Submitted to *Operations Research*
manuscript OPRE-2008-08-420.R3

Cargo Capacity Management with Allotments and Spot Market Demand

Yuri Levin and Mikhail Nediak
School of Business, Queen's University, Kingston, ON, K7L3N6, Canada, {ylevin,mnediak}@business.queensu.ca

Huseyin Topaloglu
School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853, USA, topaloglu@orie.cornell.edu

We consider a problem faced by an airline that operates a number of parallel flights to transport cargo between a particular origin destination pair. The airline can sell its cargo capacity either through allotment contracts or on the spot market where customers exhibit choice behavior between different flights. The goal is to simultaneously select allotment contracts among available bids and find a booking control policy for the spot market so as to maximize the sum of the profit from the allotments and the total expected profit from the spot market. We formulate the booking control problem on the sport market as a dynamic program and construct approximations to its value functions, which can be used to estimate the total expected profit from the spot market. We show that our value function approximations provide upper bounds on the optimal total expected profit from the spot market and they allow us to solve the allotment selection problem through a sequence of linear mixed integer programs with a special structure. Furthermore, the value function approximations are useful for constructing a booking control policy for the spot market with desirable monotonic properties. Computational experiments show that the proposed approach can be scaled to realistic problems and provides well performing allotment allocation and booking control decisions.

Key words: transportation: freight; dynamic programming: applications; programming: integer/applications.

History:

---

1. Introduction

A significant portion of revenues in the airline industry comes from transporting cargo. Indeed, the International Air Transport Association (IATA) estimates that 2008 system-wide global revenues from cargo were 64 billion versus 439 billion from passengers, see IATA (2009). Many airlines face the problem of controlling cargo bookings for both dedicated cargo and mixed passenger/cargo aircraft. For a large airline, management of passenger capacity shares many features with management of cargo capacity. Given a limited amount of cargo capacity, an airline, typically called a *carrier*, decides whether to commit to a current booking request or to reserve capacity for a possible future booking request with potentially higher revenue. This basic tradeoff is accompanied by some other concerns, such as the existing overbooking policies which deal with the fact that not all of the booked requests show up at the departure time and the allotment contracts which reserve portions of cargo capacity to numerous clients. However, while passenger capacity management has received significant attention in the revenue management literature, there is nowhere near a comparable effort on cargo revenue management, in spite of its evident importance as a potential source of improvement in revenues.

In this paper, we consider an airline that operates a number of parallel flights to transport cargo between a particular origin-destination pair. The airline faces two problems that interact tightly with each other. The first problem is to determine what contracts, if any, should be signed with potential allotment customers. The allotment contracts typically fix the shipping rate and the amount of reserved capacity and they have a duration of at least a few months so that many departures are scheduled to occur during the contract period. The customers who opt for such contracts are usually intermediaries called *forwarders*, who provide the end customers with a door to door service by handling and transporting cargo packages. The airline periodically collects bids from forwarders for allotment contracts and simultaneously decides which bids are granted, making this problem akin to combinatorial auctions. After setting up the allotment contracts, the airline faces the second problem that determines which booking requests to accept on the spot market. Since the parallel flights are quite similar from a customer standpoint, spot market demands exhibit consumer choice behavior. The two problems interact because allotment and spot market cargo eventually share the same capacity, and the airline is penalized if this capacity is oversold.

There has been extensive research on controlling passenger bookings, but relatively little attention has been directed to cargo bookings. Although the trade-offs involved in controlling passenger and cargo bookings are similar, the methods for controlling passenger bookings do not immediately apply to cargo bookings for a number of reasons. To begin with, passenger capacity is counted in the number of seats, whereas cargo capacity is counted in units of both volume and weight. In certain applications that involve wide body aircraft, it may even be necessary to use body position as the third capacity dimension. Furthermore, the capacity requirements of a booked cargo request are usually not known with certainty until loading at the departure time. Thus, the total capacity required by the booked cargo requests is random and this creates complications when formulating the booking control problem as a dynamic program. Finally, the possibility of allotments introduces challenges. The airline can sell a portion of its cargo capacity as allotments at relatively low prices and hedge the randomness of the spot market. However, the amount of capacity utilized, eventually, by the allotment contracts is random, which implies that the amount of capacity available for the spot market demand is also random.

