Single-Leg Airline Revenue Management With Overbooking

Nurşen Aydın, Ş. İlker Birbil, J. B. G. Frenk and Nilay Noyan
Sabancı University, Manufacturing Systems and Industrial Engineering, Orhanlı-Tuzla, 34956 Istanbul, Turkey

Abstract: Airline revenue management is concerned with identifying the maximum revenue seat allocation policies. Since a major loss in revenue results from cancellations and no-shows, overbooking has received a significant attention in the literature over the years. In this study, we propose new static and dynamic single-leg overbooking models. In the static case we introduce two models; the first one aims to determine the overbooking limit and the second one is about finding the overbooking limit and the booking limits to allocate the virtual capacity among multiple fare classes. Since the second static model is hard to solve, we also introduce computationally tractable models that give upper and lower bounds on its optimal expected net revenue. In the dynamic case, we propose a dynamic programming model, which is based on two streams of events. The first stream corresponds to the arrival of booking requests and the second one corresponds to the cancellations. We conduct simulation experiments to illustrate the effectiveness of the proposed models.

Keywords: Revenue management; airline; overbooking; cancellation; static model; dynamic model; dynamic programming; simulation

---

1. Introduction. Historically, airline industry plays the steering role in revenue management. This can be attributed to the quick responses of the airline executives, who have realized the importance of controlling the reservation process in order to increase their gains over a fiscal year. The main problem, then and now, in airline revenue management is to determine how to reserve the seats for the requests coming from the passengers. Naturally, the objective of this problem is to maximize the total revenue. We refer to (Talluri and van Ryzin, 2005, Section 1.2) for a historical account of the role of airline industry in revenue management.

Capacity allocation and overbooking are two main strategies used by revenue management specialists. While capacity allocation deals with reserving seats for different fare classes, overbooking is concerned with the number of additional booking requests to be accepted above the physical capacity. It is quite common that a certain percentage of the accepted requests cancel before the departure time (cancellations) or do not show-up at the departure time (no-shows). Consequently, the capacity becomes available for boarding the overbooked passengers. Thus, overbooking is used by the airline companies to protect themselves against vacant seats due to no-shows and late cancellations. On the other hand, it may also happen that some of the reservations are denied boarding due to the lack of capacity at the departure time. In such a case, the airline faces penalties like monetary compensations, and even worse, suffers from bad public relations. Even though the overbooking decision involves uncertainties regarding the no-shows and cancellations, accepting more booking requests than the available capacity is still a commonly-used, profitable strategy because the revenue collected by overbooking usually exceeds the penalties for denied boardings (Rothstein, 1985). The overbooking limit, which is also referred to as virtual capacity or total booking limit, is the maximum number of booking requests an airline company is willing to accept. An allocation policy specifies how to allocate this virtual capacity to each fare class. Although a common practice is first setting the virtual capacity and then doing the allocations (c.f. (Belobaba, 2006)), this heuristic approach in fact undermines the effects of these two decisions on each other. Therefore, it is natural to study the joint capacity allocation and overbooking problem which is, in general, difficult to solve largely because of the uncertainty in the class dependent no-show and cancellation parameters.

It is well known that many airline companies are interested in managing their revenues over a network of flights. However, solving single-leg problems is still crucial because (i) the network based seat allocation problems are quite difficult to solve, and hence, in practice, the methods that require solving a series of single-leg problems are frequently applied; (ii) some small airline companies, like charter flight companies commonly seen in Europe, accept booking requests only for single-leg itineraries.

Roughly speaking, in a static model one does not consider the dynamics of the stochastic processes representing the booking requests and the cancellations over time. On the other hand, a dynamic model accounts for the behavior of the system over time. In the remaining part of this paper, we propose new mathematical programming models for static and dynamic single-leg problems that involve no-shows, cancellations, and hence, overbooking. Our first static model focuses on finding the total overbooking limit for multiple classes under the assumption that the fare class requests are accepted as long as the total number of reservations is below the total booking limit. This model allows for class dependent cancellations and no-shows. We discuss that the proposed model is general and the resulting problem can be solved to optimality efficiently. To the best of our knowledge, our model is a first in the literature in determining the optimal total booking limit under this broad setting. As a by-product of our approach, we also discover that a well-known heuristic from the literature finds the optimal overbooking limit whenever the particular parameters dictated by our analysis are used. In the second static model, which also considers the class dependent no-shows and cancellations, we determine simultaneously the total booking limit and the partitioned allocation of the virtual capacity to each fare class. Arriving at a computationally difficult model, we propose upper and lower bounding problems to obtain approximate solutions, which have demonstrated promising performance in our computational study. Our last model involves a dynamic setting based on two independent streams of events; arrivals of booking requests and cancellations. Contrary to the static case, in the dynamic setting we deal with the class independent show-ups and cancellations. The proposed model, therefore, can be used as a heuristic in practice for the actual model with class dependent processes. We show that it is easy to solve the resulting problem with dynamic programming. After characterizing the optimal policy, we also present the nested structure of the optimal allocations.

The rest of the paper is organized as follows. Section 2 gives the literature review on static and dynamic overbooking models. We introduce our static models in Section 3. This is followed by the dynamic model in Section 4. We present our computational study in Section 5 and conclude the paper in Section 6.

2. Literature Review The early overbooking literature concentrates mainly on static models with one or two fare classes and the objective of finding the overbooking limit. The first scientific work on overbooking is proposed by Beckman (1958). Beckman proposes a static single fare class overbooking model, which determines the overbooking limit by considering the trade-off between the lost revenue due to empty seats at the departure, the total cost of denied boardings and the revenue generated by the go-show passengers. The go-shows are the passengers who show up without any reservation at the departure time. American Airlines adopted Beckman's approach and implemented a related model in 1976 and then revised it in 1987 (Smith et al., 1992). Beckman's work is succeeded by Thompson (1961), who considers a practical model ignoring the probability distribution of demand and requiring only data on the number of cancellations among the total number of reservations. Given the capacities for two fare classes, Thompson aims at determining the overbooking amount for each fare class so that the probability of overbooking equals to a specified value. He also supports his arguments by a statistical analysis of the involved distributions. The works of Beckman and Thompson are refined by Taylor (1962). Like Thompson, Taylor focuses on a service measure by constraining the number of denied boardings but considers cancellations, no-shows and group sizes explicitly. This influential work of Taylor has attracted the attention of various airlines. Consequently, the variants of this work are implemented, and then, reported in a sequence of papers. The references and the details of this history are given by Rothstein (1985).

In the first part of his thesis, Chi (1995) studies a static overbooking problem with multiple fare classes and formulates it as a dynamic programming model. However, when cancellations and no-shows are considered, the model suffers from the curse of dimensionality because one needs to keep track of the number of reservations for each class. To overcome this difficulty, Chi proposes an approximate model that can be solved in polynomial time. Coughlan (1999) also considers a overbooking problem with multiple fare classes, but he assumes that the go-show passengers are given the empty seats at the same price as in (Beckman, 1958). Unlike the majority of the studies in the literature, Coughlan does not use a Poisson distribution to model the demand but makes the simplifying assumption that both the demand and the number of bookings for each fare class are independent and normally distributed. Coughlan's discussion also supposes implicitly that the minimum of the demand and the number of bookings is also normally distributed; unfortunately, this supposition does not hold mathematically in general. Overall, the author provides a closed form formula for the revenue function and applies heuristic search methods to find a maximizer.

Several researchers have addressed dynamic overbooking models for single-leg problems. Generally, the dynamic overbooking problem is modeled as a Markov Decision Process (MDP). Rothstein (1971) proposes two such models, where only one fare class is considered. In the first model, the objective is to find the optimal expected revenue after deducting the cost of denied boardings. Following the work of Thompson (1961), the second model adds a constraint to limit the proportion of denied boardings. Alstrup et al. (1986) also use a MDP to solve an overbooking model but this time with two fare classes and the cost of downgrading (a cost that is incurred due to reserving cheaper seats for the passengers requesting more expensive fare classes). In the second part of his thesis, Chi (1995) discusses two dynamic models with multiple fare classes. Although the first model incorporates the realistic setting of cancellations occurring in time, it is computationally intractable. To ease the computational burden, Chi then assumes in his second model that the cancellations occur right before the departure time. This assumption allows him to solve the resulting model with an approximation similar to the one he uses in the static case. Chatwin (1998) analyzes the optimal solution structure of a discrete time dynamic single fare class overbooking model and discusses the conditions, under which a booking limit policy is optimal. Subramanian et al. (1999) study a more general setting than Chatwin, where they analyze the overbooking problem with multiple fare classes. The authors consider the arrival of a cancellation, the arrival of a booking request and no arrival of any type as a combined stream and assume that at most one of these events can occur at any discrete time epoch. Under this setting they present two models. In the first model, the cancellation and no-show probabilities do not depend on the fare classes. They show that the resulting problem can be equivalently modeled as a queuing system discussed in the literature (Lippman and Stidham, 1977). In their second model, they relax the class independence assumption and model a more general problem with class dependent cancellations and no-shows. Unfortunately, the resulting dynamic programming formulation cannot be solved efficiently because of the high-dimensional state space. Chatwin (1999) examines a continuous-time single fare class overbooking problem, where fares and refunds vary over time according to piecewise constant functions. In his model the arrival process of requests is assumed to be a homogeneous Poisson process, and the probabilities to identify the type of a request are independent of time. He assumes that the reservations cancel independently according to an exponential distribution with a common rate, and the arrival process of requests depends on the number of reservations. Under these assumptions, the author formulates the problem as a homogeneous birth-and-death process and shows that a piecewise constant overbooking limit policy is optimal. A closely related study is given by Feng et al. (2002). They consider a continuous-time model with cancellations and no-shows. They derive a threshold type optimal control policy, which simply states that a request should be admitted only if the corresponding fare is above the expected marginal seat revenue (EMSR). Karaesmen and van Ryzin (2004) examine the overbooking problem differently. Their model permits that fare classes can substitute for one another. They formulate the overbooking model as a two-period optimization problem. In the first period the reservations are made by using only the probabilistic information of cancellations. In the second period, after observing the cancellations and no-shows, all the remaining customers are either assigned to a reserved seat or denied by considering the substitution options. They give the structural properties of the overall optimization problem, which turns out to be highly nonlinear. Therefore, they propose to apply a simulation based optimization method using stochastic gradients to solve the problem.

In all of the above models probability distributions are used to model uncertainty in demand and cancellations. Recent studies in revenue management focus on the availability of information. Adaptive methods are used when there exists no or limited information about the demand. Most of these methods assume that there is access only to samples from demand distributions. They mainly compute the booking limits based on the past information but also react to the possible inaccuracies related to the estimates of demand (van Ryzin and McGill, 2000; Huh and Rusmevichientong, 2006). Kunnumkal and Topaloglu (2009) consider a capacity allocation problem with limited demand information and develop a stochastic approximation method to compute the optimal protection levels iteratively. They prove that the sequence of protection levels computed by using their approach converge to the optimal ones. Birbil et al. (2009) present a robust version of static and dynamic single leg problems. In their model, they take into account the inaccuracies associated with the estimated probability distributions of the demand for different fare classes. Ball and Queyranne (2009) use online algorithms to treat also a robust problem. In this way, they eliminate the need for estimating the demand and present the closed-form optimal booking limits. Lan et al. (2008) generalize Ball and Queyranne's model by assuming that the demand for each fare class lies in a given interval. By using relative regret and absolute regret as performance criteria, they provide two capacity allocation models which differ in their objective functions. They show that these two models can be analyzed in a unified manner and both models provide nested booking limits. In a related work, Lan et al. (2011) formulate a joint overbooking and seat allocation model, where both the random demand and no-shows are characterized using interval uncertainty. They focus on the seller's regret in not being able to find the optimal policy due to the lack of information. The regret of the seller is quantified by comparing the net revenues associated with the policy obtained before observing the actual demand and the optimal policy obtained under perfect information. The model aims to find a policy which minimizes the maximum relative regret.

In the present study, we develop new static and dynamic overbooking models and their associated solution methods. In the static case we discuss two models both of which allow class dependent cancellations and no-shows. The first model can be seen as a generalization of the single fare class model discussed in Phillips (2005). The second static model aims at determining both the total booking limit and the partitioned allocation of the virtual capacity to each fare class. We then propose a discrete-time dynamic model based on independent streams of arrivals of booking requests and cancellations. Our modeling approach differs from the one based on a combined stream of events (Subramanian et al., 1999) by allowing the arrival and cancellation processes to be independent. In particular, we assume that requests for different fare classes arrive according to independent nonhomogeneous Poisson processes. Moreover, the number of cancellations in any time period, given that there are n number of accepted requests at the beginning of that time period, is a binomially distributed random variable with n independent trials and a period-dependent cancellation probability. Thus, as desired, the arrival process of the booking requests are independent of the number of reservations whereas the cancellation and no-show probabilities depend on the total number of reservations.

3. Static Overbooking Models. In this section, we propose two static risk-based overbooking models and analyze them in-depth to obtain efficient solution methods. The risk-based models try to determine a policy considering the trade-off between the potential revenue from accepting an additional request and the cost of an additional denied service. The objective of our first static model is to find the optimal booking limit maximizing the expected net revenue under the assumption that the greedy policy—that is, a request for any fare class is accepted as long as the total number of reservations is below the overbooking limit— is applied. In this model, the probabilistic information comes from the aggregated demand for all fare classes. However, we assume that each booking request belongs to a fare class with a certain probability. Finding the optimal total booking limit in this way is useful in practice, since the overbooking limit can be used as an input to some well-known allocation methods. This kind of heuristic approach first determines the total booking limit and then uses one of the well-known capacity allocation methods, like the famous EMSR heuristics (Belobaba, 1987, 1989), to calculate the nested protection levels for different fare classes. In our second model, on the other hand, the probabilistic information is related to the demand for each fare class. We try to determine both the total booking limit and the partitioned allocation of the virtual capacity to each fare class in such a way that the expected net revenue is maximized. Since the second static model is quite hard to solve, we introduce two computationally tractable models that give upper and lower bounds on the proposed model's optimal expected net revenue.

In the subsequent discussion, we consider a flight with a known seat capacity of C and do not assume that the booking requests for different fare classes arrive in a certain order. In the first model, the booking requests for m different fare classes are accepted until the total booking limit $b \geq C$ is reached, whereas in the second model the booking decisions are based on the capacity allocated to each fare class. An accepted request becomes a reservation and a reservation may cancel at any time before departure or may not show up without cancelling. Let $\beta_i^s > 0$ denote the probability that an accepted fare class i request shows up at the departure time. For the remaining fare class i reservations, if we denote the probability of a cancellation by δ_i, then a fare class i reservation cancels with probability $\beta_i^c := (1-\beta_i^s)\delta_i$. We assume that a fare class i cancellation is refunded a percentage α_i of the corresponding ticket price r_i, and no-shows do not receive any refund. If the number of shows exceeds the capacity C, then exactly C shows will be on the flight and the rest will be denied boarding. For each denied service, the airline incurs a denied service cost of $\theta > 0$. We refer the interested reader to (Chatwin, 1999) for a discussion on fare class independent compensation for a denied boarding. Aside from this notation, the random variables and the vectors are denoted by uppercase and lowercase boldface letters, respectively. If $\mathbf{X}$ and $\mathbf{Y}$ are random variables, then $\mathbf{X} =^d \mathbf{Y}$ means that the cumulative distribution functions of $\mathbf{X}$ and $\mathbf{Y}$ are identical. To simplify the exposition, we also denote $\max\{x, 0\}$ by $[x]^+$.

3.1 Total Booking Limit. In this section, we propose a model to determine the optimal total booking limit $b \geq C$. We consider a model, where the probabilistic information is the random total booking requests, and denote this non-negative integer valued random variable by $\mathbf{D}$. We assume that each booking request belongs to a certain fare class according to a multinomial selection mechanism with given probabilities. These probabilities can be estimated using historical data about the overall market share of each fare class. In particular, each arriving request is for fare class i with probability p_i, $i = 1, \ldots, m$. Clearly, $p_i \geq 0$ and $\sum_{i=1}^m p_i = 1$. Thus, we assume that the random fare class i demand, denoted by $\mathbf{D}_i$, has a binomial distribution with $\mathbf{D}$ independent trials and the success probability of p_i (see Appendix A for an introduction to the Bernoulli selection scheme). We consider the greedy policy of accepting a booking request for any fare class as long as the total booking limit b is not reached. Under this policy the random total number of reservations is given by $\mathbf{N}(b) := \min\{b, \mathbf{D}\}$. Let $\mathbf{B}(p, k)$ denote a binomially distributed random variable with k independent trials each having a success probability of p and $\mathbf{D}_i^r$ designate the random number of reservations for fare class i. Since our policy accepts any request until the booking limit is reached, it is easy to prove the following lemma, which implies that the joint distribution of the random vector $(\mathbf{D}_1^r, \ldots, \mathbf{D}_m^r)$ follows a multinomial distribution with $\mathbf{N}(b)$ independent trials and the success probabilities $p_i,\ i = 1, \ldots, m$.

Lemma 3.1 *Under the greedy policy, it follows that* $\mathbf{D}_i^r =^d \mathbf{B}(p_i, \mathbf{N}(b))$.

