Our aim is to investigate how to coordinate these two models. Specifically, we propose a simple heuristic method which is based on two key approximations: 1) the airline pre-commits to virtual capacities at the start of the booking process (e.g. it uses static overbooking limits), and 2) denied service costs are approximated by assuming that the overbooking limits are always reached (e.g. every leg “sells out” its entire capacity). This decomposes the net revenue into two terms – a network revenue function and an overbooking cost function – both of which are functions of the virtual capacities. Then, the coordinated overbooking and capacity control problem becomes a two-stage optimization problem. In the first stage, the optimal virtual capacities are determined based on the overbooking costs and the network revenues, where the network revenues are computed in the second-stage given the virtual capacities. In this way, the problem decomposes into an overbooking and a capacity control problem, where information from the capacity control problem in the second stage is a vital input for the overbooking problem in the first stage, and vice versa.

This two-stage method has advantages because it applies to a wide variety of seat allocation and overbooking models. However, we focus on the case where capacity control is performed using the deterministic linear programming (DLP) model and overbooking is based on an economic model, because these are quite familiar models that are used commonly in practice. We show that policies based on this special case are in fact asymptotically optimal in a fluid scaling of the problem similar to Cooper (2002), Gallego and van Ryzin (1997) and Talluri and van Ryzin (1999a). Finally, we propose an efficient algorithm to determine the virtual capacities in our two-stage method. A numerical study shows that our method provides more consistent and robust performance than traditional, ad hoc methods.
The rest of the paper is organized as follows: We first briefly review the relevant literature in Section 2. We provide an overview the DLP model for seat inventory control and the economic overbooking model in Section 3. We then define a general model of the joint capacity control and overbooking problem and define our generic two-stage approximation method in Section 4. In Section 5, we apply our two-stage approximation to combine the DLP method with the economic model of overbooking. Detailed analysis of this special case - including structural results, asymptotic behavior of a policy obtained by using DLP - are presented in the same section. Finally, we propose an algorithm to solve the special case in Section 6 and use numerical examples to compare our two-stage approach to some common heuristic alternatives in Section 7. Further discussion on the problem and future research directions are presented in Section 8.

2. Literature review

The recent book of Talluri and van Ryzin (2004) provides a comprehensive overview of revenue management, including both capacity allocation and overbooking. Rothstein (1985) provides a very readable account of the history of overbooking in the airline industry. Similarly, Ratliff (1998) presents a survey and focuses on practical problems in overbooking decisions. The earliest overbooking models in the literature either take a cost-based approach (e.g. economic models that balance the cost of overselling with the opportunity cost of empty seats) and/or a service-level-based approach (e.g. a bound on the expected number of passengers denied service or the probability that a passenger is denied service because of overbooking). Beckmann (1958) provides a static, single period cost-based model, to determine an upper bound on the number of reservations to accept. Thompson (1961) shows a way to determine overselling probabilities for a static, single leg problem. His work is refined by Taylor (1962), and Rothstein and Stone (1967). Shlifer and Vardi (1975) provide static cost-based and service-level-based models for both a single-leg flight carrying two types of passengers and a two-leg flight. Several researchers have addressed dynamic models of overbooking for single leg flights (Chatwin (1992, 1999), Rothstein (1971), Subramanian et al. (1999), and Alstrup et al. (1986)). Models for the hotel industry are presented in Rothstein (1974), Bitran and Gilbert (1996) and Liberman and Yechiali (1978). Karaesmen and van Ryzin (2004), present a model that solves an overbooking problem where the resources are substitutable (e.g. hotel rooms, sequential flights on the same route).
Unfortunately, many of these models are not directly useful for capacity control. The exceptions are the two-leg airline model of Shlifer and Vardi (1975), and single-leg, multiple-fare-class models of Chatwin (1999) and Subramanian et al. (1999). Subramanian et al. (1999) in particular provide a very complete analysis of the single-leg model with both overbooking and capacity controls.

Likewise, the literature on combining overbooking and capacity control on a network is limited. Various approximate approaches to network problems with overbooking have been proposed. Bertsimas and Popescu (2003) provide an approximation method for dynamically controlling seat inventories with cancellations on a network. They develop a dynamic programming approximation to control seat inventories and show that it can be heuristically adjusted to handle cancellations or no-shows by adjusting problem parameters to reflect the effect of expected cancellations. Kleywegt and Bharadwaj (2001) analyze a deterministic model of the capacity control problem with cancellations when the customers choose which itinerary to fly. Similarly, Kleywegt (2001) proposes an optimal control model for the network revenue management problem, considering pricing (as a proxy to capacity controls), overbooking, and customer choice. The focus of the last two papers is on computational methods. The former proposes a derivative free search method while the latter proposes a bundle-trust method. Lim (2000) combines the well known DLP with overbooking. He models the joint decision problem by representing it as a mathematical program with equilibrium constraints, where the cost of spoilage (denied boardings) is approximated by a deterministic exponentially decreasing (increasing) function. He solves this model using a penalty interior point algorithm. Lim’s work is the most similar in spirit to ours in that it explicitly combines capacity allocation and overbooking decisions, though there are significant differences in our modelling approach and solution methodology.

3. An Overview of Traditional Capacity Control and Overbooking Models

We first briefly review the DLP model for capacity control and the economic model for overbooking. We then show that ad hoc attempts to combine them - as is often done in practice - may fail to effectively coordinate the two sets of decisions.
3.1 Deterministic Linear Programming Model for Capacity Control

Simpson (1989) first proposed the DLP model and the associated bid price policy for capacity control. The approach was subsequently analyzed in more detail by Williamson (1992).

The DLP model is defined as follows: An airline operating on a network offers \( n \) products (i.e. fare class and itinerary combinations) on \( m \) legs (resources). The capacity on leg \( i \) is denoted \( c_i \) and \( c = (c_1, \ldots, c_m) \) is the vector of network capacity. We represent the product requirements by an incidence matrix \( A = [a_{ij}] \), where \( a_{ij} = 1 \) if product \( j \) uses leg \( i \), and \( a_{ij} = 0 \) otherwise. Each product \( j \) has a revenue \( r_j \), and we let \( r = (r_1, \ldots, r_n) \) denote the vector of product revenues. The demand (reservation requests) for each product is uncertain. The mean demand for product \( j \) is known to be \( \mu_j \). DLP model is then

\[
R^{DLP}(c) = \max_x r \cdot x \quad s.t. \quad Ax \leq c, \quad 0 \leq x \leq \mu. \quad (1)
\]

where \( x = (x_1, \ldots, x_n) \) represents the allocation of seat capacity to products. Let the vector of dual variables associated with the capacity constraints \( Ax \leq c \) be \( \lambda \). Let \( x^* \) be the optimal primal solution, and \( \lambda^* \) be the optimal shadow price. Intuitively, these shadow prices approximate the marginal opportunity cost of capacity on each leg of the network.

The DLP solution can be used to define two different capacity control policies for dynamically accepting/rejecting reservation requests over time:

- **Primal allocation (PA):** \( x^* \) is used to define a partitioned allocation of capacity in which reservation requests for product \( j \) are accepted up-to the primal allocation level \( x^*_j \).

- **Bid pricing (BP):** \( \lambda^* \) is used to define a set of threshold values - called bid prices - such that a reservation request for product \( j \) is accepted if there is capacity remaining on all the legs product \( j \) uses and the fare for \( j \) exceeds the sum of bid prices (shadow prices) associated with those legs (i.e. if \( r_j \geq \sum_{i=1}^{m} a_{ij} \lambda^*_i \)).

Note the policy parameters produced by the DLP are static, being based on a simple deterministic approximation to the problem (replacing the random demand to come by its mean). To make the resulting policies more adaptive, in practice the model is resolved frequently to update the policy parameters as bookings and capacity remaining evolve.

The PA policy is not used often in airline practice. This is because the partitioned allocations do not “pool” capacity allocations effectively. Still, the derived PA policy is
shown to be asymptotically optimal for the stochastic and dynamic capacity control problem with no cancellations (see Cooper (2002)). (Yet partitioned allocations have been used in some passenger-railway applications as described in Ciancimino et al. (1999).) The BP policy is used more frequently in practice, but typically requires frequent resolving of the model to ensure bid prices track the changes in remaining capacity and demand to come.\footnote{Although the results of static models are used in a rolling horizon fashion by resolving, Cooper (2002) and Secomandi (2003) show that resolving DLP may not be beneficial.}

Although the DLP model is quite simplistic, the resulting BP policy generally has good performance (see Williamson (1992)). One can compute the opportunity cost of a seat on a leg of the network using methods that are more sophisticated than the DLP (e.g. Bertsimas and Popescu (2003), Talluri and van Ryzin (1999b) and Talluri and van Ryzin (2004)), and generally these methods improve the performance of BP policies. Still, because of its simplicity and computational efficiency, the DLP remains a popular model.