Proof. Let $\mathbf{D}_i^r$ denote the random number of fare class i reservations. By the definition of the total booking limit b and the used policy, we obtain for every integer k satisfying $k \leq b-1$ and $y \leq k$ that

$$\mathbb{P}(\mathbf{D}_i^r = y \mid \mathbf{N}(b) = k) = \mathbb{P}(\mathbf{D}_i^r = y \mid \mathbf{D} = k) = \binom{k}{y} p_i^y (1-p_i)^{k-y}. \tag{1}$$

It also follows for every $y \leq b$ that

$$\begin{aligned}\mathbb{P}(\mathbf{D}_i^r = y \mid \mathbf{N}(b) = b) &= \mathbb{P}(\mathbf{D}_i^r = y \mid \mathbf{D} \geq b) = \tfrac{\mathbb{P}(\mathbf{D}_i^r = y, \mathbf{D} \geq b)}{\mathbb{P}(\mathbf{D} \geq b)} \\ &= \tfrac{\sum_{k=b}^{\infty} \mathbb{P}(\mathbf{D}_i^r = y, \mathbf{D} = k)}{\mathbb{P}(\mathbf{D} \geq b)} = \tfrac{\sum_{k=b}^{\infty} \mathbb{P}(\mathbf{D}_i^r = y \mid \mathbf{D} = k)\mathbb{P}(\mathbf{D} = k)}{\mathbb{P}(\mathbf{D} \geq b)} \\ &= \tfrac{\sum_{k=b}^{\infty} \binom{b}{y} p_i^y (1-p_i)^{b-y} \mathbb{P}(\mathbf{D} = k)}{\mathbb{P}(\mathbf{D} \geq b)} = \binom{b}{y} p_i^y (1-p_i)^{b-y}.\end{aligned} \tag{2}$$

Applying now relations (1) and (2) yields the desired result. □

As discussed at the beginning of Section 3, we distinguish between a no-show and a cancellation to obtain an explicit expression of the revenue obtained from each reservation. By Lemma 3.1 and the properties of the Bernoulli selection mechanism as discussed in Appendix A, the random number of fare class i shows and fare class i cancellations are given by $\mathbf{B}(\beta_i^s p_i, \mathbf{N}(b))$ and $\mathbf{B}(\beta_i^c p_i, \mathbf{N}(b))$, respectively, (c.f. (Thompson, 1961; Chatwin, 1998; Coughlan, 1999) for similar uses of the Bernoulli selection scheme).

Hence, for a given booking limit b the random total revenue generated by any fare class i reservation is given by

$$r_i\mathbf{B}(p_i, \mathbf{N}(b)) - \alpha_i r_i \mathbf{B}(\beta_i^c p_i, \mathbf{N}(b)),$$

where $\alpha_i r_i$ denotes the refund paid for a fare class i cancellation. Introducing now

$$\tau_i = r_i(1 - \alpha_i \beta_i^c), \qquad i = 1, \ldots, m, \tag{3}$$

the expected total revenue over all reservations becomes

$$\sum\nolimits_{i=1}^{m} p_i \tau_i \mathbb{E}(\mathbf{N}(b)). \tag{4}$$

To incorporate the penalty cost of overbooking, we first observe adding up all the shows that the total number of denied boardings equals

$$\left[\sum\nolimits_{i=1}^{m} \mathbf{B}(\beta_i^s p_i, \mathbf{N}(b)) - C\right]^+.$$

Since the binomial random variables $\mathbf{B}(\beta_i^s p_i, \mathbf{N}(b)),\ i = 1, \ldots, m$, arise within a multinomial selection experiment with independent trials from the same population, we obtain

$$\left[\sum\nolimits_{i=1}^{m} \mathbf{B}(\beta_i^s p_i, \mathbf{N}(b)) - C\right]^+ =^d \left[\mathbf{B}\left(\sum\nolimits_{i=1}^{m} \beta_i^s p_i, \mathbf{N}(b)\right) - C\right]^+. \tag{5}$$

Then, using relations (4) and (5) the expected net revenue is obtained as

$$\psi(b) := \sum\nolimits_{i=1}^{m} p_i \tau_i \mathbb{E}(\mathbf{N}(b)) - \theta \mathbb{E}\left(\left[\mathbf{B}\left(\sum\nolimits_{i=1}^{m} \beta_i^s p_i, \mathbf{N}(b)\right) - C\right]^+\right)$$

and the optimal booking limit is found by solving

$$\max\{\psi(b) : b \geq C, b \in \mathbb{Z}_+\}. \tag{P_T}$$

To analyze the global properties of the function $b \mapsto \psi(b)$, we first observe that $\psi(b) = \mathbb{E}(f(\mathbf{N}(b)))$ with $f : \mathbb{Z}_+ \to \mathbb{R}$ given by

$$f(x) = \sum\nolimits_{i=1}^{m} p_i \tau_i x - \theta \mathbb{E}\left(\left[\mathbf{B}\left(\sum\nolimits_{i=1}^{m} \beta_i^s p_i, x\right) - C\right]^+\right). \tag{6}$$

By Lemma B.2 it follows that the function $x \mapsto \mathbb{E}([\mathbf{B}(\sum_{i=1}^{m} \beta_i^s p_i, x) - C]^+)$ is discrete convex, and this implies that the function $x \mapsto f(x)$ is discrete concave. Therefore, by Lemma B.3 the optimal solution of

$$\max\{f(b) : b \geq C, b \in \mathbb{Z}_+\}$$

coincides with the optimal solution of problem (P_T). Then, by using the discrete concavity of the function f, an optimal solution to (P_T) is given by

$$b_{opt} = \inf\{b \geq C : f(b+1) - f(b) < 0\}. \tag{7}$$

Here we use the convention that the infimum of the empty set is equal to infinity. Introduce $\beta^s := \sum_{i=1}^{m} \beta_i^s p_i$ and let $\mathbf{U}_k,\ k = 1, \ldots, b+1$, be a sequence of independent standard uniformly distributed random variables. Furthermore, let $\mathbf{1}_A$ be the indicator random variable of the event A, i.e, it takes value 1 if the event A occurs, and 0 otherwise. Then, by relation (6) and the representation of a binomial distributed random variable given in (27) we obtain for every $b \geq C$ that

$$\begin{aligned} f(b+1) - f(b) &= \sum\nolimits_{i=1}^{m} p_i \tau_i - \theta \mathbb{E}\left(\mathbf{1}_{\{\mathbf{U}_{b+1} \leq \beta^s\}}\right) \mathbb{E}\left(\mathbf{1}_{\{\sum_{k=1}^{b} \mathbf{1}_{\{\mathbf{U}_k \leq \beta^s\}} \geq C\}}\right) \\ &= \sum\nolimits_{i=1}^{m} p_i \tau_i - \theta \beta^s \mathbb{P}\left(\sum\nolimits_{k=1}^{b} \mathbf{1}_{\{\mathbf{U}_k \leq \beta^s\}} \geq C\right) \\ &= \sum\nolimits_{i=1}^{m} p_i \tau_i - \theta \beta^s \mathbb{P}\left(\mathbf{B}(\beta^s, b) \geq C\right). \end{aligned}$$

This shows using $\theta \beta^s > 0$ that

$$f(b+1) - f(b) < 0 \Leftrightarrow \mathbb{P}\left(\mathbf{B}(\beta^s, b) \geq C\right) > \frac{\mu_0}{\mu_1},$$

where

$$\mu_0 = \sum_{i=1}^{m} p_i \tau_i \text{ and } \mu_1 = \theta \beta^s. \tag{8}$$

Therefore, by using (7), the optimal solution to our optimization problem becomes

$$b_{opt} = \inf \left\{ b \geq C : \mathbb{P}\left(\mathbf{B}(\beta^s, b) \geq C\right) > \frac{\mu_0}{\mu_1} \right\}. \tag{9}$$

A surprising consequence of this result is that the optimal total booking limit does not depend on the probability distribution function of the total demand $\mathbf{D}$. It is also easy to see that the optimal solution to our overbooking problem is to set $b = \infty$ when $\mu_0 - \mu_1 \geq 0$. An intuitive interpretation of this result is as follows: Since the expected net revenue per fare class i reservation is at least equal to $\tau_i - \theta \beta_i^s$, the expected net revenue per reservation is given by

$$\sum_{i=1}^{m} p_i(\tau_i - \theta \beta_i^s) = \mu_0 - \mu_1.$$

This expression being non-negative shows that for the risk-based objective, it is always profitable to accept all requests despite the overbooking cost. Thus, the total booking limit should be set to infinity. When $\mu_0 - \mu_1 < 0$, there exists a finite optimal solution $b_{opt} \geq C$.

We next provide a computationally efficient iterative method to calculate the optimal total booking limit. To determine b_{opt}, we need to evaluate iteratively for $b \geq C$ the increasing sequence

$$\gamma_b = \mathbb{P}\left(\mathbf{B}(\beta^s, b) \geq C\right).$$

For $b = C$, it is obvious that

$$\gamma_C = \mathbb{P}\left(\mathbf{B}(\beta^s, C) \geq C\right) = (\beta^s)^C.$$

Then, we obtain the recursive relation

$$\gamma_{b+1} = \gamma_b + \beta^s \mathbb{P}\left(\mathbf{B}(\beta^s, b) = C - 1\right). \tag{10}$$

Our proposed overbooking model is related to the single fare class model discussed in Section 9.3.2 of (Phillips, 2005). Actually, the optimal booking limit of our model with multiple fare classes is equal to the booking limit obtained by the risk-based overbooking model with a single fare class, where the price is μ_0/β^s, the overbooking cost is θ and the show-up probability is β^s. In Section 9.4.2 of the same book, a heuristic is proposed to determine the total booking limit for multiple fare classes by reducing the problem to a single fare class model. Basically, this method first estimates the values of the parameters associated with a representative single fare class from the fare class dependent parameters, and then, solves the resulting single fare class model. As a direct consequence of this estimation, only a heuristic method is obtained. Contrary to Phillips, we show in this paper that under a multinomial selection scheme linking the overall demand to the demand for each fare class and the policy of accepting all the requests until the total booking limit is reached, our proposed model determines the optimal total booking limit. From a different angle, we can state that our analysis provides the values of the price, show-up probability and overbooking cost parameters for which the heuristic proposed by Phillips is exact. As mentioned before, our model can be used to provide the overbooking limit to the capacity allocation heuristics like EMSR-a and EMSR-b. Since we allow class dependent show-up probabilities, our model could perform better than those standard static models that determine the total overbooking limit when the show-up probabilities do not depend on the fare classes (Phillips, 2005). We note that the performance of the proposed model depends on the accuracy of the estimation of the model parameters. Among the parameters required to determine the optimal total booking limit (see (3),(8) and (9)), we acknowledge that the parameters p_i are the most challenging to estimate due to the non-availability of proper historical data. As emphasized in (Talluri and Ryzin, 2004), typically, the data on the arrivals is incomplete and only the purchase transaction data are available. In our case, suppose that the p_i parameters associated with more expensive fare classes, and consequently the parameter μ_0 in relation (8), are overestimated. Then, this shows by relation (9) that we may end up with a higher total booking value.

We conclude this section with two further remarks: (i) The first static model in the airline revenue management literature was proposed by Beckman (1958). Beckman considers the cost minimization for a single fare class and provides a more complex analysis. He also observes that the overbooking limit decision does not depend on the demand distribution. His model can also be analyzed with our simpler approach. (ii) As it is common in the literature (Subramanian et al., 1999; Talluri and van Ryzin, 2005), the expected total denied boarding cost may be given by an increasing convex function to represent the need to offer higher levels of compensation or incur higher goodwill costs for each additional denied boarding. Given the total booking limit b, this implies that for our model the denied boarding cost equals $\mathbb{E}(c(\mathbf{N}(b)))$, where $c : \mathbb{Z}_+ \to \mathbb{R}$ is given by

$$c(x) = \mathbb{E}(g(\mathbf{B}(\sum_{i=1}^{m} \beta_i^s p_i, x) - C))$$

and $g : \mathbb{R} \to \mathbb{R}$ is an increasing convex function satisfying $g(z) = 0$ for every $z \leq 0$. Again by Lemma B.2 the function c is discrete convex, and consequently, the function $f : \mathbb{Z}_+ \to \mathbb{R}$ given by

$$f(x) = \sum_{i=1}^{n} p_i \tau_i x - c(x)$$

is discrete concave. Therefore, as in the previous model, one can show that the optimal booking limit is in the form of (7).

3.2 Booking Limits for Individual Fare Classes. In this section we focus on a model, in which the partitioned booking limits as well as the overbooking limit are determined. This modeling approach sets us apart from other methods using capacity allocation heuristics, like EMSR-a and EMSR-b (Belobaba, 1987, 1989), after setting the overbooking limit. However, it is important to note that a policy, which strictly maintains the partitioned booking limits, is rarely applied in practice because in such a dynamic setting it is clearly suboptimal to reject a higher fare class request even if there is available capacity for lower fare classes. Therefore, the partitioned booking limits are used to obtain nested booking limits or nested protection levels. Under a nested policy, higher fare classes are allowed to use all the capacity reserved for lower fare classes. From this perspective, whenever the optimal partitioned limits that are obtained in this section are used in a nested way, the resulting method becomes another heuristic but it does not require a predefined overbooking limit.

We assume that the distribution of the demand for fare class i, denoted by $\mathbf{D}_i$, is known. If b_i is the partitioned booking limit for fare class i, then the random variable $\mathbf{N}_i(b_i) = \min\{b_i, \mathbf{D}_i\}$ denotes the number of reservations for fare class i. Using our notation in the previous section, the random number of fare class i reservations that show up at the departure time and the random number of fare class i cancellations are given by $\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i))$ and $\mathbf{B}(\beta_i^c, \mathbf{N}_i(b_i))$, respectively. Since the random total number of denied boardings is equal to $[\sum_{i=1}^{m} \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C]^+$, the expected net revenue $\phi(\mathbf{b})$ for a vector $\mathbf{b} = (b_1, \ldots, b_m) \in \mathbb{Z}_+^m$ is given by

$$\phi(\mathbf{b}) = \sum_{i=1}^{m} \tau_i \mathbb{E}(\mathbf{N}_i(b_i)) - \theta \mathbb{E}\left(\left[\sum_{i=1}^{m} \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right]^+\right). \tag{11}$$

Thus, we need to solve the following problem to obtain the optimal partitioned booking limits:

$$\max\{\phi(\mathbf{b}) : \mathbf{b} \in \mathbb{Z}_+^m\}. \tag{P_I}$$

Observe that $\sum_{i=1}^{m} b_i$ defines the overbooking limit and as suggested, the problem (P_I) provides the optimal overbooking limit and the optimal partitioned booking limits simultaneously. Unfortunately, due to the expected total overbooking cost, the expected total net revenue is not separable by the fare classes and this makes it difficult to solve the optimization problem (P_I) in an efficient way. Therefore, we consider lower and upper bounding functions on the expected total overbooking cost and develop computationally efficient methods to find approximate solutions to problem (P_I).

To compute a lower bounding function on the total expected overbooking cost, we use Jensen's inequality which leads to

$$\mathbb{E}\left(\left[\sum_{i=1}^{m} \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right]^+\right) \geq \left[\mathbb{E}\left(\sum_{i=1}^{m} \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right)\right]^+ = \left[\sum_{i=1}^{m} \beta_i^s \mathbb{E}\left(\mathbf{N}_i(b_i)\right) - C\right]^+.$$

This shows by relation (11) that for every $\mathbf{b} \in \mathbb{Z}_+^m$

$$\phi(\mathbf{b}) \leq \sum_{i=1}^{m} \tau_i \mathbb{E}(\mathbf{N}_i(b_i)) - \theta \left[\sum_{i=1}^{m} \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)) - C\right]^+ := \phi_U(\mathbf{b}). \tag{12}$$

Hence, an upper bound on the optimal objective value of problem (P_I) can be obtained by solving the optimization problem

$$\max\{\phi_U(\mathbf{b}) : \mathbf{b} \in \mathbb{Z}_+^m\}. \tag{P_I^{UB}}$$

Although its objective function is not separable, it is still possible to use dynamic programming to solve the problem (P_I^{UB}). For the solution method based on dynamic programming we refer to Appendix E. Here we present a solution method based on a mixed-integer programming formulation, which is easier to follow and seems to be computationally more efficient as demonstrated by our numerical experiments.