When applying the DLP, one typically computes a set of virtual capacities $u$ and uses these in place of the physical capacities $c$ in formulation (1), in which case the first constraint becomes $Ax \leq u$. Note in this case that both the primal allocation $x$ and the bid price value $\lambda$ will depend on the values of the virtual capacities $u$ – an example of the interdependence between the overbooking and capacity control models.

### 3.2 Economic Model for Overbooking

The economic model (Beckmann (1958)) of overbooking determines the virtual capacity (or overbooking limit) $u_i$ for leg $i$ by solving

$$\max_{u_i \geq c_i} = b_i u_i - C_i(u_i)$$  \hfill (2)

where $C_i(u_i)$ is the expected overbooking cost for leg $i$ and $b_i$ is an estimate of the marginal value of providing an additional unit of virtual capacity (the marginal value of an unsold seat). The expected overbooking cost is expressed as

$$C_i(u_i) = q_i E[Z_i(u_i) - c_i]^+]$$  \hfill (3)

where $q_i$ the cost of denied service on leg $i$ and the random variable $Z_i(u_i)$ the show demand for leg $i$ at flight time. We use the notation $(Z_i(u_i) - c_i)^+$ for $\max\{Z_i(u_i) - c_i, 0\}$.

Note that in computing the expected overbooking cost, the show demand has to be represented using a probabilistic model. One natural choice is the Binomial model.\footnote{Although the results of static models are used in a rolling horizon fashion by resolving, Cooper (2002) and Secomandi (2003) show that resolving DLP may not be beneficial.}
That is, \( Z_i^B(u_i) \) is a binomial distributed random variable with parameters \( p_i \) and \( u_i \) where \( p_i \) is the probability that a booking on leg \( i \) shows-up at flight time, i.e. \((1 - p_i)\) is the probability of cancellation. In this case, one can show that the random variable \( Z_i^B(u_i) \) is stochastically increasing and linear in \( u_i \) (see Yao (1994) for a definition of stochastic linearity) and the associated overbooking cost,

\[
C_i^B(u_i) = q_i E_{Z_i^B(u_i) - c_i} = q_i \sum_{z=c_i}^{u_i} \binom{u_i}{z} p_i^z (1 - p_i)^{u_i - z},
\]

is non-decreasing and discrete convex in \( u_i \). Hence, the optimal virtual capacity of leg \( i \) can be determined by a simple search procedure. Alternatively, one can use the Poisson approximation to binomial distribution, which is useful for analytical purposes. Approximating \( Z_i^P(u_i) \) by a Poisson random variable with mean \( p_i u_i \), the expected overbooking cost becomes

\[
C_i^P(u_i) = q_i E_{Z_i^P(u_i) - c_i} = q_i \sum_{z=c_i}^{\infty} (z - c_i) e^{-p_i u_i} \frac{(p_i u_i)^z}{z!}.
\]

One can easily show that this approximate overbooking cost is non-decreasing, continuous, differentiable and convex in \( u_i \).

While the economic model of overbooking results in a simple optimization problem, often even simpler deterministic rules are used in practice (see Ratliff (1998), Belobaba (2001)). For example, one can take the overbooking pad as a fixed percentage of the cabin capacity based on a pre-determined service level. Alternatively, one can scale the capacity of a leg using cancellation probabilities. We investigate the performance of such rules in Section 7.

### 3.3 Ad Hoc Combination of Seat Inventory Control and Overbooking

Since the economic model of overbooking requires an estimate of the “opportunity cost” of an unsold seat on a leg and the bid prices of a capacity control model provide just such an estimate, this suggests one simple idea to coordinate overbooking and capacity control models: Simply alternate between solving a capacity control model to determine opportunity costs and an economic overbooking model to determine virtual capacities, in the hope that the two converge to a good solution. Indeed, such iterative application of overbooking and capacity control models roughly mimics the way these models are often combined in practice.

Despite the intuitive appeal of this idea, a simple example shows that such procedures may not converge. The example uses the DLP model and the economic model of overbooking.
with binomial distributed cancellations. There is only one leg and two fare classes. The penalty for oversales be \( q_1 = 120 \) (refund on ticket price plus a compensation), and the probability of a reservation showing up be \( p_1 = 0.9 \). There is no cancellation fee. The fares are $100 and $50. The seat capacity is 100 and expected demand for the fare classes is 50 and 60, respectively.

Starting with an overbooking pad of zero, the DLP is

\[
\max \ 50x_1 + 100x_2 \quad \text{s.t.} \quad x_1 + x_2 \leq 100, \ 0 \leq x_1 \leq 50, \ 0 \leq x_2 \leq 60.
\]

The resulting bid price associated with the capacity constraint is \( \lambda_1^* = 50 \) (the optimal primal solution is \( x_1^* = 40, \ x_2^* = 60 \)). When this bid price is used as the opportunity cost of an unsold seat, the overbooking problem (2) becomes

\[
\max_{u_1 \geq 100} \left\{ 50u_1 - 120E[Z^B(u_1) - 100]^+ \right\}.
\]

The resulting overbooking limit is \( u_1^* = 111 \). Solving the DLP once again with the revised capacity constraint \( x_1 + x_2 \leq 111 \), we obtain a dual price of \( \lambda_1^* = 0 \) (with primal solution \( x_1^* = 50, \ x_2^* = 60 \)). Resolving the overbooking problem with the revised opportunity cost of \( \lambda_1^* = 0 \) in turn gives an overbooking pad of zero, which takes us back to the original problem. Hence, the procedure oscillates between these two solutions and never converges.

This simple example shows that ad hoc exchanges of bid prices and virtual capacities may not guarantee good coordinating values – or even convergence. A more sophisticated approach is needed. This is the focus of the next section.

4. A General Formulation of the Joint Capacity Control and Overbooking Problem and the Two-Stage Approximation

We first develop a general description of the joint overbooking and inventory control problem for a network. While this model itself makes some simplifying assumptions, it still results in a difficult joint optimization problem.

4.1 A General Model of the Problem

To begin, we break the time horizon into two periods: a reservation period followed by a service period. The reservation period spans \( (0, T] \), and is the period during which the reservations can be made for any of the \( n \) products (fare class-itinerary combinations). The fares
of the products are determined in advance and are assumed to be constant over the entire reservation period. Cancellations and no shows occur only at the end of the reservation period. We use the term survivors to refer to customers who make reservations and show-up at the service period.

During the service period, the airline may need to deny boarding to some of the survivors if legs are oversold. The airline may need to pay a compensation to the customers who are denied service and incurs goodwill costs. Collectively, we refer to these as the overbooking costs. We ignore refunds on cancellations and no-shows, though these can be included in the model by modifying the product revenue values as shown by Subramanian et al. (1999).

Reservation requests for $n$ products arrive according to a stochastic process during $(0, T]$. We consider a general continuous-time arrival process $D(\cdot)$, with mean number of reservation requests $\mu_j$ for product $j$ during $(0, T]$. We assume reservation requests of products are independent of each other and satisfy $\text{Var}(D_j((0, T])) = \sigma^2 < \infty$ for all $j$. We assume $E[D(\{t\})] = 0$ for all $t$. There are no reservations in the system at time $t = 0$.

A policy defines the rules for processing reservation requests during $(0, T]$. If a request for product $j$ is accepted, fare $r_j$ is received. Let the total number of reservations accepted during the reservation period $(0, T]$ when policy $\pi$ is in effect be represented by the $n-$ vector $N^\pi = (N_1^\pi, ..., N_n^\pi)$. $N^\pi$ is a random vector, since it depends on the demand process during $(0, T]$. Then, $AN^\pi$ is the vector of accepted bookings at the leg level at time $t = T$, where $A$ is the product-leg incidence matrix introduced in Section 3.1.

Naturally, the number of survivors on the network in the service period depends on the total number of reservations accepted during the reservation period. Consequently, the overbooking costs are determined based on the number of survivors. We make the simplifying assumption that cancellations and overbooking costs are independent across legs. That is, a customer will cancel on each leg of his/her itinerary independently. Further, we assume that the number of cancellations and no-shows on each leg is only a function of the total number of reservations on the leg and not the mix of products that are booked. Both of these assumptions are generally violated in real life, but they are common approximations in airline practice. Again, Subramanian et al. (1999) discuss the complexities that arise when there are class-dependent cancellation probabilities.

\footnote{This is essentially the pure no-show case; however, we refer to both cancellations and no-shows because when applying the model we combine both cancellations and no-shows. Indeed, in Section 7 we look at numerical examples in which cancellations occur prior to the service period.}
As a result of these assumptions, the overbooking costs are only a function of the aggregate number of reservations accepted on each leg, given by the vector $AN^\pi$. Let $C(AN^\pi)$ denote this expected overbooking cost. We assume $C(\cdot)$ is a non-decreasing function of each of its components, which is quite natural as more reservations on hand lead to higher chances of denied service and therefore higher expected overbooking costs.

The joint overbooking and capacity control problem can now be expressed (somewhat abstractly) as

$$
\nu^{**} \equiv \sup_{\pi \in \Pi} \{ E_D[ r \cdot N^\pi - C(AN^\pi) ] : 0 \leq N^\pi \leq D \},
$$

where $D = (D_1, \ldots, D_n)$ is the vector of total number of reservation requests during $(0, T]$, and $\Pi$ is the set of all history-dependent, non-anticipating reservation acceptance policies. Despite the assumptions we have made thus far, this is still a difficult problem to solve.

### 4.2 A Two-Stage Approximation Method

Our proposed approximation for problem (6) is based on two simplifying approximations: First, we assume that the airline pre-commits to the maximum number of reservations it will accept on each leg. That is, it pre-commits to virtual capacities, denoted $u = (u_1, \ldots, u_m)$, at the start of the booking process and then enforces the constraint $AN^\pi \leq u$ on its seat inventory control policy. Second, we approximate the overbooking cost $C(AN^\pi)$ by the upper bound $C(u)$. This corresponds to assuming that the number of reservations accepted on leg $i$ always reaches the upper bound $u_i$. Since the overbooking cost $C(u)$ is assumed to be non-decreasing in $u$, this results in an over-estimate of the true overbooking cost.

Using these simplifications, the joint overbooking and seat inventory control problem becomes a two-stage optimization problem. The first stage problem is to determine the virtual capacities by solving

$$
\nu \equiv \sup_{u \geq c} \{ R(u) - C(u) \}
$$

where $C(u)$ is the expected overbooking cost and $R(u)$ is the value of the second stage problem

$$
R(u) = \sup_{\pi \in \Pi} \{ E_D[ r \cdot N^\pi ] : AN^\pi \leq u \ (a.s.), \ 0 \leq N^\pi \leq D \}.
$$

In words, $R(u)$ is the maximum revenue given that the airline pre-commits to using a fixed set of virtual capacity levels $u$ and (8) is equivalent to the classical stochastic, dynamic seat
inventory control problem with no cancellations. On the other hand, if $R(u)$ were a linear function of virtual capacity $u$, i.e. $R(u) = \sum_{i=1}^{n} b_i u_i$ for some $b_i$ for leg $i$, and $C(u)$ was additive, i.e. $C(u) = \sum_{i=1}^{n} C_i(u_i)$, then the first stage problem (7) would separate into a collection of single-leg overbooking problems of the type discussed in Section 3.2.

To gain some insight into the coordinated overbooking and seat inventory control decisions produced by this two-stage approximation, suppose $R(u)$ and $C(u)$ were differentiable functions of $u$. Then (assuming a strictly interior solution), the first order conditions for the first stage problem (7) imply that the vector of optimal virtual leg capacities $u^*$ satisfies

$$\nabla R(u^*) = \nabla C(u^*).$$

This relation can be interpreted in two ways. Viewed in terms of the economic overbooking model, these first-order conditions imply $u^*$ maximizes $\nabla R(u^*) \cdot u - C(u)$. This corresponds to an overbooking problem with $\nabla R(u^*)$ representing the opportunity costs of unsold seats. On the other hand, the gradient of the revenue with respect to virtual capacity, $\nabla R(u)$, is precisely the vector of bid prices discussed in Section 3.1. Thus, the first-order conditions imply that the opportunity cost of unsold seats in the overbooking model should equal the bid price values produced by the network seat inventory control model at a set of optimal virtual capacity levels $u^*$. This equivalence is quite intuitively appealing.

While this two-stage approximation method is intuitively appealing, two immediate questions arise. First, how can the optimal virtual capacities $u^*$ be computed for problem (7)? Ideally, we would like this computation to take advantage of data and outputs available from overbooking and seat inventory control models currently in use. Second, how well does it perform, both theoretically and practically? We address these questions next.

5. Analysis of a Special Cases: DLP Combined with the Economic Overbooking Model

Our two-stage approximation method is general and can be used to combine various capacity control and overbooking models. However, here we focus on the the DLP allocation model and economic overbooking model introduced in Section 3, because these are widely used models in

---

3We note that one does have to worry about degeneracy in network optimization models producing multiple dual solutions, in which case $\nabla R(u^*)$ is not well defined. See Talluri and van Ryzin (1999a) for a discussion of this issue.
practice and are computationally efficient to solve. We also test it with a randomized version of the DLP proposed by Talluri and van Ryzin (1999b) in Section 7.

5.1 Analysis Based on Partitioning

Our approximation can be analyzed using ideas of partitioning which are well-known in non-linear programming and large-scale optimization (Bertsekas, 1999). Partitioning separates the optimization regarding virtual capacities from that of capacity allocation, and the resulting problem becomes well-suited to primal-dual based solution methods.

When an approximate method is used for the capacity control problem, we denote the resulting revenue as a function of given virtual capacity vector \( u \) as \( \bar{R}(u) \). In our case here, the second-stage revenue function \( R(u) \) is approximated by the DLP, so \( \bar{R}(u) = R^{DLP}(u) \). We assume here overbooking decisions are based on the economic model with Poisson cancellations, so the expected overbooking cost \( C(u) \) is separable,

\[
C(u) = \sum_{i=1}^{m} C_i^p(u_i),
\]

jointly convex in \( u \) and continuously differentiable.

The two-stage optimization problem is then

\[
\tilde{\nu} = \max_{u \in U} \bar{R}(u) - C(u)
\]

where

\[
\bar{R}(u) = \max F(x) \quad s.t. \quad Ax \leq u, \quad x \in X.
\]

With the change of notation, we have \( F(x) = rx \) and \( X = \{x : 0 \leq x \leq \mu\} \) and \( U = \{u | u \geq c\} \). We then have the following result from Bertsekas (1999):

**Proposition 1** (Bertsekas (1999), Proposition 6.2.1) Assume the problem

\[
\max F(x) \quad s.t. \quad Ax \leq u, \quad x \in X
\]

has an optimal solution and at least one Lagrange multiplier for each \( u \in U \). Then, the set of subgradients of \( \bar{R}(u) \) at \( u \) is the set of all Lagrange multipliers of problem (11) corresponding to the constraint \( Ax \leq u \).

The above proposition shows that the subgradients for the first-stage problem in (10) can be determined by subtracting the subgradient of the overbooking revenue function \( C(u) \) from
the “bid prices” of problem (11). One can then solve the first-stage problem by a subgradient optimization method to determine the optimal virtual capacities. Note this partitioning analysis suggests why the ad hoc combination of overbooking and DLP model of the type discussed in Section 3.3 may fail; the gradient information of the capacity control problem must be combined with the gradient information of the overbooking cost to determine an ascent direction for the combined net revenue function.

Note that the two-stage optimization problem in (10) is equivalent to

$$\tilde{\nu} = \max F(x) - C(u) \quad \text{s.t.} \quad Ax \leq u, \quad x \in X, \quad u \in U.$$  \hspace{1cm} (13)

The Karush-Kuhn-Tucker (KKT) conditions for this problem are easy to determine and are omitted. However, let \((x^*, u^*, \lambda^*)\) be a KKT point of (13), where \(\lambda^*\) is the multiplier associated with the constraints \(Ax \leq u\). The following observations follow directly from the KKT conditions:

**Proposition 2** Suppose \((x^*, u^*, \lambda^*)\) is a KKT point of problem (13). Then,

(i) \(u^*_i = \sum_{j=1}^{n} a_{ij} x^*_j\) and \(\lambda^*_i \geq 0\) if and only if \(\sum_{j=1}^{n} a_{ij} x^*_j \geq c_i\) for leg \(i\),

(ii) if \(u^*_i > c_i\), then \(\lambda^*_i = \frac{\partial}{\partial u^*_i} C(u^*) > 0\) for \(i = 1, \ldots, m\).