We introduce upper bounds on the booking limits to restrict the feasible region of the problem (P_I^{UB}) and formulate it as a mixed-integer linear program. In Appendix D, we propose a method to determine the upper bounds, denoted by $M_i,\ i = 1, \ldots, m$, so that the problem (P_I^{UB}) is solved to a desired accuracy level. Utilizing the proposed method we obtain the upper bounds, and then, restrict the feasible region of the problem (P_I^{UB}) to a box by enforcing the bounding constraints $b_i \leq M_i,\ i = 1, \ldots, m$. Let us introduce the binary variables $x_{ij},\ i = 1 \ldots, m,\ j = 0, \ldots, M_i$, where $x_{ij} = 1$ and $x_{ij} = 0$ imply that $b_i = j$ and $b_i = 0$, respectively. Then, calculating the input parameters $a_{ij} := \mathbb{E}(\mathbf{N}_i(j))$ for all $i = 1 \ldots, m,\ j = 0, \ldots, M_i$, we obtain an alternate formulation of the problem (P_I^{UB}):

$$\text{maximize} \quad \sum_{i=1}^{m} \tau_i \sum_{j=0}^{M_i} a_{ij} x_{ij} - \theta w \tag{13}$$

$$\text{subject to} \quad w \geq \sum_{i=1}^{m} \beta_i^s \sum_{j=0}^{M_i} a_{ij} x_{ij} - C, \tag{14}$$

$$w \geq 0, \tag{15}$$

$$\sum_{j=0}^{M_i} x_{ij} = 1, \quad i = 1, \ldots, m, \tag{16}$$

$$x_{ij} \in \{0, 1\}, \quad i = 1, \ldots, m,\ j = 0, \ldots, M_i, \tag{17}$$

$$\sum_{j=0}^{M_i} a_{ij} x_{ij} \leq \mathbb{E}(\mathbf{D}_i), \quad i = 1, \ldots, m. \tag{18}$$

By the definition of parameters a_{ij} and constraints (16)-(17), it is guaranteed for each fare class i that $\sum_{j=0}^{M_i} a_{ij} x_{ij} = \mathbb{E}(\mathbf{N}_i(j))$ for a single $j \in \{0, \ldots, M_i\}$. Constraints (14) and (15), and the structure of the objective function (13) ensure that at the optimal solution

$$w = [\sum_{i=1}^{m} \beta_i^s \sum_{j=0}^{M_i} a_{ij} x_{ij} - C]^+.$$

Then, it is easy to see that at the optimal solution of the problem (P_I^{UB}) with additional bounding conditions, the booking limit b_i equals to j for which $x_{ij} = 1$ and $\sum_{j=0}^{M_i} a_{ij} x_{ij} = \mathbb{E}(\mathbf{N}_i(b_i))$. Since $\mathbb{E}(\mathbf{N}_i(b_i)) \leq \mathbb{E}(\mathbf{D}_i)$, constraint (18) trivially holds and it is added as a set of valid inequalities. The number of binary variables is $\sum_{i=1}^{m} M_i \leq m \max\{K_1, \ldots, K_m\}$. In practice, the number of fare classes is a reasonably small number for a single leg problem, and therefore, the proposed formulation can be very efficiently solved by a standard mixed integer programming solver such as CPLEX. We note that restricting the feasible region by introducing sufficiently large bounds is not really a concern in determining the optimal policy. Having $b_i = M_i$ at the optimal solution of the problem (13)-(17) would imply that, in practice, all of the booking requests for fare class i are accepted, since M_i is in general a large number compared to the number of arriving booking requests. However, forcing $b_i \leq M_i$ leads to an error in calculating the objective function value, since the function $\mathbb{E}(\mathbf{N}_i(\cdot)) : \mathbb{R}_+ \to \mathbb{R}$ is increasing, and so $\mathbb{E}(\mathbf{N}_i(M_i)) < \mathbb{E}(\mathbf{N}_i(\infty))$. To this end, we provide in Appendix D an analysis to determine the upper bound values in such a way that the derivation from the optimal objective function value of the problem (P_I^{UB}) is at most $m\epsilon$ for a specified error tolerance ϵ.

To compare the quality of the revenue obtained with the approximate optimization problem (P_I^{UB}) against that provided by the optimization problem (P_I), we next find a lower bound on the optimal objective function of the problem (P_I). To compute an upper bounding function on the expected total overbooking cost, let $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{Z}_+^m$ with $\sum_{i=1}^m y_i = C$ be a partitioned allocation of available capacity C to each fare class. By the subadditivity of the function $x \mapsto [x]^+$, we observe that

$$\left[\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right]^+ = \left[\sum_{i=1}^m (\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i)\right]^+ \leq \sum_{i=1}^m [\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i]^+ .$$

Thus, for any partitioned allocation $\mathbf{y}$ such that $\sum_{i=1}^m y_i = C$, $y_i \in \mathbb{Z}_+$, we have

$$\mathbb{E}\left(\left[\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right]^+\right) \leq \sum_{i=1}^m \mathbb{E}\left([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i]^+\right),$$

and we obtain by relation (11) that

$$\phi(\mathbf{b}) \geq \sum_{i=1}^m \tau_i \mathbb{E}(\mathbf{N}_i(b_i)) - \theta \sum_{i=1}^m \mathbb{E}\left([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i]^+\right) := \phi_L(\mathbf{b}, \mathbf{y}). \tag{19}$$

Hence, a lower bound on the optimal objective value of the problem (P_I) is found by solving

$$\max\{\phi_L(\mathbf{b}, \mathbf{y}) : \sum_{i=1}^m y_i = C,\ \mathbf{b} \in \mathbb{Z}_+^m,\ \mathbf{y} \in \mathbb{Z}_+^m\}. \tag{P_I^{LB}}$$

Since the optimization problem (P_I^{LB}) is separable, it can be solved by dynamic programming. We first observe that the problem (P_I^{LB}) is equivalent to the optimization problem

$$\max\{\rho_L(\mathbf{y}) : \sum_{i=1}^m y_i = C, \mathbf{y} \in \mathbb{Z}_+^m\}$$

with

$$\rho_L(\mathbf{y}) := \max\{\phi_L(\mathbf{b}, \mathbf{y}) : \mathbf{b} \in \mathbb{Z}_+^m\}.$$

By the additivity of the function $\mathbf{b} \to \phi_L(\mathbf{b}, \mathbf{y})$ given in (19) it follows that

$$\rho_L(\mathbf{y}) = \sum_{i=1}^m \rho_i(y_i)$$

with

$$\rho_i(y_i) = \max\{\tau_i \mathbb{E}(\mathbf{N}_i(b_i)) - \theta \mathbb{E}([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i]^+) : b_i \in \mathbb{Z}_+\}.$$

Since the random variable $\mathbf{B}(\beta_i^s, \mathbf{N}_i(b))$ is bounded above by b and the function $b \to \tau_i \mathbb{E}(\mathbf{N}_i(b))$ is increasing, we can restrict the feasible region $\{b_i \in \mathbb{Z}_+\}$ by adding the valid inequality $b_i \geq y_i$ and obtain

$$\rho_i(y_i) = \max\{\tau_i \mathbb{E}(\mathbf{N}_i(b_i)) - \theta \mathbb{E}([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - y_i]^+) : b_i \geq y_i, b_i \in \mathbb{Z}_+\}.$$

Observe that the above problem is in the form of the problem (P_T) presented in the previous section. Then, by using relation (9), the optimal solution of the above problem becomes

$$b_i^*(y_i) = \min\left\{b \geq y_i : \mathbb{P}(\mathbf{B}(\beta_i^s, b) \geq y_i) > \frac{\tau_i}{\theta \beta_i^s}\right\}.$$

This yields

$$\rho_i(y_i) = \tau_i \mathbb{E}(\mathbf{N}_i(b_i^*(y_i))) - \theta \mathbb{E}([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i^*(y_i))) - y_i]^+). \tag{20}$$

Therefore, the problem (P_I^{LB}) boils down to a simple allocation problem

$$\max\left\{\sum_{i=1}^m \rho_i(y_i) : \sum_{i=1}^m y_i = C, \mathbf{y} \in \mathbb{Z}_+^m\right\}$$

that can be solved by dynamic programming with a one-dimensional state space, where the stages correspond to the fare classes. The associated dynamic programming recursion can be formulated as follows: We consider for $j \in \{1, \ldots, m\}$ and $n \in \{0, 1, \ldots, C\}$, the parameterized optimization problems

$$R_j(n) = \max\left\{\sum_{i=j}^m \rho_i(y_i) : \sum_{i=j}^m y_i = n,\ y_i \in \mathbb{Z}_+,\ i = j, \ldots, m\right\}. \tag{21}$$

By relation (21), the boundary condition for $n \in \{0, 1, \ldots, C\}$ becomes

$$R_m(n) = \rho_m(n).$$

Then, by the dynamic programming optimality principle, the recursive relation for every $j \in \{1, \dots, m-1\}$ and $n \in \{0, 1, \dots, C\}$ is given by

$$R_j(n) = \max\{\rho_j(y_j) + R_{j+1}(n - y_j) : \ y_j \le n, \ y_j \in \mathbb{Z}_+\}.$$

Notice that this solution method requires evaluating the value of the function $\rho_i(y_i)$ given in (20) for all $i \in \{1, \dots, m\}$ and $y_i \in \{0, 1, \dots, C\}$. It is easy to find $b_i^*(y_i)$ using the recursive relation (10). Then, we need to efficiently calculate $\mathbb{E}([\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i^*(y_i)) - y_i]^+)$ for all $y_i \in \{0, 1, \dots, C\}$. To achieve this, we derive the distribution function of the bounded random variable $\mathbf{N}_i(b_i)$ and compute $P(\mathbf{B}(\beta_i^s, n) = k)$ for $n \in \{0, \dots, b_i\}$ and $k \in \{0, \dots, n\}$ using the following recursion:

$$P(\mathbf{B}(\beta_i^s, n) = k) = (1 - \beta_i^s)P(\mathbf{B}(\beta_i^s, n - 1) = k) + \beta_i^s P(\mathbf{B}(\beta_i^s, n - 1) = k - 1)$$

with the boundary condition $P(\mathbf{B}(\beta_i^s, 0) = 0) = 1$.

We remark that the lower bounding problem (P_I^{LB}) has a nice interpretation. The decision maker first determines the y_i, $i = 1, \dots, m$, values representing a partitioned allocation of the available capacity to each fare class. Then, the risk she takes is the possibility of observing that the total number of fare class i shows exceeds the preallocated capacity y_i, in which case she ends up paying a penalty cost. This means that a penalty is incurred even if a reservation occupies a preallocated seat belonging to a different fare class. With this interpretation, it is clear that by solving the problem (P_I^{LB}), we obtain a lower bound on the actual optimal expected total net revenue that would be secured by solving the actual problem (P_I).

An interesting question at this point is how to formally estimate the error committed by solving (P_I^{LB}) or (P_I^{UB}) instead of the originally proposed problem (P_I). We partly answer this question in the case of the upper bounding problem. The details of this analysis is given in Appendix C, where the concluding result is summarized in Lemma C.6. The main argument in this analysis is based on deriving a bound on the error introduced by Jensen's equality under the assumption that the random demands for different fare classes are independent. Lemma C.5 of Appendix C provides a key insight to explain what happens to the upper bound on the error committed by using Jensen's inequality: when the expectation of the total number of show-ups is close to C or its variance is high, the upper bound on the error turns out to be large.

As discussed in the beginning of this section, the practitioners prefer to use the partitioned booking limits in a nested way. Therefore, one can use the partitioned booking limits obtained by our lower and upper bounding models to calculate the nested booking limits, or equivalently, the nested protection levels that could be used in a dynamic setting. To be precise, the nested booking limit for fare class i is determined as $\sum_{j=1}^{i} b_j$, $i = 1, \dots, m$. In fact, this shall also be our approach in our computational study given in Section 5.

4. Dynamic Overbooking Model. We are next interested in solving the dynamic overbooking problem, where the seats need to be allocated to the fare classes from the start of the reservation horizon until the departure time. Since overbooking is allowed, the total number of reservations may exceed the actual capacity but the consequences, like denying boarding or departing with vacant seats, are faced at the time of departure. As time progresses during the reservation period the booking requests arrive randomly, and when a request arrives into the system we need to decide whether to accept or reject that request. The sequence of these accept or reject decisions leading to the highest net revenue is the optimal policy that we are after in this section.

4.1 Dynamics of The System. We introduce a discrete-time dynamic overbooking model, where time 0 represents the beginning of the reservation horizon and time T represents the departure time of the flight. The request arrivals only occur at discrete time points $t_k = kh$, $k = 1, \dots, K-1$, with h being chosen sufficiently small, $T = Kh$, $K \in \mathbb{N}$, and $t_0 = 0$. At most one booking request occurs at each time period $I_k = [t_{k-1}, t_k)$. A sample path of this discrete time arrival process is represented by a realization of a finite random vector $(\xi_1, \dots, \xi_{K-1})$, where $\xi_k = i$ designates that a request for fare class i arrives at time t_k, $i \in \{0, \dots, m\}$, $k = 1, \dots, K-1$. Note that a request for fare class 0 is also added to represent a no arrival at a given time point. The probability that a request for fare class i arrives at time t_k is $p_i(t_k) := \mathbb{P}(\xi_k = i)$, $i \in \{0, \ldots, m\}$, $k = 1, \ldots, K-1$. Clearly, $p_i(t_k) \geq 0$ and $\sum_{i=0}^{m} p_i(t_k) = 1$ for all time points $t_1, \ldots, t_{K-1}$.

To model the cancellation process, we assume that each reservation, independently of other reservations, cancels in period I_k with probability $c(I_k)$, $k = 2, \ldots, K$. Thus, the number of cancellations in period I_k, given that there are n accepted requests at time t_{k-1}, is a binomial distributed random variable $\mathbf{B}(c(I_k), n)$. Consequently, the number of accepted requests just before time t_k becomes $\mathbf{B}(1 - c(I_k), n)$. Observe that when

$$c(I_k) = 1 - \exp\left(-\lambda^c h\right),$$

the cancellation process is represented by a homogeneous Markovian death process with departure rate $\lambda^c > 0$, and hence, the cancellation probability does not depend on when the reservation was made. This property is coined as "forgetfulness property" and it is empirically confirmed to hold in practice (Rothstein, 1985).

As before r_i is the price of a fare class i ticket, $i = 1, \ldots, m$. We also introduce $r_0 = 0$ to represent the price for the no-arrival case. Without loss of generality, we take $r_0 < r_1 < \ldots < r_m$. We assume that each cancelled reservation receives a fixed refund of κ, and the airline incurs a fixed cost of θ for each denied boarding. At each time epoch t_k, we decide to accept or reject a possible request after the number of cancellations in the time interval I_k is realized. We might observe some no-shows just before the departure of the flight. It is assumed that the show-up probability of each reservation does not depend on its fare class, and it is denoted by β^s.

At this point we should note that some aspects of our model are covered by Subramanian et al. (1999) and Chatwin (1999). Subramanian et al. consider the arrival of a cancellation, the arrival of a booking request and no-arrival of any type as a combined stream. That is, they assume that only a booking request, a cancellation or a null event (no booking request, no cancellation) can be realized at each time epoch. This implies that the arrival and cancellation events are dependent and hence the probability measure of the arrival process of requests depends implicitly on the total number of reservations. However, their discretization approach allows for the independence of these two stochastic processes up to a $o(h)$ error in the associated probabilities, where h is the length of each time interval. In other words, in the discrete time setting of their model the independence between the arrival and cancellation processes holds as h goes to zero. On the other hand, our approach avoids this technical issue by modeling the arrival and cancellation processes as two different streams and allows naturally the independence between these two stochastic processes. Moreover, our alternative modeling approach yields a simpler mathematical proof of the discrete concavity of the expected optimal net revenue as a function of the total number of reservations. Chatwin (1999) avoids the discretization approach and assumes that the overall arrival process of the requests is a continuous time homogeneous Poisson process, and the probabilities to identify the class of a request are independent of time. Under this assumption, the arrival processes of requests for different fare classes are independent homogeneous Poisson processes. Also he models the cancellation process as a homogenous Markovian death process, and therefore, (although Chatwin applies the Bellman-Jacobi differential approach) it is possible to use a regenerative approach to analyze his model. However, for nonhomogeneous stochastic processes it is more difficult to apply the Bellman-Jacobi or regenerative approach (essentially we need to use a two dimensional state space in our optimal control problem) and since the corresponding continuous optimal value equation needs to be solved by discretization, it seems to be more natural to start at the beginning with a discrete time nonhomogenous arrival process.

4.2 Analysis of The Proposed Model. We now present the detailed mathematical description of the proposed dynamic model. Let us denote by t_k^+ the time epoch just after an accept or reject decision for a request that arrives at time t_k, $k = 1, \ldots, K-1$. Similarly, the time epoch just after the departure of the flight is denoted by t_K^+. Let $J_k(n)$, $k = 1, \ldots, K-1$, denote the expected optimal net revenue from t_k^+ up to t_K^+ given that the number of reservations at t_k^+ is n. To determine $J_k(n)$, $n \in \mathbb{Z}_+$, $k = 1, \ldots, K-1$, we first observe that after an accept or reject decision at t_k yielding a total of n reservations at time

t_k^+, the number of cancelled reservations in the interval I_{k+1} is a binomially distributed random variable $\mathbf{B}(c(I_{k+1}), n)$. Hence, the total number of reservations just before time t_{k+1} is $\mathbf{B}(1 - c(I_{k+1}), n)$. This implies that the total number of reservations just before the departure time is $\mathbf{B}(1-c(I_K), n)$ and the total number of shows is given by $\mathbf{B}(\beta^s(1 - c(I_K)), n)$. Then, by $\mathbb{E}(\mathbf{B}(c(I_k), n)) = nc(I_k)$, the independence of the arrival and cancellation processes and the dynamic programming optimality principle we obtain for every $k = 1, \ldots, K - 2$, and $n \in \mathbb{Z}_+$

$$\begin{aligned} J_k(n) &= -\kappa nc(I_{k+1}) + p_0(t_{k+1})\mathbb{E}\left(J_{k+1}(\mathbf{B}(1 - c(I_{k+1}), n))\right) \\ &+ \textstyle\sum_{i=1}^m p_i(t_{k+1})\mathbb{E}\left(\max\left\{r_i + J_{k+1}(\mathbf{B}(1 - c(I_{k+1}), n) + 1), J_{k+1}(\mathbf{B}(1 - c(I_{k+1}), n))\right\}\right) \end{aligned} \qquad (P_{DM})$$

and the boundary condition

$$J_{K-1}(n) = -\kappa nc(I_K) - \theta\mathbb{E}\left([\mathbf{B}(\beta^s(1 - c(I_K)), n) - C]^+\right). \qquad (22)$$

Clearly, for $n = 0$ we obtain $P(\mathbf{B}(1 - c(I_{k+1}), 0) = 0) = 1$, and the above recursion reduces to

$$J_k(0) = p_0(t_{k+1})J_{k+1}(0) + \sum_{i=1}^m p_i(t_{k+1}) \max\left\{r_i + J_{k+1}(1), J_{k+1}(0)\right\}.$$

We next obtain the optimal policy of the above dynamic programming model by showing that the function $n \mapsto J_k(n)$ is a discrete concave function on $\mathbb{Z}_+$ for every $k = 1, \ldots, K - 1$.