The first property shows that the virtual capacity constraint is binding and the bid price is non-negative when the total partitioned allocation on a leg exceeds the leg capacity. Also, there is no “slack” capacity on a leg unless the total partitioned allocation is lower than the physical capacity of a leg. This is natural since the overbooking cost is non-decreasing in virtual capacities. The second property shows that the bid-price for leg \(i\) is equal to the gradient of the overbooking cost function. This is exactly what we argued earlier in Section 4 by equation (9).

The next observation is elementary but important.

**Proposition 3** Suppose \((x^*, u^*, \lambda^*)\) is a KKT point of (13). The optimal virtual capacity satisfies \(c \leq u^* \leq \max(A\mu, c)\).

This, unfortunately, illustrates a key drawback of using the DLP model; it does not allow any oversales when the total mean demand on a leg is below the capacity. In reality, demand may be highly variable and actual reservation requests may exceed the seat capacity even though the mean demand is low. Yet the optimal virtual capacity fails to take this into account. This drawback may be overcome by using other seat inventory control models as we discuss later in this section.

14
5.2 Asymptotic Analysis

When DLP is used to approximate \( R(u) \), the first stage solution \( u^\ast \) provides virtual capacities and the corresponding second-stage solution \( x^\ast \) can be used to control the network capacity in a PA policy (see the definition in Section 3.1). We next look at how good the PA policy is (asymptotically) as a solution to our original two-stage problem (7).

Suppose a product class \( j \) demand arrives at time \( t, t \in (0,T] \), then the PA policy, denote \( \pi^{PA} \), is defined by:

\[
\pi^{PA}_j \equiv 1(D_j((0,t]) \leq x^*_j),
\]

where \( 1(\cdot) \) is the indicator function. That is, the policy accepts a reservation request of class \( j \) at time \( t \) if the total number of class \( j \) reservation requests prior to time \( t \) have not exceeded the primal allocation limit \( x^*_j \). The total number of customers accepted in product class \( j \) using \( \pi^{PA} \) is a random variable given by

\[
N^{PA}_j = \min(D_j, x^*_j).
\]

The random revenue generated by this policy is:

\[
R'(N^{PA}) = F(N^{PA}) - C(AN^{PA}) = r \cdot N^{PA} - C(AN^{PA}).
\]

Notice that the number of accepted reservations satisfies \( N^{PA} \leq x^* \leq \mu \) and \( AN^{PA} \leq u^* \). Note that \( E_D[R'(N^{PA})] \neq \nu \), where \( \nu \) is the optimal objective function value of the approximation in (10) (and its equivalent in (13)).

The next two lemmas give upper and lower bounds on the expected net revenue obtained by the PA policy. The proofs of the results in this section are presented in the Appendix unless noted otherwise. Since the number of requests accepted is at most \( Ax^* \) and \( C(\cdot) \) is increasing componentwise, \( C(Ax^*) \) is an upper bound on the overbooking cost under the PA policy. Hence, \( LB = E_D[r \cdot N^{PA}] - C(Ax^*) \) is a lower bound on the policy’s net revenue. Thus, we have:

**Lemma 1** Define \( LB = E_D[r \cdot N^{PA}] - C(Ax^*) \). Then, \( \nu \geq E_D[R'(N^{PA})] \geq LB \).

The DLP model (10) solves a “deterministic” version of (7) by replacing \( D \) in the constraint set with its mean over the entire horizon, \( \mu \). This leads to the next result which holds by the concavity of the functions and Jensen’s inequality (proof omitted).

**Lemma 2** \( \bar{\nu} \geq \nu \).
Following Cooper (2002), define a sequence of “scaled” parameters and associated problems. Let the $k^{th}$ optimization problem be defined as

$$\tilde{v}^k = \max_{u \geq kc} \tilde{R}^k(u) - C^k(u)$$

(15)

where

$$\tilde{R}^k(u) = \max_x rx \ s.t. \ Ax \leq u, 0 \leq x \leq k\mu,$$

$$C^k(u) = \sum_{i=1}^m q_i E[Z_i(u_i) - kc]^+].$$

That is, the mean demand and the capacity parameters are scaled as $k\mu$ and $kc$, respectively.

We denote the optimal solution of the scaled problem by $(x^{*k}, u^{*k})$ and the PA-policy using results of the $k^{th}$ problem yields $N^{PA(k)}$. Consequently, the lower bound in Lemma 1, denoted $LB^k$, becomes:

$$LB^k = E[D(N^{PA(k)}) - C^k(Ax^{*k}).$$

The net expected revenue of an optimal policy for the two-stage problem (7) with scaled parameters is denoted $\nu^k$. We consider a general continuous-time arrival process $D(\cdot)$ as before. For the scaled problems, we need to define the demand such that $k^{-1}D^k((0,T]) \rightarrow E[D(0,T)]$ in distribution. We can obtain such a sequence of demand processes by $D^k(\cdot) = \tilde{D}^1(\cdot) + \tilde{D}^2(\cdot) + \cdots + \tilde{D}^k(\cdot)$, where each $\tilde{D}^i, i = 1, \ldots, k$ are independent copies of $D(\cdot)$. This process satisfies $k^{-1}D^k((0,T]) \rightarrow D((0,T])$ in distribution. This construction allows the shape of the expected buildup of demand to be preserved. Properties of the scaled problems are stated below.

**Lemma 3** If $\mu < \infty$, then $(1/k)LB^k$ and $(1/k)\tilde{v}^k$ are bounded.

**Lemma 4** $(1/k)[\tilde{v}^k - LB^k] \rightarrow 0$ as $k \rightarrow \infty$.

Finally, we establish the asymptotic property of the PA policy.

**Theorem 1** $\frac{\nu^k}{E[D(N^{PA(k)})]} \rightarrow 1$ as $k \rightarrow \infty$.

**Proof** Using the results in Lemmas 1 through 4, we have $\tilde{v}^k \geq v^k \geq E[R'(N^{PA(k)})] \geq LB^k$ and $(\tilde{v}^k/LB^k) \rightarrow 1$ as $k \rightarrow \infty$. Therefore,

$$\frac{\tilde{v}^k}{LB^k} \geq \frac{\nu^k}{E[R'(N^{PA(k)})]} \geq 1$$
which completes the proof.

This shows that, as the demand and the capacity get large, the ratio of the expected net revenues of the PA policy obtained by DLP to that of an optimal policy for our two-stage problem goes to one. The above result does not imply that the difference between expected net revenues of the PA policy and the best policy diminish, although the difference as proportional to the scaling parameter $k$ diminishes. Unfortunately, the same property is not easy to establish for the BP policy.

5.3 Alternative Capacity Control Models

While we have focused on the case where DLP is used for network capacity control, there are several other models that can be used in our general two-stage approach. One alternative is the static, probabilistic non-linear programming (PNLP) model (see Williamson (1992), Talluri and van Ryzin (1999a) and Talluri and van Ryzin (2004) for details) that approximates the second stage problem by solving

$$\tilde{R}(u) = R^{PNLP}(u) = Max \sum_{j=1}^{n} r_j E[D][\max\{x_j, D_j\}] \text{ s.t. } Ax \leq u, \quad x \geq 0.$$ 

Note that, the partitioning approach introduced in Section 5.1 also applies to PNLP. The idea is to solve the two-stage problem introduced in (10) by adjusting the second-stage formulation in (11) so that $F(x) = r E[D][\min(x, D)]$ and $X = \{x|x \geq 0\}$. Consequently, Proposition 1 holds. Investigating a KKT solution of the problem shows that Proposition 2 also holds for PNLP. Further, PNLP does not suffer from the bad behaviour noted in Proposition 3 in which DLP does not permit oversales when mean demand is only slightly under capacity. In contrast, PNLP may permit oversales in such cases.

Another alternative is the randomized linear programming (RLP) procedure introduced by Talluri and van Ryzin (1999b). The idea is to solve the problem

$$R^{RLP}(u) = E_D \left[ \max_x \{ r \cdot x : Ax \leq u, 0 \leq x \leq D \} \right].$$

Here, the random demand $D$ is in the constraint right-hand-side of the linear program and allocations are made with perfect knowledge of each realization of $D$. The function $R^{RLP}(u)$ can be computed approximately by sampling. With $S$ samples of the demand $d^1, ..., d^S$, the revenue function can be computed by

$$\tilde{R}(u) = R^{RLP(S)}(u) = \frac{1}{S} \sum_{s=1}^{S} \max_x \{ r \cdot x^s : Ax^s \leq u, 0 \leq x^s \leq d^s \}.$$  

(17)
The analysis for DLP can be repeated for RLP, keeping in mind that given \( u \) and the

demand samples \( d^1, \ldots, d^S \), we solve \( S \) replicas of the DLP. Using the partitioning approach

and Proposition 1, we see that the average of the Lagrange multipliers associated with

the virtual capacity constraints in \( S \) problems is the subgradient of \( \bar{R}(u) \) at \( u \). RLP also

overcomes the DLP problem noted in Proposition 3.

Similar to DLP, the solution to PNLP defines both a PA and a BP policy for network

capacity control. RLP, in contrast, only defines a BP policy, where the bid prices are

computed by averaging the shadow prices obtained by solving \( S \) samples of the DLP. (See

Talluri and van Ryzin (1999b) for details.)

### 5.4 Alternative Overbooking Models

One can also use alternative overbooking models in the two-stage approach. For example,

one model of cancellations used in practice is based on the estimation of cancellation rates

(see Ratliff (1998)). In this model, the number of show-ups on leg \( i \) is \( Z_i^\alpha(u_i) = u_i\alpha_i \) where

\( \alpha_i \) is a random variable with known probability distribution \( G_i(\cdot) \) and support \([0, 1]\). In fact,

\( (1 - \alpha_i) \) is the fraction of bookings on leg \( i \) that are cancelled, and the probability distribution

\( G_i(\cdot) \) can be estimated using historical data. The corresponding expected overbooking cost

for leg \( i \) is

\[
C_i^\alpha(u_i) = q_i E_{Z_i}[(Z_i^\alpha(u_i) - c_i)^+] = q_i \int_{c_i/u_i}^1 (u_i z - c_i) dG_i(z). 
\] (18)

Note that the expected overbooking cost in this case is non-decreasing, continuous, differentiable

and convex with respect to \( u_i \). Hence, our analytical results would hold if this cancellation rate based model were combined with any of the network capacity control models discussed above.

### 6. A Computational Algorithm Based on the Alternating Direction Method of Multipliers

The structure produced by our two-stage method is well suited to computation using an alternating
direction method of multipliers for the augmented Lagrangian function. The method

proceeds as follows: First, rewrite problem (10) such that the virtual capacity constraints

are represented as equalities

\[
\bar{\nu} = \max \ r \cdot x - C(u) \quad s.t. \ Ax + x^d = u, \ 0 \leq x \leq \mu, \ u \geq c, \ x^d \geq 0 \quad (19)
\]
using the slack variable \( x^l \). By changing the notation, we write \( A'x' = u \) where \( x' = (x, x^l) \) and \( A' = [AI] \), \( I \) being the \( m \)-by-\( m \) identity matrix. We define the sets \( U = \{ u : u \geq c \} \), \( X' = \{ x' = (x, x^l) : 0 \leq x \leq \mu, x^l \geq 0 \} \). We also use the notation \( F'(x') = r \cdot x \) for the sake of completeness. The resulting augmented Lagrangian function is:

\[
\tilde{L}(x', u, \lambda) = F'(x') - C(u) - \lambda^T (A'x' - u) - \frac{\zeta}{2} \|A'x' - u\|^2.
\]

An alternating direction method to find the maximizers of this augmented Lagrangian proceeds at iteration \( k + 1 \) as follows:

\[
x'(k + 1) = \arg \max_{x' \in X'} \{F'(x') - \lambda(k)^T A'x' - \frac{\zeta}{2} \|A'x' - u(k)\|^2\} \tag{20}
\]

\[
\begin{align*}
\lambda(k + 1) &= \lambda(k) + \frac{\zeta}{2} A'x'(k + 1) - u(k + 1). \\
u'(k + 1) &= \arg \max_{u \in U} \{\lambda(k)^T u - C(u) - \frac{\zeta}{2} \|A'x'(k + 1) - u\|^2\} \tag{21}
\end{align*}
\]

The parameter \( \zeta \) is any positive number, and initial vectors \( u(0) \) and \( \lambda(0) \) are arbitrary.

The next result assures the convergence of the alternating direction method. The proof is available in Bertsekas and Tsitsiklis (1997), Section 3.4.

**Theorem 2** A sequence generated by \( \{x'(k), u(k), \lambda(k)\} \) by the algorithm of equations (20), (21), and (22) is bounded and every limit point of \( \{x'(k), u(k)\} \) is an optimal solution to the original problem (19). Furthermore \( \{\lambda(k)\} \) converges to the optimal dual variable associated with \( A'x' = u \).

The alternating direction method proves to be very efficient for the special case of revenue functions \( F'(x') \) and \( C(u) \) based on the DLP and economic overbooking models, respectively. The method iteratively solves a problem for \( x' \), another one for \( u \) and updates the multiplier \( \lambda \). We use \( \zeta = 1 \) below for illustration.

**Finding \( x'(k+1) \).** We have to solve the following problem:

\[
\max_{x' \in X'} F'(x') - \lambda(k)^T A'x' - \frac{1}{2} \|A'x' - u(k)\|^2
\]

which is equivalent to

\[
(QP) \max_{x' \in X'} F'(x') + (u(k) - \lambda(k))^T A'x' - \frac{1}{2} (x')^T (A')^T A'x'. \tag{23}
\]
Problem (QP) is a quadratic programming problem since $F'(x')$ is linear. Hence, it can be solved by a standard solution procedure (e.g. see Bazaraa et al. (1993)).

Finding $u(k+1)$. This requires solving

\begin{equation}
(SP) \quad \max_{u \in c} \lambda(k)^T u - C(u) - \frac{1}{2} ||A'x'(k + 1) - u||^2 \\
= -C(u) + (\lambda(k) + A'x'(k + 1))^T u - \frac{1}{2} u^T u + \frac{1}{2} x'(k + 1)^T (A')^T A' x'(k + 1).
\end{equation}

For our special case, the problem (SP) is separable. Therefore, for leg $i$ we solve

\[
\min_{u_i \geq c_i} \psi(u_i) = C_i^P(u_i) - (\lambda_i(k) + [A'x'(k + 1)]_i)u_i + \frac{1}{2}u_i^2,
\]

where $[A'x'(k + 1)]_i$ is the $i^{th}$ coefficient of the matrix-vector product $A'x'(k + 1)$. When the Poisson approximation to binomial is used in the overbooking cost function, we have $\frac{d}{du_i} C_i^P(u_i) = q_i p_i P(Z_i^P(u_i) \geq c_i)$ and $\psi$ is a convex function. Thus, one can compute the optimal $u_i(k + 1)$ by a simple search procedure.

Below we summarize the alternating direction method of multipliers for our problem.

**Step 0** Initialize: $\zeta = 1, \ u(0) = c, \ \lambda(0) = 0, \ k = 0$.

**Step 1** Solve problem $(QP)$ in (23) and get $x'(k + 1)$.

**Step 2** Solve problem $(SP)$ in (24) and get $u(k + 1)$.

**Step 3** Compute $\lambda(k + 1)$ by (22).

**Step 4** If “stopping criteria” is satisfied, STOP. Otherwise, set $k \leftarrow k + 1$, GOTO Step 1.

There are several choices for the stopping criteria: (i) check that $x(k), u(k), \lambda(k)$ satisfy the KKT conditions, (ii) check that $x(k), u(k), \lambda(k)$ do not significantly vary from one iteration to another during execution of the algorithm, or (iii) reach a pre-set number of iterations; this can be done if one has prior numerical experience with the algorithm. To solve the problem numerically, we used our own source codes together with the codes available from Numerical Algorithms Group (NAG). The resulting algorithm has proved to be fast (convergence in about 30-50 iterations, with a negligible run-time) on the problems we have tested.
We note that, the alternating direction method for the augmented Lagrangian can also be applied when RLP is used in network capacity control. When randomization is used with $S$ samples from the demand distribution, the problem in Step 1 separates by scenario, and we solve $S$ quadratic programs. Then, Steps 2 and 3 are executed with minor adjustments.