Lemma 4.1 *The function $n \mapsto J_k(n)$ is discrete concave on $\mathbb{Z}_+$ for every $k = 1, \ldots, K - 1$.*

Proof. For ease of exposition we introduce the function $n \mapsto \Gamma_{k+1}(i, n)$ given by

$$\Gamma_{k+1}(i, n) := \begin{cases} \max\left\{r_i + J_{k+1}(n + 1), J_{k+1}(n)\right\}, & \text{for } i \in \{1, \ldots, m\}; \\ J_{k+1}(n), & \text{for } i = 0, \end{cases} \qquad (23)$$

Then, the recursion of the dynamic model (P_{DM}) for every $k = 1, \ldots, K - 2$, becomes

$$J_k(n) = -\kappa nc(I_{k+1}) + \textstyle\sum_{i=0}^m p_i(t_{k+1})\mathbb{E}\left(\Gamma_{k+1}\left(i, \mathbf{B}(1 - c(I_{k+1}), n)\right)\right). \qquad (24)$$

Using Lemma B.2, it follows that the function $n \mapsto J_{K-1}(n)$ listed in relation (22) is discrete concave on $\mathbb{Z}_+$. Suppose now for a given $k+1 < K$ that the function $n \mapsto J_{k+1}(n)$ is discrete concave on $\mathbb{Z}_+$. Our proof is then completed once we show that the function $n \mapsto J_k(n)$ is discrete concave on $\mathbb{Z}_+$. Applying our induction hypothesis and Lemma B.1, we first obtain that the function $n \mapsto \Gamma_{k+1}(i, n)$ given in (23) is discrete concave for any $i \in \{0, 1 \ldots, m\}$. This implies using Lemma B.2 that the function

$$n \mapsto \mathbb{E}\left(\Gamma_{k+1}\left(i, \mathbf{B}(1 - c(I_{k+1}), n)\right)\right)$$

is discrete concave on $\mathbb{Z}_+$ and by relation (24) the result follows. □

Let us now introduce

$$b_{ki} := \max\left\{n \in \mathbb{Z}_+ : r_i \geq J_{k+1}(n) - J_{k+1}(n + 1)\right\}.$$

Since a discrete concave function has decreasing differences by definition, it follows by Lemma 4.1 that the following dynamic booking limit policy is optimal:

"accept the request for fare class i at t_k $\Leftrightarrow$ total number of reservations $\leq b_{ki}$"

As the fares are assumed to be ordered, we then obtain the following nested structure:

$$b_{k1} \leq b_{k2} \leq \cdots \leq b_{km}.$$

5. Computational Experiments. We devote this section to a computational study for discussing different aspects of the models proposed in the previous sections. In particular, we conduct simulation experiments to benchmark the policies obtained with our lower bounding model (P_I^{LB}), upper bounding model (P_I^{UB}) and the dynamic model (P_{DM}) against some well-known approaches used in the literature (Lan et al., 2008, 2011). We next explain our simulation setup in detail and then present our numerical results.

5.1 Simulation Setup. We simulate the arrival of requests and cancellations over the discrete time points $t_k,\ k = 1, \ldots, K-1$. The probability that there is a request for fare class i at time point t_k is $p_i(t_k)$. If we accept a request for fare class i, then we generate a revenue of r_i. Without loss of generality, we take $r_0 < r_1 < \ldots < r_m$. Each accepted fare class i request cancels with probability $c_i(I_k)$ in period $I_k = [t_{k-1}, t_k),\ k = 2, \ldots, K$. Hence, the number of fare class i cancellations at time point t_k is binomially distributed with a success probability $c_i(I_{k+1})$. Each cancellation is refunded with an amount of $r_i\alpha_i,\ i = 1, \ldots, m$. At the end of the reservation period, each reservation shows up with probability β_i^s and the penalty cost of denying boarding to a reservation for fare class i is νr_i.

To generate these arrival and cancellation probabilities we shall mimic the actual stochastic processes. We assume that the booking requests arrive according to a homogeneous Poisson process with rate λ^a, and the cancellations for fare classes $i = 1, \ldots, m$, are modeled by a Markovian death process with departure rates λ_i^c. Then, we have for $k = 1, \ldots, K-1$

$$p_0(t_k) = \exp(-\lambda^a h)$$

and

$$c_i(I_k) = 1 - \exp\left(-\lambda_i^c h\right).$$

Given a request arrives at time t_k, this request is for fare class i with probability $f_i(t_k)$ satisfying, $f_i(t_k) \geq 0$ and $\sum_{i=1}^m f_i(t_k) = 1$. In other words, upon an arrival at time t_k, the different fare class requests are generated according to a multinomial selection scheme with probabilities $f_i(t_k)$, $i = 1, \ldots, m$, $1 \leq k \leq K-1$. Assuming that in reality the lower fare class requests arrive more frequently in the early periods than the higher fare classes, we set the multinomial probabilities as

$$f_i(t_k) = \frac{\pi_i(t_k)}{\sum_{i=1}^m \pi_i(t_k)},\ \ i = 1, \ldots, m,$$

where $\pi_i(t_k)$ are simple linear functions. This way of setting the multinomial probabilities complies with the desired demand pattern. As illustrated in Figure 1, we set

$$p_i(t_k) = f_i(t_k)(1 - p_0(t_k)), \quad i = 1, \ldots, m,\ k = 1, \ldots, K-1.$$

In our simulation setup, the following class-dependent parameters are given: fares (r_i), refund percentages

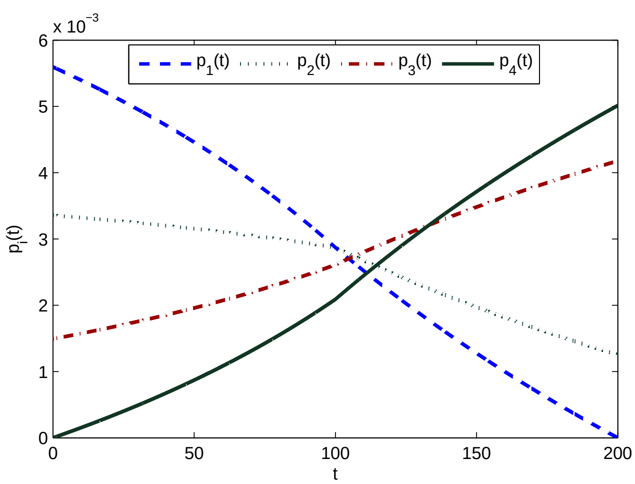

Figure 1: An example of the changes in multinomial probabilities over time

(α_i), cancellation probabilities (β_i^c), and show-up probabilities (β_i^s). In order to test the performances of the booking policies against varying arrival intensities, we use the load factor parameter ρ, which is given by

$$\rho = \frac{(K-1)(1 - \exp(-\lambda^a h))}{C}. \tag{25}$$

Observe that the denominator is the expected number of booking requests. To conform with our simulation setup, we tie the arrival rate to a given load factor and obtain λ^a by solving (25) for a specified value of ρ. When it comes to the cancellation rates, we assume that the behaviour of the customers towards cancellation is independent of whether they have reserved a ticket or not. Using this assumption and simple conditioning, we can relate the cancellation probabilities to the cancellation rates and acquire λ_i^c, $i = 1, \ldots, m$, from

$$\frac{\sum_{k=1}^{K-1}(1 - \exp(-\lambda_i^c(T - t_k)))f_i(t_k)}{\sum_{k=1}^{K-1} f_i(t_k)} = \beta_i^c, \; i = 1, \ldots, m.$$

Letting $\mathbf{D}_i$ be the aggregated fare class i demand, we obtain the probabilities $p_i = \frac{\mathbb{E}(\mathbf{D}_i)}{\mathbb{E}(\mathbf{D})}$, $i = 1, \ldots, m$, denoting the fractions of the aggregated demand allocated to different fare classes.

Recall that in our dynamic model the cancellation and show-up probabilities do not depend on the fare classes. By applying a simple conditioning, we estimate the class-independent show-up and cancellation probabilities as

$$\beta^s = \sum_{i=1}^{m} \beta_i^s p_i \text{ and } \beta^c := \sum_{i=1}^{m} \beta_i^c p_i, \tag{26}$$

respectively. Using now the class independent cancellation probability, we obtain the cancellation rate, λ^c by solving

$$\frac{\sum_{k=1}^{K-1}(1 - \exp(-\lambda^c(T - t_k)))}{K - 1} = \beta^c.$$

5.2 Numerical Results. In this section, we apply a benchmarking study including several approaches from the literature as well as our static and dynamic models. We also provide an experimental design, similar to the one in (Topaloglu et al., 2011), for different parameters used in our simulation. All the contender methods that we use for benchmarking apply the EMSR-b heuristic but they mainly differ in terms of the way the virtual capacity is obtained:

- EMSR/Risk: Our total booking limit given by relation (9) is used as the virtual capacity.
- EMSR/MP: The virtual capacity is set according to the deterministic rule described by Belobaba (2006). However, this rule requires a class independent show up rate. Therefore, we use β^s as described at the end of the previous section and the virtual capacity is equal to C/β^s.
- EMSR/SL: The virtual capacity is based on a type-I service level constraint using the actual capacity. This constraint imposes that probability of overbooking is less than or equal to $1.0e-3$ (Phillips, 2005, Section 9.3).
- EMSR/NO: Overbooking is not allowed. Therefore, EMSR-b heuristic is applied with the actual capacity.

In the sequel, we simulate the arrival process for many replications and refer to the average revenues obtained by the optimal policies of our static models (P_I^{UB}) and (P_I^{LB}) as UB and LB, respectively. Likewise, we denote the average revenue of the dynamic policy obtained with our model (P_{DM}) by DM. We note once again that both of the static models provide partitioned booking limits but we use these limits in a nested way in all our simulations.

In all our numerical experiments, we set the capacity of the plane, the planning horizon, the discretization mesh lengths and the number of discrete time points to $C = 150$, $T = 200$, $h = 1.0e-2$, $K = 20,000$, respectively. The refund percentages $(\alpha_1, \ldots, \alpha_m)$ and the cancellation probabilities $(\beta_1^c, \ldots, \beta_m^c)$ are evenly distributed in the intervals $[0.00, 0.30]$ and $[0.05, 0.17]$. For our dynamic programming implementation to solve the DP model, an upper bound sufficiently larger than C was imposed on the total number of reservations. This allows us to restrict the state space for computational purposes. In the implementation for solving the DP model, setting such an upper bound means that a booking request would be rejected if the total number of reservations reaches this upper bound. As required by formulation (13)-(18), we also need to impose an upper bound M_i on the booking limit b_i for each $i = 1, \ldots, m$. To serve this purpose, we choose sufficiently large M_i values by setting $\epsilon = 1.0e-7$ in relation (63).

Our experimental design is based on various factors of the fares (r_i), the overbooking cost θ, the load factor ρ, the number of fare classes m, and the show-up probabilities (β_i^s). The lowest price is fixed to 50 and the prices of the other fare classes are evenly distributed in the interval $[50, \eta 50]$, where $\eta \in \{4, 7\}$ gives two sets of fares. The overbooking cost is determined by

$$\theta = \nu \sum_{i=1}^{m} r_i p_i,$$

where $\nu \in \{3, 5\}$ is used for creating two factors indicating low and high overbooking costs. We use load factor values $\rho \in \{1.4, 1.8\}$ corresponding to medium and high loads. We also apply sensitivity analysis with respect to the number of fare classes selected as $m \in \{4, 8\}$. The last parameter set comes from the show-up probabilities $\boldsymbol{\beta}_\bullet^s := (\beta_1^s, \ldots, \beta_m^s)$. We give two sets of show-up probabilities to represent possibly low and high show-up rates. These are $\boldsymbol{\beta}_L^s := (0.95, 0.92, 0.80, 0.77)$ and $\boldsymbol{\beta}_H^s := (0.98, 0.95, 0.83, 0.80)$ for $m = 4$; $\boldsymbol{\beta}_L^s := (0.95, 0.93, 0.91, 0.89, 0.83, 0.81, 0.79, 0.77)$ and $\boldsymbol{\beta}_H^s := (0.98, 0.96, 0.94, 0.92, 0.86, 0.84, 0.82, 0.80)$ for $m = 8$. Under this setup, we evaluate the solutions of all the approaches under consideration for all 32 test problem instances. Then, the policies obtained by these solutions are compared for each instance by taking 50 simulation runs.

Table 1 presents the optimal objective function values of (P_I^{UB}) and (P_I^{LB}) and the gap between them for all test instances. This gap is defined as the relative difference with respect to the optimal objective function value of (P_I^{LB}). As seen from this table, the relative differences are mostly affected by the number of fare classes. Recall that (P_I^{LB}) partitions the actual capacity to each fare class and incurs a penalty even if a reservation occupies a preallocated seat belonging to a different fare class. This treatment of the capacity does not allow sharing the seats among the fare classes efficiently. Consequently, the performance of (P_I^{LB}) deteriorates more than that of (P_I^{UB}) and the percentage gap increases with a higher number of fare classes. We also observe that the overbooking cost coefficient ν slightly affects the percentage gap. The results indicate that the optimal objective function value of (P_I^{LB}) tends to decrease as ν gets higher. On the other hand, the changes in the optimal objective function values of (P_I^{UB}) are insignificant when the overbooking cost becomes higher. Consequently, the percentage gap tends to increase with ν. Regarding the effect of the parameter η, we observe that the optimal objective function values of both models increase with η. However, the increase in the optimal objective function value is larger for (P_I^{LB}) compared to (P_I^{UB}). Therefore, the percentage gap tends to decrease as η gets higher.

Figures 2 to 5 present average net revenues over all simulation runs for the booking policies obtained by different methods with varying factors. In these figures, we compare the performances of the booking policies obtained by our proposed models to those of the benchmarking methods with respect to the high and low show-up probabilities, denoted by H and L, and the overbooking penalty factor. The detailed results related to these numerical experiments are given in Table 2, where the dynamic model is used as a base approach to report the relative gap of the remaining approaches with respect to the revenue obtained by the dynamic model.

The first observation we have is that the proposed upper bounding model (P_I^{UB}) performs better than all the EMSR-based heuristics for any combination of the parameters. There are even cases when the average revenues of the booking policies obtained by (P_I^{UB}) and (P_{DM}) are relatively close (see Figure 2). We caution the reader that these relatively small gaps between DM and UB implicitly demonstrates the importance of using class dependent show-up and cancellation probabilities. Lacking this attribute, the dynamic model treats all cancellations and no-shows the same way, and consequently, may fail to capture the actual dynamics of the system. As Figures 3-5 illustrate, the lower bounding problem (P_I^{LB}) performs slightly better when the load factor is high. As we mentioned before, (P_I^{LB}) is more conservative than the upper bounding problem and its overbooking policy is based on reserving more seats only for the expensive fare classes. Therefore, when the load-factor is high, it benefits from the increase in the number of booking requests for the expensive fare classes. Comparing the performance of (P_I^{LB}) in Figures 2 and 4 with those in Figures 3 and 5, we note that the average revenue associated with the policy obtained by (P_I^{LB}) is closer to the revenue obtained by EMSR/SL for the lower load-factor value.

Table 1: The optimal objective function values of P_I^{UB} and P_I^{LB}

Instances							
m	ρ	$\boldsymbol{\beta}_{\bullet}^s$	η	ν	P_I^{LB}	P_I^{UB}	$(P_I^{UB}$-$P_I^{LB})/P_I^{LB}$
4	1.4	$\boldsymbol{\beta}_H^s$	4	3	21,444.88	22,815.67	6.39%
			4	5	21,337.41	22,815.67	6.93%
			7	3	35,601.98	37,265.54	4.67%
			7	5	35,464.30	37,268.98	5.09%
		$\boldsymbol{\beta}_L^s$	4	3	21,654.65	23,071.37	6.54%
			4	5	21,528.50	23,071.38	7.17%
			7	3	35,834.72	37,527.54	4.72%
			7	5	35,702.62	37,527.54	5.11%
	1.8	$\boldsymbol{\beta}_H^s$	4	3	24,434.11	26,106.48	6.84%
			4	5	24,186.78	26,106.48	7.94%
			7	3	41,014.06	43,672.63	6.48%
			7	5	40,618.44	43,672.63	7.52%
		$\boldsymbol{\beta}_L^s$	4	3	24,904.36	26,674.35	7.11%
			4	5	24,622.59	26,674.35	8.33%
			7	3	41,714.30	44,537.86	6.77%
			7	5	41,277.23	44,537.86	7.90%
8	1.4	$\boldsymbol{\beta}_H^s$	4	3	20,403.45	22,657.20	11.05%
			4	5	20,215.49	22,657.20	12.08%
			7	3	33,653.33	36,990.55	9.92%
			7	5	33,396.48	36,990.55	10.76%
		$\boldsymbol{\beta}_L^s$	4	3	20,670.32	23,053.38	11.53%
			4	5	20,436.40	23,053.38	12.81%
			7	3	34,022.66	37,502.04	10.23%
			7	5	33,702.95	37,502.04	11.27%
	1.8	$\boldsymbol{\beta}_H^s$	4	3	23,141.81	25,606.24	10.65%
			4	5	22,873.36	25,606.24	11.95%
			7	3	38,817.54	42,726.61	10.07%
			7	5	38,399.12	42,726.61	11.27%
		$\boldsymbol{\beta}_L^s$	4	3	23,542.35	26,135.89	11.02%
			4	5	23,209.54	26,137.07	12.61%
			7	3	39,384.75	43,504.99	10.46%
			7	5	38,917.66	43,502.70	11.78%

However, it performs better and the average revenues stay close to the revenues provided by EMSR/Risk and EMSR/MP, when the load factor is high. There are even instances when (P_I^{LB}) outperforms both EMSR/MP and EMSR/Risk. However, when the number of fare classes increases, its performance quickly deteriorates (see Figures 4 and 5).