7. Numerical Examples

We use several numerical examples to evaluate our two-stage approximation method. For comparison purposes, we test our procedure against several ad hoc approaches. Each ad hoc approach defines a different heuristic way to determine virtual capacities, primal allocation values and/or bid prices.

7.1 Policy Definitions

For each of the ad hoc approaches, the virtual capacities are determined first. Then these values are then fed into the DLP denoted $R^{DLP}(u^\theta)$ where $u^\theta$ is the given virtual capacity for approach $\theta$. Once the DLP is solved, one can use the optimal primal solution $x^\theta$, the optimal shadow prices $\lambda^\theta$ and the given overbooking level $u^\theta$ to apply the PA policy or BP policy to control bookings.

Several ad hoc procedures were considered: Two of the approaches (DET-1, DET-2) choose the virtual capacities using deterministic rules. The others use the economic model of overbooking and differ in terms of their estimation of the opportunity cost for each leg. They reflect the typical practice of using heuristic opportunity cost values (see Belobaba (2001)) and are denoted approaches OBC-1, OBC-2, OBC-3 and OBC-4.

We tested these against two variations of our two-stage approximation method, denoted POI and BIN, which combine economic model of overbooking with DLP. Two other variations, denoted R-POI and R-BIN, combine economic model of overbooking with the RLP network capacity control model.

The precise definition of the policies is as follows:

- DET-1 uses the virtual capacities $u^{DET-1} = c/p$.
- DET-2 uses $u^{DET-2} = c + c(1 - p)$; note that $u^{DET-2} < u^{DET-1}$ when $p < 1$.
- OBC-1 through OBC-4 determine the virtual capacity of each leg of the network separately by solving the economic model in (2). The overbooking cost function uses the
binomial distribution and is given in (4). In each of these approaches the choice of unit revenue per seat, $b_i$, is different.

- OBC-1 uses $b_i = \max\{f_{ij}^c : j = 1, \ldots, m\}$ for leg i where $f_{ij}^c = 0$ for $a_{ij} = 0$ and $f_{ij}^c = r_j / (\sum_{k=1}^{n} a_{kj})$ for $a_{ij} > 0$. That is, the approximate unit revenue on leg i is taken as the maximum of the pro-rated itinerary revenues on that leg.

- OBC-2 uses $b_i = \min\{f_{ij}^c : j = 1, \ldots, m\}$ for leg i where $f_{ij}^c = 0$ for $a_{ij} = 0$ and $f_{ij}^c = r_j / (\sum_{k=1}^{n} a_{kj})$ for $a_{ij} > 0$. That is, the approximate unit revenue on leg i is taken as the minimum of the pro-rated itinerary revenues on that leg.

- OBC-3 uses $b_i = \max\{f_{ij}^c : j = 1, \ldots, m\}$ for leg i where $f_{ij}^c = r_j$ for all $j = 1, \ldots, n$. That is, the approximate unit revenue on leg i is taken as the maximum of revenues of the itineraries using that leg.

- OBC-4 uses $b_i = \min\{f_{ij}^c : j = 1, \ldots, m\}$ for leg i where $f_{ij}^c = r_j$ for all $j = 1, \ldots, n$. That is, the approximate unit revenue on leg i is taken as the minimum of revenues of the itineraries using that leg.  

- POI solves the two-stage problem (13) assuming the expected overbooking cost is $C(u) = \sum_{i=1}^{m} q_i E[(Z_i^P(u_i) - c_i)^+]$. The alternating direction method is used to solve this continuous optimization problem.

- BIN solves two-stage problem (13) when cancellations are binomial distributed and the expected overbooking cost is $C(u) = \sum_{i=1}^{m} q_i E[(Z_i^B(u_i) - c_i)^+]$. Hence, the virtual capacities are restricted to non-negative integers in this case. We only provide a heuristic solution to the problem: We use the alternating direction method introduced in the previous section and compute an integer $u$ in Step 2 of the algorithm. The rest of the algorithm remains unchanged. The virtual capacity obtained, $u^{BIN}$, is then used in the DLP to determine $x^{BIN}$ and $\lambda^{BIN}$.

- R-POI combines the Poisson approximation based economic model of overbooking with the RLP to control network capacity. The problem is solved using the alternating direction method. Note that the method is adjusted to accommodate $S$ scenarios for the network capacity control problem. The resulting virtual capacity $u^{R-POI}$ and

---

4One can even use average fare contributions instead of minima and maxima and define variations of the economic model. Results are no different for average fare contributions, hence, they are omitted.
associated Lagrange multiplier $\lambda^{R-POI}$ are used in a BP policy to control network capacity.

- R-BIN is similar to R-POI, but the expected overbooking cost is represented using the binomial distribution. Heuristic solution to this discrete optimization problem is obtained using the alternating direction method adjusted as in BIN and R-POI. The virtual capacity $u^{R-BIN}$ and Lagrange multiplier $\lambda^{R-BIN}$ are used in a BP policy.

### 7.2 An Example Without Reoptimization

We first consider an example where each policy is applied statically without any reoptimization (i.e. without resolving the models periodically during the booking process). We assume reservation requests arrive according to a Poisson process during the reservation period. Following the reservation period, cancellations occur and the number of survivors is binomial distributed. We evaluate both PA and BP policies.

**Example 1.** This is a network with 4 legs as presented in Figure 1. The network was used in numerical experiments by Lim (2000) and Williamson (1992). There are seven products, each of which represents an itinerary (ATL/BOS, ATL/MIA, ATL/SAV, LAX/ATL, LAX/BOS, LAX/MIA and LAX/SAV).

We use this example to show the differences in performances of the heuristics. In order to do that, we use experiments with different sets of parameters and evaluate the performance of the heuristics using simulation. At the end of each experiment, we compute the relative performance of each of the heuristics. The measure we use is the percentage difference in average net revenues compared to that of the highest average net revenue obtained in an
experiment, which we call the “relative suboptimality” of an approach:

\[
\text{relative suboptimality} = 100 \frac{\text{max revenue in the experiment} - \text{revenue of the heuristic}}{\text{max revenue in the experiment}}.
\]

For instance, if POI has an average of $100 as the highest average net revenue among the PA policies in one experiment whereas the average net revenue of OBC-1 is $90 as a PA policy in the same experiment, then the relative suboptimality for POI and OBC-1 are 0 and 10%, respectively.

In the experiments, the capacity of each leg is set to 100, the overbooking penalty for each leg is 800. The survival probability for each leg is the same in an experiment. The demand is Poisson distributed. The mean demand for each product is 80. The fare of single-leg products is 100, whereas the fare for two-leg products is 200. In Table 1, we present the results of three experiments which use survival probabilities \(\{0.6, 0.75, 0.9\}\). Note that we only highlight the possible differences in performances of the heuristics in this example.

When the survival probability is 0.6, BIN yields the highest average net revenue among the PA policies, followed by POI. Other approaches yield average net revenues with relative suboptimalities ranging from 2.21% to 6.48%. When the survival probability is 0.75, the variability among approaches reduces. In fact, when the survival probability is 0.9, OBC-1, OBC-2 and OBC-3 provide the highest average net revenues, followed closely by BIN and DET-2. Average net revenues of POI and DET-1 are lower by more than 5% of the best heuristic in this experiment. The results are similar for BP policies; one heuristic that yields the highest average net revenue in one experiment, may yield relatively low revenues in a different experiment.

### 7.3 Examples with Reoptimization

Each of our test policies can be applied dynamically by resolving the models sequentially over time. Although re-solving the models is not guaranteed to improve their performance (see Cooper (2002) and Secomandi (2003)), this is the most common way such models are applied in practice. Hence, these example represent the best test of the practical performance of the methods.

We consider \(K\) “update” periods, at the start of which the control parameters are updated by resolving the models based on the current state of the system. This breaks the problem into \(K\) smaller periods within the reservation period. These \(K\) reservation periods are followed by the service period. The cancellations from one update period to the next follow
a Binomial distribution. The demand for each product class arrives according to a Poisson distribution.

The sequence of events in an update period is as follows: At the beginning of a reservation period, policy control parameters are computed based on the state of the system. Then, reservation requests arrive one at a time during the period. They are accepted or rejected based on the policy parameters. The reservations on-hand at the end of the current period may be cancelled before the next update period. If the current update period is followed by the service period, then reservations that are not cancelled become the survivors; otherwise, the clock moves to the next reservation period.

While we did not allow for cancellations during the reservation period in our original formulation, in these examples we test all the policies in a more realistic setting where existing reservations may cancel from one update period to the next.

**Example 2.** This represents a small airline network with 2 legs, $A$ to $B$ and $B$ to $C$. There are 3 itineraries on the network: $AB$, $BC$ and $ABC$. We use $K = 5$ update periods. The probability that a reservation in an update period survives to the next period is 0.9. The overbooking cost per leg is fixed at $2000$. The following factors were varied for each example to create a number of experiments:

- Price differentials: The itinerary fares vary from $100$ to $1000$ (which may lead to significantly different overbooking limits for the ad hoc methods).

<table>
<thead>
<tr>
<th>Approach $(\vartheta)$</th>
<th>PA Policy</th>
<th>BP Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p_i = 0.6$</td>
<td>$p_i = 0.75$</td>
</tr>
<tr>
<td>DET-1</td>
<td>2.21</td>
<td>4.48</td>
</tr>
<tr>
<td>DET-2</td>
<td>6.48</td>
<td>1.16</td>
</tr>
<tr>
<td>OBC-1</td>
<td>3.47</td>
<td>0.93</td>
</tr>
<tr>
<td>OBC-2</td>
<td>3.47</td>
<td>0.93</td>
</tr>
<tr>
<td>OBC-3</td>
<td>2.50</td>
<td>1.89</td>
</tr>
<tr>
<td>OBC-4</td>
<td>3.47</td>
<td>0.93</td>
</tr>
<tr>
<td>POI</td>
<td>0.03</td>
<td>1.61</td>
</tr>
<tr>
<td>BIN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R-POI</td>
<td>$--$</td>
<td>$--$</td>
</tr>
<tr>
<td>R-BIN</td>
<td>$--$</td>
<td>$--$</td>
</tr>
</tbody>
</table>

Table 1: Relative suboptimality of heuristics in Example 1
• Stationary vs. non-stationary demand: We generate the demand based on two rules: either the demand for the itineraries arrives according to a time-homogeneous process, or the local traffic (AB and/or BC) arrives before the network-through traffic ABC.

• Demand balance: We choose demand parameters such that the aggregate demand for the legs vary as well as the intensity of local traffic vs. network traffic. For instance, suppose leg AB has high demand. This may happen because (i) the local traffic AB is high while network traffic ABC is low, (ii) the network traffic ABC is high while local traffic AB is low, (iii) both local and network traffic are high. We prepared experiments to test the effect of each of such cases.

• Load Factor: We define the load factor as the ratio of the expected demand net of cancellations to the leg capacities. The leg capacities vary from 40 to 110 in our experiments. We use moderate load factors where the leg level load factors are between 1.0 to 1.2. The mean demand of an itinerary in a period is chosen based on the load factors, as well as the demand balance conditions mentioned above.

Based on the above factors, we created 51 instances with moderate load factors and different fares, capacities and demand intensities. Then, we used these instances in experimenting with both stationary and non-stationary demand. We tested PA policies and BP policies separately. The results are summarized in Tables 2 and 3.

From Table 2, note the average performance of PA policies is not significantly different from each other. However, BIN stands out as the most robust; it is within 0.66% of the best heuristic on the average in all the experiments.

For the BP policies, the performance is more variable. BIN and R-BIN have the lowest average relative suboptimality. Looking at the worst relative performance of the BP policies, we see that except for the approaches that use randomization (R-POI, R-BIN), the maximum relative suboptimality is as high as 12-16%. BIN is robust as a BP policy as well; it has the lowest average relative suboptimality.

The robustness of BIN is more evident when we look at the detailed analysis in Table 3. The figures in this table show the “fractile values” of relative suboptimality of the heuristics in the designed experiment. For instance, in 10% of the experiments, DET-2, OBC-1, OBC-2, OBC-4 and BIN have a relative suboptimality of 0% as PA policies. In fact, BIN provides the highest average net revenue among the PA policies for at least 50% of the experiments.
<table>
<thead>
<tr>
<th>Approach ((\theta))</th>
<th>(\text{PA Policy})</th>
<th>(\text{BP Policy})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Maximum</td>
</tr>
<tr>
<td>DET-1</td>
<td>1.45</td>
<td>5.74</td>
</tr>
<tr>
<td>DET-2</td>
<td>0.88</td>
<td>4.06</td>
</tr>
<tr>
<td>OBC-1</td>
<td>1.23</td>
<td>5.74</td>
</tr>
<tr>
<td>OBC-2</td>
<td>0.5</td>
<td>3.39</td>
</tr>
<tr>
<td>OBC-3</td>
<td>1.93</td>
<td>5.74</td>
</tr>
<tr>
<td>OBC-4</td>
<td>0.44</td>
<td>2.05</td>
</tr>
<tr>
<td>POI</td>
<td>1.86</td>
<td>5.33</td>
</tr>
<tr>
<td>BIN</td>
<td>0.12</td>
<td>0.66</td>
</tr>
<tr>
<td>R-POI</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>R-BIN</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 2: Relative suboptimality of heuristics in Example 2

<table>
<thead>
<tr>
<th>Approach</th>
<th>Fractiles for PA Policies</th>
<th>Fractiles for BP Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta)</td>
<td>(10^{th})</td>
<td>(30^{th})</td>
</tr>
<tr>
<td>DET-1</td>
<td>0.08</td>
<td>0.61</td>
</tr>
<tr>
<td>DET-2</td>
<td>0</td>
<td>0.12</td>
</tr>
<tr>
<td>OBC-1</td>
<td>0</td>
<td>0.59</td>
</tr>
<tr>
<td>OBC-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OBC-3</td>
<td>0.11</td>
<td>1.39</td>
</tr>
<tr>
<td>OBC-4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>POI</td>
<td>0.02</td>
<td>0.58</td>
</tr>
<tr>
<td>BIN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R-POI</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>R-BIN</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 3: Fractile values for relative suboptimality (%) of the heuristics in Example 2
Its relative suboptimality is 0.48% in 90% of the experiments. Similarly, as a BP policy, BIN is the best in at least 30% of the experiments, and has a relative suboptimality of 0.61% in at least 70% of the experiments. Even though its worst-case relative suboptimality is 12%, its average net revenues are within 3.6% of the highest in at least 90% of the experiments. R-BIN is the only other approach that is superior to BIN as a BP policy.

**Example 3.** This example uses the same network as in Example 1 which has 4 legs and 7 itineraries. There is a single fare class for each itinerary. In the experiments, the capacity of each leg is set to 100, the overbooking penalty for each itinerary and leg is $1300. We use $K = 4$ update periods. The survival probabilities are $\{0.75, 0.9\}$. All itineraries have the same cancellation probability in a given experiment. We are interested in understanding the effect of “network parameters” on the performance of the heuristics. In order to do that, we partition the fare classes into two sets where products use a single leg (local traffic) or two legs (network traffic). Then we vary the demand such that (i) network traffic is high, (ii) local traffic is high or (iii) local and network traffic load is balanced. The demand is Poisson distributed in all instances and the mean demand per period is chosen based on the demand scenario. Single leg itinerary fares are chosen between $100 and $250 and two-leg itinerary fares range from $200 to $350. To be practical, we limit the fare choices such that the fare of a two-leg itinerary is no less than the fare of each of the legs it uses and no more than the sum of the fares of the legs. In this experiment, we used high load factors to show the effect of overbooking decisions more clearly. A total of 48 experiments are used.