When we look into the performances of the EMSR-based heuristics, we observe that EMSR/Risk and EMSR/MP are better than the remaining two heuristics, EMSR/NO and EMSR/SL. This difference is more striking when the load factor is high and the show-up probabilities are low as designated by Figures 3 and 5 (see also the rows corresponding to $\boldsymbol{\beta}_L^s$ in Table 2). The average revenue obtained by EMSR/MP is slightly higher than that of EMSR/Risk. Unlike EMSR/Risk, EMSR/MP does not consider the overbooking penalty when determining the virtual capacity. Therefore, the difference between the average revenues of the policies obtained by these models increases with the overbooking cost factor. It turns out that the proposed weighted average of the class-dependent show-up rates given in relation (26) captures the nature of the show-up behavior accurately. We observe in our numerical study that

EMSR/MP reserves slightly more seats than EMSR/RISK (at most 3 seats over all instances), and these additional seats are effective for collecting extra revenues. This success of EMSR/MP is also in accordance with the observation made by (Phillips, 2005, Section 9.3). Table 2 and Figures 2 to 5 illustrate that, like our bounding models, the performances of the EMSR-based heuristics deteriorate with respect to the dynamic model with a higher number of fare classes. The deterioration in the performances of the EMSR-based heuristics can be attributed to the fact that they are mainly based on comparing two fare classes. To obtain such a structure, each fare-class is compared against the aggregation of the classes with lower fares. As the number of fare-classes increases, the aggregation does not capture the stochastic nature of the problem well. It is also important to note that the percentage gaps between DM and the revenues of the remaining strategies are more striking when the load factor is high. This intensity can be attributed to the reactions of the models to the low fare class requests, especially, in the early periods. As the load factor becomes higher, we observe many requests throughout the planning horizon. The dynamic policy then reacts in a more conservative way and rejects the early low fare requests. Such behaviour allows reserving seats for more expensive fare classes arriving in later periods, and hence, results with an increase in the total revenue. However, working with aggregated demands, the static models cannot react to the changes within different time intervals.

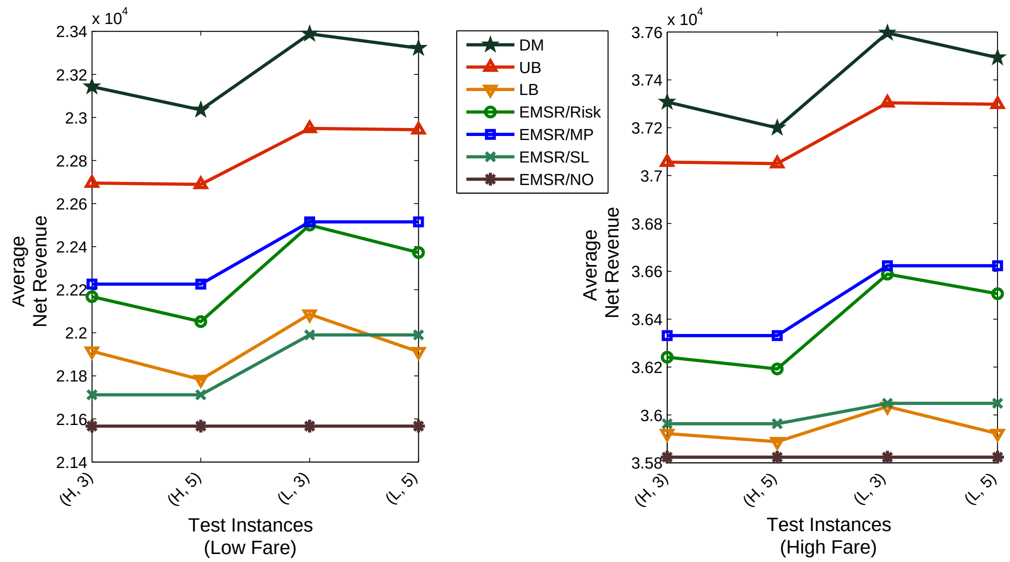

Figure 2: Average net revenues ($\rho = 1.4$, $m = 4$)

We next report an encouraging result about the error we introduce by solving the upper bounding problem. As in Lemma C.6, we denote the optimal solutions of the original static problem (P_I), the upper bounding problem (P_I^{UB}) and the lower bounding problem (P_I^{LB}) by $\mathbf{b}^*$, $\mathbf{b}_U^*$ and $\mathbf{b}_L^*$, respectively. We evaluate the theoretical upper bounds on the ratio $\phi_U(\mathbf{b}_\mathbf{U}^*)/\phi(\mathbf{b}^*)$ by using Lemma C.6 and the relations in (61). We also calculate the actual error bounds as discussed in Appendix C. Table 3 presents these theoretical and actual error bounds given by (59). Let $\mathbf{Z}(\mathbf{b}_U^*)$ denote the random total number of show-ups associated with the optimal solution $\mathbf{b}_U^*$ (see relation (35)). As seen in Lemma (C.5) the quality of the theoretical error bound depends on the volatility of the random variable $\mathbf{Z}(\mathbf{b}_U^*)$ and how close its expectation is to the capacity. The figures in the table confirm our analysis as the calculated theoretical error bound increases with the variance. We also observe a similar behaviour for the calculated actual error bound. However, it is important to note that although the theoretical upper bound overestimates the actual difference between (P_I) and (P_I^{UB}), this overestimation improves as the load-factor increases. This also signals that the optimal objective function value of the upper bounding problem could be close to the original problem when the load-factor is high.

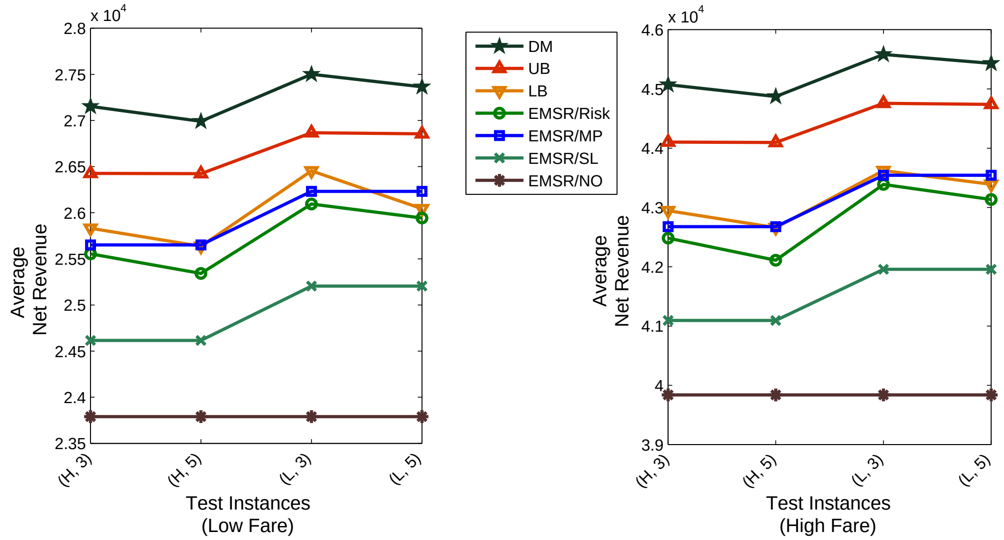

Figure 3: Average net revenues ($\rho = 1.8$, $m = 4$)

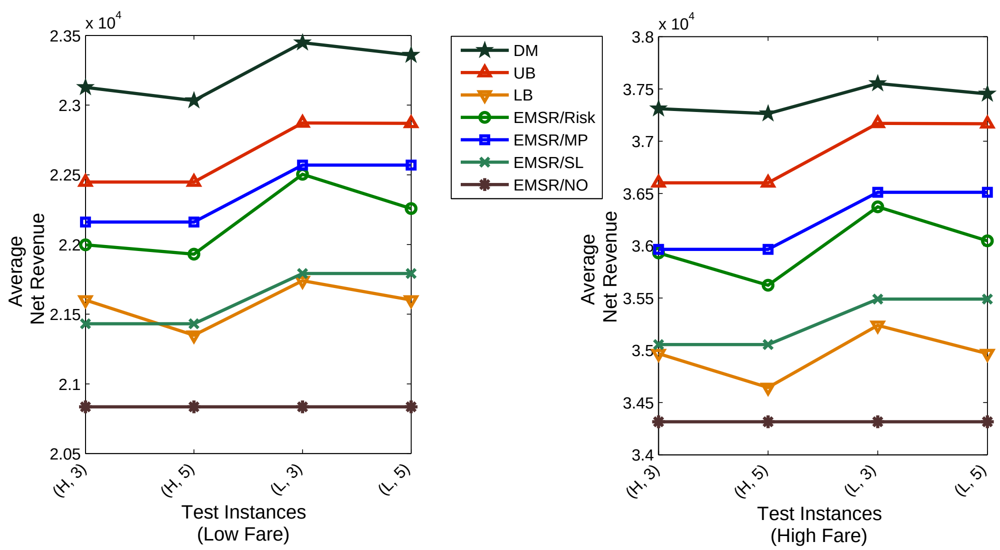

Figure 4: Average net revenues ($\rho = 1.4$, $m = 8$)

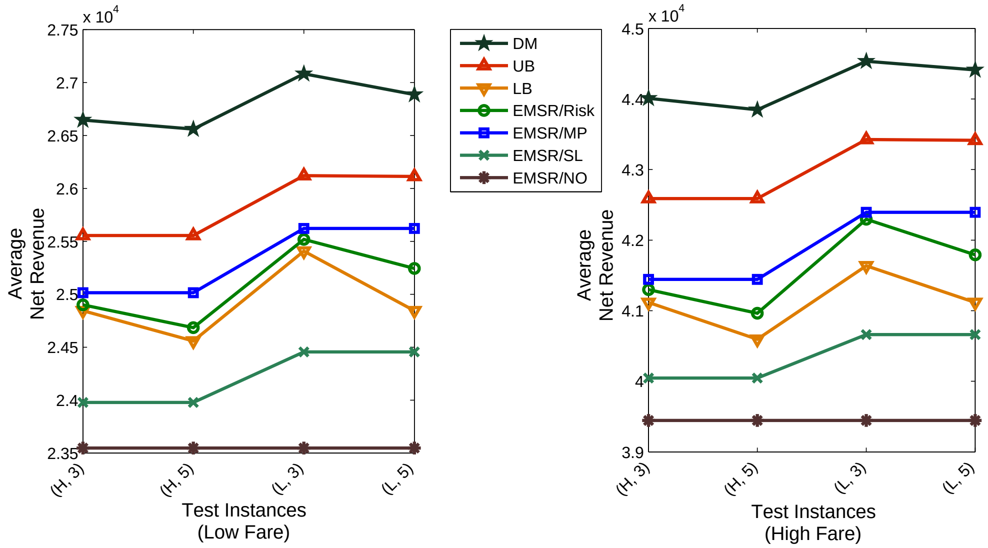

Figure 5: Average net revenues ($\rho = 1.8$, $m = 8$)

We conclude the presentation of our numerical results by reporting the wall-clock times of the proposed solution methods. We used a computer with 2.4 GHz Intel Core 2 Quad processor and 3024 MB of RAM. The codes are written in MATLAB 7.6.0 running under Windows XP operating system. EMSR/NO, EMSR/SL, EMSR/MP, and EMSR/Risk heuristics require on average less than 0.1 seconds. It takes on average 1.10 and 0.40 seconds to solve the lower and the upper bounding problems, respectively. Thus, our heuristics are comparable to the widely-applied EMSR-based heuristics in terms of computational efficiency. The most computational effort is invested in finding the optimal policy of dynamic model, which takes on average 2260 seconds. Clearly, this time depends on the mesh-size parameter h and the length of the planning horizon T.

6. Conclusion In this study, we develop new optimization models for static and dynamic single-leg revenue management problems that involve no-shows, cancellations, and hence, overbooking. In the static case we discuss two risk-based models both of which allow class dependent cancellations and no-shows. Our first static model determines the optimal total booking limit under the greedy policy. Finding the optimal total booking limit under such a general setting is useful in practice, since the overbooking limit can be used as an input to some well-known capacity allocation methods like the EMSR heuristics. In the second static model, we determine both the total booking limit and the partitioned booking limits. Arriving at a computationally difficult model, we propose upper and lower bounding problems to obtain approximate solutions. As preferred in practice, we propose to use the partitioned booking limits obtained by our upper and lower bounding models in a nested way. Thus, the resulting method becomes a heuristic to obtain nested booking limits but it does not require a predefined overbooking limit like the EMSR heuristics. In the dynamic case we propose a model based on two independent streams of events; arrivals of booking requests and cancellations. Our modeling approach allows the arrival process of the booking requests to be independent of the number of reservations. Moreover, the number of cancellations in any time period, given the number of accepted requests at the beginning of that time period, is a binomially distributed random variable. We show that it is easy to solve the resulting problem with dynamic programming. After characterizing the optimal policy, we also present the nested structure of the optimal allocations.

We conduct a computational study to compare the performances of the booking policies obtained by our proposed models to those of some well-known EMSR-based approaches used in the literature. The numerical results demonstrate that the proposed upper bounding model outperforms the EMSR-based heuristics for the generated test problem instances and perform reasonably well with respect to the DP

Table 2: Percentage differences relative to the expected net revenue of (P_{DM}) $(C = 150)$

Instances					DM *versus*					
m	ρ	$\boldsymbol{\beta}^s_\bullet$	η	ν	EMSR/NO	EMSR/SL	EMSR/MP	EMSR/Risk	LB	UB
4	1.4	$\boldsymbol{\beta}^s_H$	4	3	6.91%	6.29%	4.06%	4.31%	5.39%	2.02%
			4	5	6.46%	5.84%	3.59%	4.35%	5.52%	1.58%
			7	3	4.04%	3.67%	2.67%	2.93%	3.78%	0.73%
			7	5	3.76%	3.39%	2.39%	2.75%	3.59%	0.45%
		$\boldsymbol{\beta}^s_L$	4	3	7.88%	6.08%	3.82%	3.88%	5.66%	1.95%
			4	5	7.61%	5.80%	3.54%	4.14%	6.13%	1.69%
			7	3	4.79%	4.20%	2.65%	2.75%	4.23%	0.83%
			7	5	4.53%	3.93%	2.38%	2.71%	4.26%	0.57%
	1.8	$\boldsymbol{\beta}^s_H$	4	3	12.48%	9.43%	5.61%	5.96%	4.94%	2.75%
			4	5	11.96%	8.89%	5.04%	6.19%	5.10%	2.19%
			7	3	11.71%	8.92%	5.39%	5.83%	4.80%	2.23%
			7	5	11.32%	8.52%	4.97%	6.26%	5.01%	1.81%
		$\boldsymbol{\beta}^s_L$	4	3	13.58%	8.43%	4.67%	5.08%	3.87%	2.37%
			4	5	13.16%	7.98%	4.21%	5.28%	4.91%	1.94%
			7	3	12.69%	8.01%	4.54%	4.88%	4.37%	1.87%
			7	5	12.40%	7.71%	4.23%	5.12%	4.56%	1.59%
8	1.4	$\boldsymbol{\beta}^s_H$	4	3	10.01%	7.42%	4.25%	4.95%	6.67%	3.02%
			4	5	9.64%	7.04%	3.85%	4.86%	7.38%	2.61%
			7	3	8.09%	6.11%	3.66%	3.75%	6.32%	1.94%
			7	5	7.97%	5.98%	3.54%	4.45%	7.07%	1.82%
		$\boldsymbol{\beta}^s_L$	4	3	11.23%	7.14%	3.81%	4.09%	7.35%	2.54%
			4	5	10.90%	6.79%	3.45%	4.78%	7.59%	2.18%
			7	3	8.68%	5.54%	2.82%	3.18%	6.20%	1.07%
			7	5	8.43%	5.29%	2.55%	3.79%	6.67%	0.81%
	1.8	$\boldsymbol{\beta}^s_H$	4	3	11.72%	10.12%	6.22%	6.65%	6.84%	4.18%
			4	5	11.45%	9.83%	5.92%	7.16%	7.63%	3.88%
			7	3	10.46%	9.12%	5.93%	6.25%	6.66%	3.31%
			7	5	10.13%	8.78%	5.58%	6.68%	7.51%	2.95%
		$\boldsymbol{\beta}^s_L$	4	3	13.15%	9.80%	5.48%	5.85%	6.26%	3.63%
			4	5	12.51%	9.15%	4.79%	6.19%	7.67%	2.95%
			7	3	11.50%	8.78%	4.88%	5.09%	6.57%	2.55%
			7	5	11.27%	8.54%	4.63%	5.99%	7.50%	2.31%

model. We also observe that the policies proposed by our upper bounding model are robust, even if we switch from low to high show-up probabilities or increase the overbooking cost. On the other hand, the performance of proposed lower bounding model deviates depending on the number of fare classes and the load factor. We also derive theoretical and actual bounds on the error introduced by solving the upper bounding problem instead of the corresponding original static model. Computational experiments demonstrate that the error bounds are tighter when the load-factor is higher. As a future work we are planning to study the extensions of our proposed models in the network environment.

Acknowledgments. We would like to thank the anonymous referees and the associate editor for their helpful suggestions and comments. The second author also acknowledges the faculty and staff at the Department of Computer Engineering, Boğaziçi University for allowing him to use their facilities.

Table 3: Error Bound in Jensen's inequality

Instances					$\mathbb{E}(\mathbf{Z}(\mathbf{b}_U^*))$	$\sigma^2(\mathbf{Z}(\mathbf{b}_U^*))$	Theoretical Error Bound	Actual Error Bound
m	ρ	$\boldsymbol{\beta}_\bullet^s$	η	ν				
4	1.4	$\boldsymbol{\beta}_H^s$	4	3	150.00	111.66	1.28	1.08
			4	5	150.00	111.66	1.47	1.13
			7	3	149.98	111.05	1.27	1.07
			7	5	150.00	111.66	1.45	1.12
		$\boldsymbol{\beta}_L^s$	4	3	149.66	105.81	1.27	1.07
			4	5	149.66	105.81	1.45	1.12
			7	3	149.66	105.81	1.26	1.07
			7	5	149.66	105.81	1.43	1.12
	1.8	$\boldsymbol{\beta}_H^s$	4	3	150.00	75.59	1.24	1.06
			4	5	150.00	75.59	1.40	1.10
			7	3	150.00	75.59	1.22	1.05
			7	5	150.00	75.59	1.38	1.09
		$\boldsymbol{\beta}_L^s$	4	3	149.68	80.27	1.23	1.06
			4	5	149.68	80.27	1.39	1.10
			7	3	149.68	80.27	1.22	1.05
			7	5	149.68	80.27	1.36	1.09
8	1.4	$\boldsymbol{\beta}_H^s$	4	3	149.88	111.56	1.29	1.08
			4	5	149.88	111.56	1.49	1.13
			7	3	149.88	111.56	1.28	1.08
			7	5	149.88	111.56	1.47	1.13
		$\boldsymbol{\beta}_L^s$	7	3	149.98	112.02	1.29	1.08
			7	5	149.98	112.02	1.49	1.14
			7	3	149.98	112.02	1.28	1.08
			7	5	149.98	112.02	1.47	1.13
	1.8	$\boldsymbol{\beta}_H^s$	7	3	149.98	108.86	1.26	1.07
			7	5	149.98	108.86	1.43	1.12
			7	3	149.98	108.86	1.24	1.07
			7	5	149.98	108.86	1.41	1.11
		$\boldsymbol{\beta}_L^s$	4	3	150.01	100.23	1.26	1.07
			4	5	150.00	100.05	1.43	1.12
			7	3	150.01	100.23	1.24	1.06
			7	5	150.01	100.23	1.41	1.11

Appendix A. Review on Bernoulli Selection Scheme. In this appendix, we first define a Bernoulli selection type random variable. If $\mathbf{X}$ denotes the non-negative integer random size of a population, then the random variable $\mathbf{B}(p, \mathbf{X})$ denotes the total number within the population of size $\mathbf{X}$ having a certain property under the condition that each member in the population has this property with probability p independent of each other. Hence, the random variable $\mathbf{B}(p, \mathbf{X})$ is given by

$$\mathbf{B}(p, \mathbf{X}) := \begin{cases} \sum_{k=1}^{\mathbf{X}} \mathbf{1}_{\{\mathbf{U}_k \leq p\}}, & \text{if } \mathbf{X} \geq 1; \\ 0, & \text{if } \mathbf{X} = 0, \end{cases} \tag{27}$$

where $\mathbf{U}_k, k \in \mathbb{N}$, is a sequence of independent standard uniformly distributed random variables, and the random variable $\mathbf{X}$ is independent of the sequence $\mathbf{U}_k,\ k \in \mathbb{N}$. By relation (27), we obtain

$$\mathbb{E}\left(\mathbf{B}(p, \mathbf{X})\right) = p\mathbb{E}\left(\mathbf{X}\right).$$

Furthermore, it is well-known that the generating function of the random variable $\mathbf{B}(p, \mathbf{X})$ is given by

$$\mathbb{E}\left(z^{\mathbf{B}(p,\mathbf{X})}\right) = \mathbb{E}\left((1 - p + pz)^{\mathbf{X}}\right) \tag{28}$$

and

$$\mathbf{B}\left(q, \mathbf{B}(p, \mathbf{X})\right) =^d \mathbf{B}(pq, \mathbf{X})$$

for any $0 \leq p, q \leq 1$ (Feller, 1968).

Appendix B. Results on Discrete Concave Functions. In this appendix, we shall mention some results related to the discrete concavity (convexity) that are used in our analysis of the proposed models. We start with a definition.

Definition B.1 *A function $f : \mathbb{Z}_+ \mapsto \mathbb{Z}$ is discrete concave if and only if the differences $n \mapsto f(n+1) - f(n)$ are decreasing. A function f is discrete convex if and only if -f is discrete concave.*

The proof of the following lemma is given by Lippman and Stidham (1977).

Lemma B.1 *Let $r \geq 0$ and $f : \mathbb{Z}_+ \mapsto \mathbb{R}$ be a discrete concave function. Then the function $h : \mathbb{Z}_+ \mapsto \mathbb{R}$ given by $h(n) = \max\{r + f(n+1), f(n)\}$ is also discrete concave.*

In the next lemma we derive an important property of expectations of discrete concave functions of the random variable $\mathbf{B}(p, n)$.

Lemma B.2 *If the function $f : \mathbb{Z}_+ \mapsto \mathbb{R}$ is discrete concave (convex), then the function $n \mapsto \mathbb{E}\left(f\left(\mathbf{B}(p, n)\right)\right)$ is also discrete concave (convex).*

Proof. We need to show that $n \mapsto \mathbb{E}\left(f\left(\mathbf{B}(p, n+1)\right)\right) - \mathbb{E}\left(f\left(\mathbf{B}(p, n)\right)\right)$ is decreasing (increasing). By the definition of $\mathbf{B}(p, n+1)$ given in relation (27) and the conditional expectation formula we obtain that

$$\begin{aligned} \mathbb{E}\left(f\left(\mathbf{B}(p, n+1)\right)\right) - \mathbb{E}\left(f\left(\mathbf{B}(p, n)\right)\right) &= p\mathbb{E}\left(f\left(\mathbf{B}(p, n+1)\right) - f\left(\mathbf{B}(p, n)\right) | \mathbf{U}_{n+1} \leq p\right) \\ &= p\mathbb{E}\left(f\left(1 + \mathbf{B}(p, n)\right) - f\left(\mathbf{B}(p, n)\right) | \mathbf{U}_{n+1} \leq p\right) \\ &= p\mathbb{E}\left(f\left(1 + \mathbf{B}(p, n)\right) - f\left(\mathbf{B}(p, n)\right)\right). \end{aligned} \tag{29}$$

Since $\mathbf{B}(p, n+1) \geq \mathbf{B}(p, n)$ and f is discrete concave (convex) we obtain that $n \mapsto f\left(1 + \mathbf{B}(p, n)\right) - f\left(\mathbf{B}(p, n)\right)$ is decreasing (increasing) and by relation (29) the result follows. □

For any non-negative random variable $\mathbf{D}$, we define the random variable $\mathbf{N}(n) = \min\{n, \mathbf{D}\}$.

Lemma B.3 *If $f : \mathbb{Z}_+ \mapsto \mathbb{R}$ is a discrete concave function and the optimization problem $\max\{f(n) : n \geq C\}$ has a finite optimal solution n_{opt}, then this is also an optimal solution of the problem $\max\left\{\mathbb{E}\left(f\left(\mathbf{N}(n)\right)\right) : n \geq C\right\}$.*

Proof. The discrete concavity of f implies its discrete unimodality and so we obtain for every $n \geq n_{opt}$ that

$$f(n+1) \leq f(n) \tag{30}$$

and for every $n < n_{opt}$

$$f(n+1) \geq f(n). \tag{31}$$

By the definition of $\mathbf{N}(n)$ it follows that

$$f(\mathbf{N}(n+1)) - f(\mathbf{N}(n)) \quad = \quad (f(n+1) - f(n))\mathbf{1}_{\{\mathbf{D} \geq n+1\}}.$$

This shows

$$\mathbb{E}\left(f(\mathbf{N}(n+1)) - f(\mathbf{N}(n))\right) = (f(n+1) - f(n))\mathbb{P}(\mathbf{D} \geq n+1) \tag{32}$$

and by relations (30),(31) and (32) we obtain

$$\mathbb{E}\left(f\left(\mathbf{N}(n+1)\right)\right) \leq \mathbb{E}\left(f\left(\mathbf{N}(n)\right)\right)$$

for every $n \geq n_{opt}$, and

$$\mathbb{E}\left(f\left(\mathbf{N}(n+1)\right)\right) \geq \mathbb{E}\left(f\left(\mathbf{N}(n)\right)\right)$$

for every $n < n_{opt}$. Hence, n_{opt} is also an optimal solution of problem $\max\left\{\mathbb{E}\left(f\left(\mathbf{N}(n)\right)\right) : n \geq C\right\}$. □

Appendix C. Bounds on Error Introduced by Using Jensen's Inequality. In this section, we derive upper bounds on the error committed by using Jensen's inequality in constructing the upper bounding approximation of the problem (P_I) when the random demands for fare classes, $\mathbf{D}_i$, $i = 1, \ldots, m$, are independent. We also analyze under which conditions the upper bounds on the error committed by using Jensen's inequality are loose or tight in detail.

We denote the nonnegative error that results from replacing the objective function in optimization problem (P_I) by the objective function in the upper bounding problem (P_I^{UB}) as

$$e_J(\mathbf{b}) = \mathbb{E}\left(\left[\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) - C\right]^+\right) - \left[\sum_{i=1}^m \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)) - C\right]^+. \tag{33}$$

In this section, we derive upper bounds on $e_J(\mathbf{b})$ by approximating the distribution of the random total number of show-ups by a Poisson distribution with the same expectation. This approximation provides us with an upper bound on the error term $e_J(\mathbf{b})$ by the next lemma, which is an immediate consequence of the fact that a summation of the binomial distributed random variables with different success probabilities is less variable than a Poisson distributed random variable with the same expectation. Recall that $\mathbf{N}_i(b_i) = \min\{\mathbf{D}_i, b_i\}$ for all fare classes $i = 1, \ldots, m$ with $\mathbf{D} = (\mathbf{D}_1, \ldots, \mathbf{D}_m)$ denoting the demand vector.

Lemma C.1 *If $\mathbf{Y}(\vartheta)$ denotes a Poisson distributed random variable with parameter ϑ, then it follows for every increasing convex function f that*

$$\mathbb{E}\left(f\left(\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i))\right)\right) \leq \mathbb{E}\left(f\left(\mathbf{Y}(\sum_{i=1}^m \beta_i^s \mathbf{N}_i(b_i))\right)\right).$$

Proof. It is known that (page 502, Ross (1996)) for any sequence of n independent standard uniform distributed random variables $\mathbf{U}_i$, $i = 1, \ldots, n$, and any sequence $p_i, i = 1, \ldots, n$, satisfying $0 < p_i < 1$

$$\mathbb{E}\left(f\left(\sum_{i=1}^n \mathbf{1}_{\{\mathbf{U}_i \leq p_i\}}\right)\right) \leq \mathbb{E}\left(f\left(\mathbf{Y}(\sum_{i=1}^n p_i)\right)\right)$$

holds true for any finite increasing convex function. This implies by using the independence of the random variables $\mathbf{N}_i(b_i), i = 1, \ldots, m$, and

$$\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i)) = \sum_{i=1}^m \sum_{k=1}^{\mathbf{N}_i(b_i)} \mathbf{1}_{\{\mathbf{U}_{ik} \leq \beta_i^s\}}$$

with $\mathbf{N}_i(b_i), i = 1, \ldots, m$, being also independent of the double infinite sequence $\mathbf{U}_{ik}$ that

$$\begin{aligned}\mathbb{E}\left(f\left(\sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i))\right)\right) \quad &\leq \quad \mathbb{E}\left(\mathbb{E}\left(f(\mathbf{Y}(\sum_{i=1}^m \beta_i^s \min\{\mathbf{D}_i, b_i\})) \mid \mathbf{D}\right)\right) \\ &= \quad \mathbb{E}(f(\mathbf{Y}(\sum_{i=1}^m \beta_i^s \mathbf{N}_i(b_i)))).\end{aligned}$$

Hence, we have shown the desired inequality. □

By Lemma C.1 applied to the increasing convex function $f(x) := [x - C]^+$ and relation (33) it follows for every $\mathbf{b} \in \mathbb{Z}_+^m$ that

$$e_J(\mathbf{b}) \leq \mathbb{E}(f(\mathbf{Y}(\mathbf{Z}(\mathbf{b})))) - f(\mu(\mathbf{b})) \tag{34}$$

with

$$\mathbf{Z}(\mathbf{b}) := \sum_{i=1}^{m} \beta_i^s \mathbf{N}_i(b_i), \quad \mu(\mathbf{b}) := \mathbb{E}(\mathbf{Z}(\mathbf{b})) = \sum_{i=1}^{m} \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)). \tag{35}$$

Introducing

$$h(\vartheta) = \mathbb{E}\left(f(\mathbf{Y}(\vartheta))\right), \tag{36}$$

we obtain

$$e_J(\mathbf{b}) \leq \mathbb{E}\left(f(\mathbf{Y}(\mathbf{Z}(\mathbf{b})))\right) - f(\mu(\mathbf{b})) = \left[\mathbb{E}\left(h(\mathbf{Z}(\mathbf{b}))\right) - h(\mu(\mathbf{b}))\right] + \left[h(\mu(\mathbf{b})) - f(\mu(\mathbf{b}))\right]. \tag{37}$$

Thus, the upper bound in (37) consists of the sum of the error caused by replacing the random $\mathbf{Z}(\mathbf{b})$ with its expectation $\mu(\mathbf{b})$ and the error caused by replacing a Poisson distributed random variable $\mathbf{Y}(\vartheta)$ with its expectation ϑ.

Before analyzing the upper bound in relation (37) in detail we present a useful result. Let us denote the jth derivative of a function h at x by $h^{(j)}(x)$.

Lemma C.2 *If $\mathbf{Y}(\vartheta)$ denotes a Poisson distributed random variable with parameter ϑ, then for any function $f : (0, \infty) \to \mathbb{R}$ and $h : (0, \infty) \to \mathbb{R}$ given by (36) the derivative of h with respect to ϑ exists for every $\vartheta > 0$ and it is given by*

$$h^{(1)}(\vartheta) = \mathbb{E}(f(\mathbf{Y}(\vartheta) + 1)) - \mathbb{E}(f(\mathbf{Y}(\vartheta))). \tag{38}$$

The proof easily follows from expressing the expectation of $f(\mathbf{Y}(\vartheta))$ for any function $f : (0, \infty) \to \mathbb{R}$ and taking the first derivative. Moreover, by a standard sample path argument and the relation that $\mathbf{Y}(\vartheta_1)$ is stochastically larger than $\mathbf{Y}(\vartheta_2)$ for $\vartheta_1 \geq \vartheta_2$ it follows by (38) that for any convex (concave) function f the function $\vartheta \mapsto \mathbb{E}(f(\mathbf{Y}(\vartheta)))$ is convex (concave).

By the convexity of the function h given in relation (36) and Jensen's inequality we obtain

$$\mathbb{E}\left(h(\mathbf{Z}(\mathbf{b}))\right) - h(\mu(\mathbf{b})) \geq 0.$$

Since $f(x) := [x - C]^+$ is also a convex function, it follows again by Jensen's inequality that

$$h(\mu(\mathbf{b})) - f(\mu(\mathbf{b})) = \mathbb{E}\left(f(\mathbf{Y}(\mu(\mathbf{b}))) - f(\mathbb{E}(\mathbf{Y}(\mu(\mathbf{b}))))\right) \geq 0.$$

We focus on these two types of nonnegative errors to analyze the upper bound given in (37). We first analyse the error term $h(\mu(\mathbf{b})) - f(\mu(\mathbf{b}))$. To do this we introduce the function $\epsilon : (0, \infty) \to \mathbb{R}$ given by

$$\epsilon(\vartheta) = \mathbb{E}\left([\mathbf{Y}(\vartheta) - C]^+\right) - [\vartheta - C]^+ \tag{39}$$

with $\mathbf{Y}(\vartheta)$ denoting a Poisson distributed random variable with parameter ϑ. Clearly,

$$\epsilon(\mu(\mathbf{b})) = \mathbb{E}\left([\mathbf{Y}(\mu(\mathbf{b})) - C]^+\right) - [\mu(\mathbf{b}) - C]^+ = h(\mu(\mathbf{b})) - f(\mu(\mathbf{b})). \tag{40}$$

In the next result, we provide the value of ϑ maximizing the error function $\vartheta \mapsto \epsilon(\vartheta)$ and tight upper bounds for it.