The results are summarized in Tables 4 and 5. Note BIN achieves the lowest average suboptimality in this example among PA and BP policies. Its worst case performance is very good; its relative suboptimality did not exceed 0.29% as a PA policy and 0.1% as a BP policy in any of the experiments. The worst case performance of POI is also very strong compared to other BP policies. While the highest relative suboptimality of other approaches ranges from 16% to 19%, the procedures that use our two-stage approach have maximum relative suboptimality gaps of only 5.59%. This differences in worst-case performance of the approaches indicate the robustness of our two-stage approximation approach.
<table>
<thead>
<tr>
<th>Approach (ε)</th>
<th>( \text{PA Policy} )</th>
<th>( \text{BP Policy} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Maximum</td>
</tr>
<tr>
<td>DET-1</td>
<td>6.23</td>
<td>10.31</td>
</tr>
<tr>
<td>DET-2</td>
<td>1.72</td>
<td>6.23</td>
</tr>
<tr>
<td>OBC-1</td>
<td>0.66</td>
<td>2.26</td>
</tr>
<tr>
<td>OBC-2</td>
<td>0.46</td>
<td>1.96</td>
</tr>
<tr>
<td>OBC-3</td>
<td>2.28</td>
<td>5.24</td>
</tr>
<tr>
<td>OBC-4</td>
<td>0.64</td>
<td>2.26</td>
</tr>
<tr>
<td>POI</td>
<td>2.61</td>
<td>5.07</td>
</tr>
<tr>
<td>BIN</td>
<td>0.01</td>
<td>0.29</td>
</tr>
<tr>
<td>R-POI</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>R-BIN</td>
<td>- -</td>
<td>- -</td>
</tr>
</tbody>
</table>

Table 4: Relative suboptimality of heuristics in Example 3

<table>
<thead>
<tr>
<th>Approach ( \vartheta )</th>
<th>Fractiles for PA Policies</th>
<th>Fractiles for BP Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10(^{th})</td>
<td>30(^{th})</td>
</tr>
<tr>
<td>DET-1</td>
<td>4.85</td>
<td>5.63</td>
</tr>
<tr>
<td>DET-2</td>
<td>0.38</td>
<td>0.65</td>
</tr>
<tr>
<td>OBC-1</td>
<td>0.13</td>
<td>0.28</td>
</tr>
<tr>
<td>OBC-2</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>OBC-3</td>
<td>1</td>
<td>1.64</td>
</tr>
<tr>
<td>OBC-4</td>
<td>0.13</td>
<td>0.27</td>
</tr>
<tr>
<td>POI</td>
<td>1.09</td>
<td>1.67</td>
</tr>
<tr>
<td>BIN</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R-POI</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>R-BIN</td>
<td>- -</td>
<td>- -</td>
</tr>
</tbody>
</table>

Table 5: Fractile values for relative suboptimality (%) of the heuristics in Example 3
7.3.1 Summary of numerical results

These example networks, tested under various (and possibly extreme) conditions, indicate that coordinating the overbooking and capacity control decisions using our two-stage approximation method provides robust and consistent performance. Although the ad hoc approaches included in our experiments are very practical and achieve very good results on certain problem instances, their performance is much more sensitive to problem parameters. In contrast, BIN has a remarkably consistent performance as a PA policy; its relative suboptimality is no more than 0.48% and it provides the highest average net revenues in the majority of the experiments. When the load factors are high, BIN is also superior to other approaches as a BP policy. In Example 3, its highest relative suboptimality is only 0.1%. Despite its analytical tractability and theoretical properties, POI does not usually match BIN’s performance in numerical examples. Still, the effect of coordinated decisions is visible in POI in Example 3 where its highest relative suboptimality is significantly better than the other ad hoc approaches.

Overall, the results on BP policies show more volatility and the relative suboptimalities are higher among BP policies. Note that in Example 2 we used moderate load factors. While BIN had good performance on the average in that example, its worst case relative suboptimality is still about 12%. In that example, R-BIN and R-POI provide drastic improvements over BIN and POI, especially in terms of their worst-case performance. This is consistent with intuition, since BIN and POI suffer from not overbooking when the mean demand is low. The benefit of using RLP is less visible when the load factors are high and the relative suboptimality of R-BIN and R-POI are no better than that of BIN and POI in Example 3. Still, using RLP appears to be a more robust approach overall.

8. Conclusion

In this paper, we looked at a simple approximate approach for combining capacity control and overbooking decisions for an airline network. Our approach uses a two-stage approximation method to determine the virtual capacities and to control the network capacity. It provides a practical and intuitive way to combine commonly used overbooking and network capacity control models. For the special case where the economic model of overbooking is combined with the DLP model, we showed a control policy derived from this model is asymptotically optimal in a fluid scaling of the problem. We also provided an efficient algorithm to determine
the optimal policy parameters based on the alternating direction method of multipliers, which is especially well-suited to the form of our optimization problem. Our numerical study showed that policies based on our approach are generally more robust than simpler heuristics representative of those used in practice.

While these results are promising, it appears that our approximation method is somewhat sensitive to the choice of both the overbooking and network control model. This is to be expected, however, as the method requires feedback (in the form of bid prices) from the capacity allocation model when computing overbooking limits and feedback in terms of virtual capacities from the overbooking model when computing capacity allocations. For example, while the Poisson approximation to the binomial has nice theoretical properties, the binomial model itself generally performs better in numerical experiments. Also, the DLP capacity control model does not lead to overbooking on a leg unless mean capacity is strictly greater than demand, which is an undesirable behavior that is most visible when load factors are lower. Our preliminary results suggest that randomized linear programming is somewhat better-behaved in this regard and performs better when the load factors are low.

Even though we only discussed the use of static approximations - DLP, PNLP, RLP - to network capacity control, a wide variety of network capacity control models could be embedded in our two-stage approach. Further, overbooking models which allow substitution among resources in the event of oversales, such as that proposed and analyzed in Karaesmen and van Ryzin (2004), can be incorporated in the overbooking cost function. Since the two-stage approach separates overbooking from capacity control, it can be used with more realistic and accurate representations of either problem.

Acknowledgements

We thank Aydin O. Balkan for his assistance in developing the optimization codes and the simulator used in our computational experiments.

Appendix

The proofs of the results in Section 5.2 are given below.
Proof of Lemma 1

The first inequality holds since policy PA may not be the optimal policy for the two-stage problem. The second inequality follows from \( N^{PA} \leq x^* \), and the relations
\[
E_D[R'(N^{PA})] = E_D[r \cdot N^{PA} - C(AN^{PA})] \geq E_D[r \cdot N^{PA} - C(Ax^*)] = E_D[r \cdot N^{PA}] - C(Ax^*)
\]
hold since the overbooking cost function is non-decreasing.

Proof of Lemma 3

First we prove that \((1/k)LB^k\) is bounded:

\[
|LB^k| = |E_D[r \cdot N^{PA(k)}] - C^k(Ax^k)| \\
\leq |E_D[r \cdot N^{PA(k)}]| + |C^k(Ax^k)| \\
= |E_D[r \cdot \min(D^k, x^k)]| + |C^k(Ax^k)| \\
\leq r \cdot E_D[D^k] + \sum_{i=1}^{m} q_i E[Z_i(\sum_{j=1}^{n} a_{ij} x_j^k)] \\
\leq \sum_{j=1}^{m} r_j (k \mu_j) + \sum_{i=1}^{m} q_i E[Z_i(\sum_{j=1}^{m} a_{ij} k \mu_j)] \\
\leq \sum_{j=1}^{m} r_j (k \mu_j) + \sum_{i=1}^{n} q_i \sum_{j=1}^{k \mu_j}.
\]

Therefore,
\[
|(1/k)LB^k| \leq \sum_{j=1}^{m} r_j \mu_j + \sum_{i=1}^{n} \sum_{j=1}^{m} q_i \mu_j < \infty.
\]

Boundedness of \( \hat{\nu}^k \) follows from the properties of the optimal solution to (13) (see Section 5.1, Propositions 2 and 3), and can be shown similarly.

Proof of Lemma 4

For a random variable \( D \) with mean \( \mu \) and standard deviation \( \sigma \), and for any real number \( \rho \), we have (see Gallego and van Ryzin, (1997)):
\[
E[(D - \rho)^+] \leq (1/2)\sigma + (\mu - \rho)^+.
\]

For the scaled problems, the mean demand is \( k\mu \), and the variance is \( k\sigma^2 \). Using these:
\[
\hat{\nu}^k - LB^k = r \cdot x^k - C^k(u^k) - (r \cdot E_D[N^{PA(k)}] - C^k(Ax^k))
\]

32
\[
\begin{align*}
&= r \cdot x^{*k} - r \cdot E_D[N^{PA(k)}] - (C^k(u^{*k}) - C^k(Ax^{*k})) \\
&\leq r \cdot x^{*k} - r \cdot E_D[\min(x^{*k}, D^k)] \\
&= r \cdot E_D[(x^{*k} - D^k)^+] \\
&\leq \sum_{j=1}^{m} r_j[(1/2)\sigma_j\sqrt{k} + (x^{*k}_j - k\mu_j)^+] \\
&= \sum_{j=1}^{m} r_j(1/2)\sigma_j\sqrt{k}.
\end{align*}
\]

The last term follows from the fact that \(x^{*k}_j \leq k\mu_j\). Therefore \((1/k)[\tilde{\nu}^k - LB^k] \to 0\) as \(k \to \infty\).

References


Athena Scientific, Belmont, Massachusetts.