Lemma C.3 *The function $\epsilon : (0, \infty) \to \mathbb{R}$ attains its maximum at $\vartheta = C$. Moreover, for $\vartheta \leq C$ it follows that*

$$\epsilon(\vartheta) \leq \frac{\exp(C)C!}{C^C} \mathbb{P}(\mathbf{Y}(\vartheta) \geq C + 1), \tag{41}$$

while for $\vartheta > C$

$$\epsilon(\vartheta) \leq \frac{\exp(C-1)(C-1)!}{(C-1)^{C-1}} \mathbb{P}(\mathbf{Y}(\vartheta) \leq C - 1). \tag{42}$$

Proof. For $f(x) = [x - C]^+$ it follows by relation (38) that the derivative of the function $\vartheta \mapsto \mathbb{E}(f(\mathbf{Y}(\vartheta)))$ is given by the continuous function $\vartheta \to \mathbb{P}(\mathbf{Y}(\vartheta) \geq C)$. This implies using relation (39) that for $0 < \vartheta < C$

$$\epsilon^{(1)}(\vartheta) = \mathbb{P}(\mathbf{Y}(\vartheta) \geq C) > 0 \tag{43}$$

and for $\vartheta > C$

$$\epsilon^{(1)}(\vartheta) = \mathbb{P}(\mathbf{Y}(\vartheta) \geq C) - 1 = -\mathbb{P}(\mathbf{Y}(\vartheta) \leq C - 1) < 0.$$

Thus, the the function $\epsilon : (0, \infty) \to \mathbb{R}$ attains its maximum at $\vartheta = C$.

To show the first inequality in (41) we note, using $\mathbf{Y}(0) = 0$ with probability 1, that $\mathbb{E}f(\mathbf{Y}(0)) = 0$. This implies by the main theorem of integration and (43) that for any $\vartheta \leq C$

$$\begin{aligned} \epsilon(\vartheta) &= \mathbb{E}(f(\mathbf{Y}(\vartheta))) = \mathbb{E}(f(\mathbf{Y}(\vartheta))) - \mathbb{E}(f(\mathbf{Y}(0))) \\ &= \int_0^\vartheta \mathbb{P}(\mathbf{Y}(v) \geq C)dv. \end{aligned} \tag{44}$$

To bound the probability $\mathbb{P}(\mathbf{Y}(v) \geq C)$ in (44) we observe applying Markov's inequality and the moment generating function of a Poisson distributed random variable $\mathbf{Y}(v)$ given by

$$\mathbb{E}(\exp(s\mathbf{Y}(v))) = \exp(-v(1 - \exp(s))), s \in \mathbb{R},$$

that for every $s \geq 0$

$$\mathbb{P}(\mathbf{Y}(v) \geq C) = \mathbb{P}(\exp(s\mathbf{Y}(v)) \geq \exp(sC)) \leq \exp(-v(1 - \exp(s)) - sC). \tag{45}$$

Since this upper bound holds for every $s \geq 0,\ 0 < v \leq \vartheta \leq C$, and the function $s \mapsto -v(1 - \exp(s)) - sC$ attains its minimum at $\ln(Cv^{-1}) \geq 0$, it follows by (45) that

$$\mathbb{P}(\mathbf{Y}(v) \geq C) \leq \frac{\exp(-v)v^C \exp(C)}{C^C}. \tag{46}$$

This implies by relation (44) that

$$\epsilon(\vartheta) \leq \frac{\exp(C)C!}{C^C} \int_0^\vartheta \exp(-v)\frac{v^C}{C!}dv = \frac{\exp(C)C!}{C^C}\mathbb{P}\left(\sum_{i=1}^{C+1} \mathbf{X}_i \leq \vartheta\right), \tag{47}$$

where $\mathbf{X}_i,\ i = 1, \ldots, C$, are independent and exponentially distributed random variables with parameter 1. Introducing now a Poisson process with arrival rate 1 by $\mathbf{W} = \{\mathbf{W}(t) : t \geq 0\}$ and using

$$\mathbb{P}(\sum_{i=1}^{C+1} \mathbf{X}_i \leq \vartheta) = \mathbb{P}(\mathbf{W}(\vartheta) \geq C + 1)$$

and $\mathbf{W}(\vartheta)$ has a Poisson distribution with parameter ϑ, the inequality in (41) follows from (47).

To show the second inequality in (42), we first observe by relation (39) that for every $\vartheta > C$

$$\epsilon(\vartheta) = -\mathbb{E}\left(\min\{\mathbf{Y}(\vartheta) - C, 0\}\right). \tag{48}$$

According to relation (38) the derivative of the function $\vartheta \mapsto \mathbb{E}(\min\{\mathbf{Y}(\vartheta) - C, 0\})$ is given by $\vartheta \mapsto \mathbb{P}(\mathbf{Y}(\vartheta) \leq C-1)$. This implies using $\lim_{\vartheta\uparrow\infty} \mathbb{E}(\min\{\mathbf{Y}(\vartheta)-C, 0\}) = 0$ and the main theorem of integration that

$$-\mathbb{E}(\min\{\mathbf{Y}(\vartheta) - C, 0\}) = \int_\vartheta^\infty \mathbb{P}(\mathbf{Y}(v) \leq C - 1)dv. \tag{49}$$

By reapplying Markov's inequality and the moment generating function of a Poisson distributed random variable $\mathbf{Y}(v)$, we obtain for every $s > 0$ and $\vartheta > C$

$$\begin{aligned} \mathbb{P}(\mathbf{Y}(v) \leq C - 1) &= \mathbb{P}(\exp(-s\mathbf{Y}(v)) \geq \exp(-s(C - 1))) \\ &\leq \exp(-v(1 - \exp(-s)) + s(C - 1)). \end{aligned}$$

Since this upper bound holds for every $s > 0,\ v > \vartheta > C$, and the function $s \mapsto -v(1-\exp(-s))+s(C-1)$ attains its minimum at $\log(v(C - 1)^{-1}) > 0$, it follows that

$$\mathbb{P}(\mathbf{Y}(v) \leq C - 1) \leq \frac{\exp(-v)v^{C-1} \exp(C - 1)}{(C - 1)^{C-1}}. \tag{50}$$

Hence, by relation (49)

$$
\begin{aligned}
-\mathbb{E}(\min\{\mathbf{Y}(\theta)-C,0\}) &\leq \tfrac{\exp(C-1)(C-1)!}{(C-1)^{C-1}}\int_{\vartheta}^{\infty}\exp(-v)\tfrac{v^{C-1}}{(C-1)!}dv \\
&= \tfrac{\exp(C-1)(C-1)!}{(C-1)^{C-1}}\mathbb{P}(\textstyle\sum_{i=1}^{C}\mathbf{X}_i>\vartheta)
\end{aligned}
\tag{51}
$$

holds true. Since

$$\mathbb{P}(\sum_{i=1}^{C}\mathbf{X}_i>\vartheta)=\mathbb{P}(\mathbf{W}(\vartheta)\leq C-1),$$

the desired relation (42) follows. □

Using Lemma C.3 and Stirling's formula, given by

$$\lim_{n\uparrow\infty}\frac{n!}{\sqrt[2]{2\pi}n^{n+\frac{1}{2}}\exp(-n)}=1,$$

we also derive approximate upper bounds on the error function $\epsilon(\vartheta)$ given in relation (39). Applying Stirling's formula and observing that this ratio approaches very fast to 1 (see formula 9.15 of Feller (1968)) we obtain for every $\vartheta\leq C$

$$\frac{\exp(C)C!}{C^C}\mathbb{P}(\mathbf{Y}(\vartheta)\geq C+1)\approx\sqrt[2]{2\pi C}\ \mathbb{P}(\mathbf{Y}(\vartheta)\geq C+1)$$

and for every $\vartheta>C$

$$\frac{\exp(C-1)(C-1)!}{(C-1)^{C-1}}\mathbb{P}(\mathbf{Y}(\vartheta)\leq C-1)\approx\sqrt[2]{2\pi(C-1)}\ \mathbb{P}(\mathbf{Y}(\vartheta)\leq C-1).$$

This shows that the maximum value of the error function $\epsilon(\vartheta)$ is of the order $\sqrt[2]{C}$, since the maximum is attained at $\vartheta=C$. Moreover, by the central limit theorem applied to the random variables $\mathbf{Y}(C)$ for $C\uparrow\infty$ this bound is asymptotically tight.

We next derive upper bound on the error term $\mathbb{E}(h(\mathbf{Z}(\mathbf{b})))-h(\mu(\mathbf{b}))$ in relation (37) by analyzing the difference $\mathbb{E}(h(\mathbf{X}))-h(\mathbb{E}(\mathbf{X}))$ for any nonnegative random variable $\mathbf{X}$ having a finite variance. Our proposed bound, presented in the next lemma, is based on a second order Taylor approximation, but note that it can be improved using a fourth order Taylor approximation.

Lemma C.4 *If $\mathbf{X}$ is a random variable on $\mathbb{R}_+$ with a finite variance $\sigma^2(\mathbf{X})$ then*

$$0\leq\mathbb{E}(h(\mathbf{X}))-h(\mathbb{E}(\mathbf{X}))\leq\frac{\sigma^2(\mathbf{X})}{2}\exp(-(C-1))\frac{(C-1)^{C-1}}{(C-1)!}. \tag{52}$$

Proof. Since $h(\vartheta)=\mathbb{E}(f(\mathbf{Y}(\vartheta)))$ with $f(x)=[x-C]^+$ it follows by relation (38) that

$$h^{(1)}(\vartheta)=\mathbb{P}(\mathbf{Y}(\vartheta)\geq C).$$

Then, we can easily obtain the second and the third derivatives of the function h as follows:

$$h^{(2)}(\vartheta)=\mathbb{P}(\mathbf{Y}(\vartheta)=C-1)=\exp(-\vartheta)\frac{\vartheta^{C-1}}{(C-1)!} \tag{53}$$

and

$$h^{(3)}(\vartheta)=\frac{\exp(-\vartheta)\vartheta^{C-2}}{(C-2)!}\left(1-\frac{\vartheta}{C-1}\right).$$

Since the function $\vartheta\mapsto h^{(3)}(\vartheta)$ is positive on $(0,C-1)$ and negative on $(C-1,\infty)$, the function $\vartheta\mapsto h^{(2)}(\vartheta)$ is increasing on $(0,C-1)$ and decreasing on $(C-1,\infty)$ with the maximum objective value of

$$\mathbb{P}(\mathbf{Y}(C-1)=C-1)=\exp(-(C-1))\frac{(C-1)^{C-1}}{(C-1)!}.$$

Using this maximum as an upper bound on the value of the function $h^{(2)}(\vartheta)$ for any ϑ, there exists by Taylor's theorem (see, e.g., (Goldberg, 1965) for every $\vartheta>0$ some point ξ_ϑ between ϑ and $\mathbb{E}(\mathbf{X})$ satisfying

$$
\begin{aligned}
h(\vartheta)-h(\mathbb{E}(\mathbf{X})) &= (\vartheta-\mathbb{E}(\mathbf{X}))h^{(1)}(\mathbb{E}(\mathbf{X}))+\tfrac{(\vartheta-\mathbb{E}(\mathbf{X}))^2}{2}h^{(2)}(\xi_\vartheta) \\
&\leq (\vartheta-\mathbb{E}(\mathbf{X}))h^{(1)}(\mathbb{E}(\mathbf{X}))+\tfrac{(\vartheta-\mathbb{E}(\mathbf{X}))^2}{2}\exp(-(C-1))\tfrac{(C-1)^{C-1}}{(C-1)!}.
\end{aligned}
\tag{54}
$$

By relation (54) the assertion follows. □

We now combine the results on the two types of errors, which are presented in Lemma C.3 and C.4, to derive (theoretical) upper bounds on the error $e_J(\mathbf{b})$ induced by using Jensen's inequality. As seen in the next lemma, our proposed upper bounds depend both on the variance of $\mathbf{Z}(\mathbf{b}) = \sum_{i=1}^m \beta_i^s \mathbf{N}_i(b_i)$, and the closeness of the expectation of $\mathbf{Z}(\mathbf{b})$ to C.

Lemma C.5 *It follows for every* $\mathbf{b} \in \mathbb{Z}_+^m$ *and* $\mu(\mathbf{b}) = \mathbb{E}(\mathbf{Z}(\mathbf{b})) \leq C$ *that*

$$e_J(\mathbf{b}) \leq \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2} \exp(-(C-1)) \frac{(C-1)^{C-1}}{(C-1)!} + \frac{\exp(C)C!}{C^C} \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b})) \geq C+1) \tag{55}$$

while for $\mu(\mathbf{b}) = \mathbb{E}(\mathbf{Z}(\mathbf{b})) > C$

$$e_J(\mathbf{b}) \leq \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2} \exp(-(C-1)) \frac{(C-1)^{C-1}}{(C-1)!} + \frac{\exp(C-1)(C-1)!}{(C-1)^{C-1}} \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b})) \leq C-1). \tag{56}$$

Proof. By relations (37) and (40) we obtain that

$$e_J(\mathbf{b}) \leq \mathbb{E}(h(\mathbf{Z}(\mathbf{b}))) - h(\mu(\mathbf{b})) + \epsilon(\mu(\mathbf{b})).$$

Then, the desired inequalities follow from Lemma C.3 and C.4. □

Clearly, by the independence of the random demand variables $\mathbf{D}_i,\ i = 1, \ldots, m$, and hence, the independence of the random variables $\mathbf{N}_i(b_i),\ i = 1, \ldots, m$, we have

$$\sigma^2(\mathbf{Z}(\mathbf{b})) = \sum_{i=1}^m (\beta_i^s)^2 \sigma^2(\mathbf{N}_i(b_i)).$$

Lemma C.5 explains under which conditions the upper bound on the error committed by using Jensen's inequality is large. It is easy to see that if $\mu(\mathbf{b})$ is closer to C and/or the variability in the random variable $\mathbf{Z}(\mathbf{b})$ is higher, we have a larger upper bound value.

Calculating The Actual Error Introduced by Using Jensen's Inequality. The actual error committed by using Jensen's inequality to obtain the upper bounding problem is given by (33). When the random demands for fare classes, $\mathbf{D}_i,\ i = 1, \ldots, m$, are independent, for a given booking policy denoted by $\mathbf{b} \in Z_+^m$ we can numerically calculate the value of the exact error $e_J(\mathbf{b})$ using the Fast Fourier Transform (FFT) method (see, e.g., Tijms, H.C, 2003). Basically, we need to compute numerically the distribution function of the bounded random variable

$$\mathbf{\Delta}(\mathbf{b}) := \sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N_i}(b_i)).$$

To achieve this, we compute the generating function of the random variable $\Delta(\mathbf{b})$. By the independence of the random demand variables $\mathbf{D}_i,\ i = 1, \ldots, m$, and hence, the independence of the random variables $\mathbf{N}_i(b_i), i = 1, \ldots, m$, and relation (28), we obtain the generating function as follows:

$$\begin{aligned} \mathbb{E}(z^{\mathbf{\Delta}(\mathbf{b})}) &= \Pi_{i=1}^m \mathbb{E}(z^{\mathbf{B}(\beta_i^s, \mathbf{N}_i(b_i))}) \\ &= \Pi_{i=1}^m \mathbb{E}\left((1 - \beta_i^s + \beta_i^s z)^{\mathbf{N}_i(b_i)}\right) \\ &= \Pi_{i=1}^m \mathcal{P}_i(1 - \beta_i^s + \beta_i^s z), \end{aligned}$$

where $\mathcal{P}_i(w) := \mathbb{E}(w^{\mathbf{N}_i(b_i)})$. Notice that $\mathcal{P}_i(w)$ can be easily calculated for given distributions of the random demand variables $\mathbf{D}_i, i = 1, \ldots, m$. When $\mathbf{D}_i$ is Poisson distributed with parameter λ_i for all $i = 1, \ldots, m$, it follows that

$$\mathcal{P}_i(w) = \exp(-\lambda) \sum_{k=0}^{b_i-1} w^k \frac{\lambda_i^k}{k!} + w^{b_i} \left(1 - \sum_{k=0}^{b_i-1} \frac{\exp(-\lambda_i)\lambda_i^k}{k!}\right).$$

Since the random variable $\Delta(\mathbf{b})$ is bounded with possible values $\{0, \ldots, \sum_{i=1}^m b_i\}$, we apply the standard FFT method for a finite sequence using $\mathbb{E}(z^{\mathbf{\Delta}(\mathbf{b})})$ and obtain the distribution function of $\Delta(\mathbf{b})$. Then, we simply compute the expectation $\mathbb{E}([\Delta(\mathbf{b}) - C]^+)$. The second term $[\sum_{i=1}^m \beta_i \mathbb{E}(\mathbf{N}_i(b_i)) - C]^+$ in (33) can easily be computed by a more simpler way; either by directly computing the cdf of $\mathbf{N}_i(b_i)$ or using the FFT method to compute the cdf of the bounded random variable $\mathbf{N}_i(b_i)$.

Error Introduced by Solving The Upper Bounding Problem. Here we present bounds to quantify the magnitude of the error introduced by solving the approximate optimization problem (P_I^{UB}) instead of the originally proposed problem (P_I) when the random demands for fare classes, $\mathbf{D}_i$, $i = 1, \ldots, m$, are independent. To derive these bounds we use the results obtained so far in the beginning of this section and the optimal function value of the lower bounding problem.

To denote the upper bounds presented in Lemma C.5, we introduce $\epsilon_1(\mathbf{b})$ and $\epsilon_2(\mathbf{b})$ given by

$$\epsilon_1(\mathbf{b}) = \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2} \exp(-(C-1)) \frac{(C-1)^{C-1}}{(C-1)!} + \frac{\exp(C)C!}{C^C} \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b}) \geq C+1) \tag{57}$$

and

$$\epsilon_2(\mathbf{b}) = \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2} \exp(-(C-1)) \frac{(C-1)^{C-1}}{(C-1)!} + \frac{\exp(C-1)((C-1)!}{(C-1)^{C-1}} \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b}) \leq C-1). \tag{58}$$

Lemma C.6 *If $\mathbf{b}^*$, $\mathbf{b}_U^*$ and $(\mathbf{b}_L^*, \mathbf{y}_L^*)$ denote the optimal solutions of the original problem (P_I), the upper bounding problem (P_I^{UB}), and the lower bounding problem (P_I^{LB}), respectively, then*

$$1 \leq \frac{\phi_U(\mathbf{b}_U^*)}{\phi(\mathbf{b}^*)} \leq 1 + \frac{\theta e_J(\mathbf{b}_U^*)}{\phi_L(\mathbf{b}_L^*, \mathbf{y}_L^*)} \leq 1 + \frac{\theta\left(\epsilon_1(\mathbf{b}_U^*)\mathbf{1}_{\{\mu(\mathbf{b}_U^*) \leq C\}} + \epsilon_2(\mathbf{b}_U^*)\mathbf{1}_{\{\mu(\mathbf{b}_U^*) > C\}}\right)}{\phi_L(\mathbf{b}_L^*, \mathbf{y}_L^*)}, \tag{59}$$

where the exact error $e_J(\mathbf{b})$ is given in relation (33) and $\mu(\mathbf{b}) = \sum_{i=1}^m \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i))$.

Proof. Since $\phi(\mathbf{b}^*) \geq \phi(\mathbf{b}_U^*)$ and $\phi_U(\mathbf{b}_U^*) \geq \phi(\mathbf{b}^*) \geq \phi_L(\mathbf{b}_L^*, \mathbf{y}_L^*) \geq 0$, we have

$$1 \leq \frac{\phi_U(\mathbf{b}_U^*)}{\phi(\mathbf{b}^*)} = \frac{\phi(\mathbf{b}_U^*) + \theta e_J(\mathbf{b}_U^*)}{\phi(\mathbf{b}^*)} \leq 1 + \frac{\theta e_J(\mathbf{b}_U^*)}{\phi_L(\mathbf{b}_L^*, \mathbf{y}_L^*)}. \tag{60}$$

By Lemma (C.5) it follows for every $\mathbf{b} \in \mathbb{Z}_+^m$ that

$$e_J(\mathbf{b}) \leq \epsilon_1(\mathbf{b})\mathbf{1}_{\{\mu(\mathbf{b}) \leq C\}} + \epsilon_2(\mathbf{b})\mathbf{1}_{\{\mu(\mathbf{b}) > C\}}.$$

This shows by (60) that the last inequality in (59) holds. □

This lemma demonstrates how the lower bounding problem is used to compare the quality of the solution obtained by the approximate optimization problem (P_I^{UB}) against the one obtained by the exact optimization problem (P_I). Note that a tighter bound can be obtained by using the second fraction in relation (59). However, this bound, which we refer to as the actual upper bound, requires computing $e_J(\mathbf{b}_U^*)$. As discussed in Section C, we can numerically evaluate this actual error term by using the FFT method which computes numerically the distribution of the bounded random variable

$$\mathbf{\Delta}(\mathbf{b}_U^*) := \sum_{i=1}^m \mathbf{B}(\beta_i^s, \mathbf{N}(b_{iU}^*)).$$

Note that it is computationally challenging to calculate the upper bounds $\epsilon_1(\mathbf{b})$ and $\epsilon_2(\mathbf{b})$ when C is large. Therefore, in the performed computational study, we use the Stirling's approximation and calculate the approximated upper bounds:

$$\begin{aligned} \epsilon_1(\mathbf{b}) &\approx \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2\sqrt[2]{2\pi(C-1)}} + \sqrt[2]{2\pi C}\ \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b})) \geq C+1) \qquad \text{if } \mu(\mathbf{b}) = \mathbb{E}(\mathbf{Z}(\mathbf{b})) \leq C \\ \epsilon_2(\mathbf{b}) &\approx \frac{\sigma^2(\mathbf{Z}(\mathbf{b}))}{2\sqrt[2]{2\pi(C-1)}} + \sqrt[2]{2\pi(C-1)}\ \mathbb{P}(\mathbf{Y}(\mu(\mathbf{b})) \leq C-1) \quad \text{if } \mu(\mathbf{b}) = \mathbb{E}(\mathbf{Z}(\mathbf{b})) > C. \end{aligned} \tag{61}$$

Appendix D. Determining Upper Bounds on The Booking Limits. In Section 3.2 we introduce upper bounds on the booking limits to formulate the upper bounding problem (P_I^{UB}) as a mixed-integer linear program. In this section, we propose a method to determine those upper bounds in a proper way. Our objective is to restrict the feasible region of the upper bounding problem to a box, in other words, introduce bounding constraints $b_i \leq M_i$, $i = 1, \ldots, m$, in such a way that the error we make in calculating the objective function is significantly small. Our proposed approach is based on the next lemma.

LEMMA D.1 *Suppose that we consider the optimization problem* $\max\{h(\mathbf{b}) : \mathbf{b} \in \mathbb{Z}_+^m\}$ *with*

$$h(\mathbf{b}) = \sum_{i=1}^m f_i(b_i) - g(\mathbf{b}).$$

If the functions $f_i,\ i = 1, \ldots, m$, *and* g *are increasing and* $\lim_{b\uparrow\infty} f_i(b_i) = f_i(\infty) < \infty,\ i = 1, \ldots, m$, *then for every* $\epsilon > 0$ *there exists a box* B *such that for every* $\mathbf{b} \in \mathbb{Z}_+^m$ *one can find a* $\hat{\mathbf{b}} \in B \subseteq \mathbb{Z}_+^m$ *satisfying*

$$h(\mathbf{b}) - h(\hat{\mathbf{b}}) \leq m\epsilon.$$

PROOF. Since $\lim_{b\uparrow\infty} f_i(b_i) = f_i(\infty)$, there exists for every $\epsilon > 0$ some $b_i(\epsilon)$ such that

$$f_i(\infty) \leq f_i(b_i(\epsilon)) + \epsilon \quad \forall i = 1, \ldots, m.$$

Consider the box $B = \{\mathbf{b} \in \mathbb{Z}_+^m : b_i \leq b_i(\epsilon),\ i = 1, \ldots, m\}$ and let $\mathbf{b} \notin B$. This shows that the set $I = \{i = 1, \ldots, m \ :\ b_i > b_i(\epsilon)\}$ is nonempty and take $\hat{\mathbf{b}} = \{\hat{b}_1, \ldots, \hat{b}_m\}$ with

$$\hat{b}_i = \begin{cases} b_i(\epsilon) & \text{if } i \in I \\ b_i & \text{otherwise} \end{cases}$$

Clearly $\hat{\mathbf{b}}$ belongs to B and $\mathbf{b} \geq \hat{\mathbf{b}}$. Using now the assumption that the functions $f_i,\ i = 1, \ldots, m$, and g are increasing we obtain

$$\begin{aligned} h(\mathbf{b}) - h(\hat{\mathbf{b}}) &= \textstyle\sum_{i=1}^m (f_i(b_i) - f_i(\hat{b}_i)) + g(\hat{\mathbf{b}}) - g(\mathbf{b}) \\ &\leq \textstyle\sum_{i=1}^m (f_i(\infty) - f_i(\hat{b}_i)) + g(\hat{\mathbf{b}}) - g(\mathbf{b}) \\ &\leq m\epsilon, \end{aligned}$$

and this shows the desired result. □

Observe that the objective function of the upper bounding problem can be written in the form of the function h given in Lemma D.1:

$$\phi_U(\mathbf{b}) = \sum_{i=1}^m f_i(b_i) - g(\mathbf{b})$$

with

$$f_i(b_i) = \tau_i \mathbb{E}(\mathbf{N}_i(b_i)) \text{ and } g(\mathbf{b}) = \theta \left[\sum_{i=1}^m \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)) - C\right]^+. \tag{62}$$

It is easy to see that the functions $f_i,\ i = 1, \ldots, m$, and g given in (62) are increasing. Since we assume that $\mathbb{E}(\mathbf{D}_i) < \infty$ for all $i = 1, \ldots, m$, we have $f_i(\infty) = \tau_i \mathbb{E}(\mathbf{D}_i) < \infty,\ i = 1, \ldots, m$. Thus, for a specified error term ϵ and given demand distributions we can find

$$f_i(\infty) - f_i(b_i(\epsilon)) \leq \epsilon \quad \forall i = 1, \ldots, m,$$

and considering the feasible region $\{b \in \mathbb{Z}_+^m \ :\ b_i \leq b_i(\epsilon),\ i = 1, \ldots, m\}$ instead of $\{b \in \mathbb{Z}_+^m\}$ would result in a deviation of at most $m\epsilon$ from the optimal objective function value.

In our computational study, we assume that $\mathbf{D}_i$ follows a Poisson distribution with parameter λ_i for all $i = 1, \ldots, m$ and they are independent. Under these assumptions, to restrict the feasible region of the upper bound problem to a box we first observe for $b_i \geq \lambda_i,\ i = 1, \ldots, m$, by relations (39) and (41) that

$$0 \leq f_i(\infty) - f_i(b_i) = \tau_i(\lambda_i - \mathbb{E}(\mathbf{N}_i(b_i))) = \tau_i \mathbb{E}(\,[\mathbf{D}_i - b_i]^+) \leq \tau_i \frac{\exp(b_i) b_i!}{b_i^{b_i}} \mathbb{P}(\mathbf{D}_i \geq b_i + 1). \tag{63}$$

Selecting now some integer $b_i(\epsilon) \geq \lambda_i$ satisfying

$$\tau_i \frac{\exp(b_i) b_i!}{b_i^{b_i}} \mathbb{P}(\mathbf{D}_i \geq b_i + 1) \leq \epsilon \tag{64}$$

we know by inequality (63) that for every $b_i > b_i(\epsilon)$ it must hold that

$$0 \leq \tau_i(\lambda_i - \mathbb{E}(\mathbf{N}_i(b_i))) \leq \tau_i(\lambda_i - \mathbb{E}(\mathbf{N}_i(b_i(\epsilon)))) \leq \epsilon.$$

Then, by Lemma D.1 it is guaranteed that $\phi_U(\mathbf{b}) - \phi_U(\hat{\mathbf{b}}) \leq m\epsilon$ for any $\hat{\mathbf{b}} \in B$ and $\mathbf{b} \geq \hat{\mathbf{b}}$.

Appendix E. Alternative Solution Method for The Upper Bounding Problem. Although its objective function $\max\{\phi_U(\mathbf{b})$, given in (12), is not separable, it is still possible to use dynamic programming to solve the problem (P_I^{UB}). The main idea is to partition the set of integers into two sets. Let

$$S_1 = \left\{\mathbf{b} \in \mathbb{Z}_+^m : \sum\nolimits_{i=1}^m \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)) \geq C\right\} \text{ and } S_2 := \left\{\mathbf{b} \in \mathbb{Z}_+^m : \sum\nolimits_{i=1}^m \beta_i^s \mathbb{E}(\mathbf{N}_i(b_i)) \leq C\right\}.$$

Clearly, $S_1 \cup S_2 = \mathbb{Z}_+^m$. Therefore, we have

$$\max\{\phi_U(\mathbf{b}) : \mathbf{b} \in \mathbb{Z}_+^m\} = \max\left\{\max\left\{\phi_U(\mathbf{b}) : \mathbf{b} \in S_1\right\}, \max\left\{\phi_U(\mathbf{b}) : \mathbf{b} \in S_2\right\}\right\}.$$

Thus, to compute $\phi_U(\mathbf{b})$, we need to take the maximum of the objective function values of the following two optimization problems

$$\max\left\{\phi_U(\mathbf{b}) : \mathbf{b} \in S_1\right\} = \theta C + \max\left\{\sum\nolimits_{i=1}^m (\tau_i - \theta\beta_i^s)\mathbb{E}(\mathbf{N}_i(b_i)) : \mathbf{b} \in S_1\right\} \tag{65}$$

and

$$\max\left\{\phi_U(\mathbf{b}) : \mathbf{b} \in S_2\right\} = \max\left\{\sum\nolimits_{i=1}^m \tau_i \mathbb{E}(\mathbf{N}_i(b_i)) : \mathbf{b} \in S_2\right\}. \tag{66}$$

Note that both problems (65) and (66) are separable and they can be solved by dynamic programming. However, we note that the implementation for solving problem (65) demands a special treatment. This is because of the greater-than-equal-to constraint, since one can check this constraint at each stage only when the bookings for all fare classes are known. To overcome this difficulty, we formulate (65) as a constrained shortest path problem and solve it using the well-known K-shortest path algorithm (Yen, 1971). This algorithm returns successively the first K paths from origin to destination on a graph. We apply the same algorithm to return several paths in decreasing order of $\phi_U(\mathbf{b})$ values until we find the first one that satisfies the constraint in (65). We also note that our upper bounding problem is similar to the approximate model proposed in (Chi, 1995, Section 2.3.4). However, Chi applies one more approximation to solve the resulting model, whereas we solve it to optimality.

References

Alstrup, J., Boas, S., Madsen, O. B. G., and Vidal, R. V. V. (1986). Booking policy for flights with two types of passengers. *European Journal of Operational Research*, 27:274–288.

Ball, M. O. and Queyranne, M. (2009). Toward robust revenue management: Competitive analysis of online booking. *Operations Research*, 57:950–963.

Beckman, J. M. (1958). Decision and team problems in airline reservations. *Econometrica*, 26:134–145.

Belobaba, P. (2006). Flight overbooking: Models and practice. Lecture notes, MIT, Boston, MA. Website.

Belobaba, P. P. (1987). Air travel demand and airline seat inventory management. Ph.D. thesis, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, MA.

Belobaba, P. P. (1989). Application of a probabilistic decision model to airline seat inventory control. *Operations Research*, 37:183–197.

Birbil, Ş. İ., Frenk, J. B. G., Gromicho, J. A. S., and Zhang, S. (2009). The role of robust optimization in single-leg airline revenue management. *Management Science*, 55:148–163.

Chatwin, R. E. (1998). Multi-period airline overbooking with a single fare class. *Operations Research*, 46:805–819.

Chatwin, R. E. (1999). Continuous-time airline overbooking with time dependent fares and refunds. *Transportation Science*, 33:182–191.

Chi, Z. (1995). *Airline Yield Management in a Dynamic Network Environment*. PhD thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Coughlan, J. (1999). Airline overbooking in the multi-class case. *Journal of the Operational Research Society*, 50:1098–1103.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol 1 (Third edition)*. Wiley, New York.

Feng, Y., Lin, P., and Xiao, B. (2002). An analysis of airline seat control with cancellations. In Yao, D. D. and H. Zhang, X. Y. Z., editors, *Stochastic Modeling and Optimization, with Applications in Queues, Finance, and Supply Chains*. Springer-Verlag, Berlin.

Goldberg, R. R. (1965). *Methods of Real Analysis*. Blaisdell Publishing Company, New York.

Huh, W. T. and Rusmevichientong, P. (2006). Adaptive capacity allocation with censored demand data: Application of concave umbrella functions. Technical report, Cornell University, School of Operations Research and Information Engineering, Ithaca, USA.

Karaesmen, I. and van Ryzin, G. J. (2004). Overbooking with substitutable inventory classes. *Operations Research*, 52:83–104.

Kunnumkal, S. and Topaloglu, H. (2009). A stochastic approximation method for the single-leg revenue management problem with discrete demand distributions. *Mathematical Methods of Operations Research*, 70:477–504.

Lan, Y., Ball, M. O., and Karaesmen, I. Z. (2011). Regret in overbooking and fare-class allocation for single leg. *Manufacturing & Service Operations Management*, 13(2):194–208.

Lan, Y., Gao, H., Ball, M. O., and Karaesmen, I. Z. (2008). Revenue management with limited demand information. *Management Science*, 54(9):1594–1609.

Lippman, S. A. and Stidham, S. (1977). Individual versus social optimization in exponential congestion systems. *Operations Research*, 25:233–247.

Phillips, R. L. (2005). *Pricing and Revenue Management*. Stanford University Press, Stanford, CA.

Ross, S. M. (1996). *Stochastic Processes*. Wiley.

Rothstein, M. (1971). An airline overbooking model. *Transportation Science*, 5:180–192.

Rothstein, M. (1985). OR and the airline overbooking problem. *Operations Research*, 33:237–248.

Smith, B. C., Leimkuhler, J. F., and Darrow, R. M. (1992). Yield management at American Airlines. *Interfaces*, 22:8–31.

Subramanian, J., Stidham, S., and Lautenbacher, C. (1999). Airline yield management with overbooking, cancellations, and no-shows. *Transportation Science*, 33:147–167.

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

Talluri, K. T. and van Ryzin, G. J. (2005). *The Theory and Practice of Revenue Management*. Springer, New York, NY.

Taylor, C. J. (1962). The determination of passenger booking levels. Technical report, AGIFORS Symposium Proceedings, American Airlines, New York, USA.

Thompson, H. R. (1961). Statistical problems in airline reservation control. *Operational Research Quarterly*, 12:167–185.

Tijms, H.C (2003). *A First Course in Stochastic Models.* Wiley, New York.

Topaloglu, H., Birbil, Ş. İ., Frenk, J., and Noyan, N. (2011). Tractable open loop policies for joint overbooking and capacity control over a single flight leg with multiple fare classes. *Transportation Science (to appear)*.

van Ryzin, G. J. and McGill, J. I. (2000). Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science*, 46:760–775.

Yen, J. Y. (1971). Finding the K shortest loopless paths in a network. *Management Science*, 17:712–716.