In general, the booking control problem is difficult by itself as it requires solving a dynamic program with a high dimensional state variable. Our proposed method approximates the value functions in the booking control problem and uses these approximations to estimate the total expected profit from the spot market. The value function approximations for the spot market problem provides a practical booking control policy. Furthermore, we make the allotment allocation decisions by maximizing the sum of the gross profit from the allotments and the estimated total expected profit from the spot market.

In this paper, we make the following research contributions. 1) We present a tractable model that integrates multiple allotment contracts and spot market bookings of an airline for a group of parallel flights. As far as we know, this appears to be the first operational model that treats such integration. 2) We construct well performing booking control policies for the booking requests that occur on the spot market. These policies are useful even when there are no allotment contracts and the airline sells all of its capacity on the spot market. 3) We show how to obtain upper bounds on the optimal total expected profit from the spot market. The upper bounds become useful when testing the quality of a booking control policy. Moreover, we demonstrate that the upper bounds provide a valuation tool for the spot market that can aid the airline in allotment contract negotiations. 4) The booking control policies used for the spot market are essentially bid price policies that use a revenue barrier for each cargo type. We show that the revenue barriers increase as we commit to more booking requests. This monotonic property is in alignment with the expectation that as we commit to more booking requests, we are less likely to accept a new booking request. 5) Computational experiments indicate that our approach can be applied to realistic problems and provides well performing allotment allocation and booking control decisions.

The rest of the paper is organized as follows. In §2, we provide a literature review. In §3, we formulate a basic optimization problem to maximize the total profit by allocating the capacity between the allotment contracts and the spot market. This problem requires knowledge of the total expected profit from the spot market and we capture this total expected profit by formulating a dynamic program for the spot market booking control problem. In §4, we develop a method to approximate the value functions in the dynamic programming formulation of the spot market booking control problem. In §5, we use these value function approximations to allocate the capacity between the allotment contracts and the spot market. In §6, we develop a booking control policy for the requests that occur on the spot market and establish desirable monotonic properties of this policy. In §7, we provide computational experiments. In §8, we conclude.

2. Literature Review

Booking control in cargo transportation is an active area of research. Kasilingam (1996) compares passenger and cargo booking control and discusses potential modeling approaches. This paper is the first one to lay out the main features of cargo booking control, including multiple dimensions of capacity and the presence of allotments. Blomeyer (2006) provides additional details on cargo booking practices. Several recent papers also review air cargo revenue management from a practical point of view. Billings et al. (2003) compare business models in the cargo and passenger cases and formulate the challenges for cargo revenue management. The authors emphasize key questions such as managing cargo capacity through allotment contracts, the tradeoff between the allotments and the spot market, as well as uncertainty in the actual cargo shipped compared to capacity booked. Slager and Kapteijns (2003) discuss the existing cargo revenue management implementation at

KLM. In particular, the authors discuss margin management on allotment contracts and on spot market shipments. According to this paper, the current practice of KLM is to collect bids for contracted capacity twice a year, at the start of IATA summer and winter schedules. The authors also mention that performance of the contracts is monitored on a weekly basis and contract utilization is expected to attain certain performance targets. An enhanced list of complexities for air cargo revenue management from a business perspective is given by Becker and Dill (2007).

Several papers explore the cargo booking problem over a single flight leg. Huang and Hsu (2005) address random capacity requirements of the booked cargo requests, but they assume that capacity is counted only in units of weight. Luo et al. (2009) and Moussawi and Cakanyildirim (2005) construct protection level policies for the booking control problem, where capacity is counted in units of both volume and weight. The models in these two papers assume that demand is a static random variable, ignoring the detailed temporal dynamics of the booking requests. Xiao and Yang (2010) give a continuous time formulation for the booking control problem in the ocean transport setting. Amaruchkul et al. (2007b) provide a dynamic programming formulation of the booking control problem that captures the detailed temporal dynamics of the booking requests. They develop numerous approximations to the value functions. Their dynamic programming formulation is similar to ours, but we endogenize the *off-loading problem* that decides which cargo bookings should not be loaded on the flight when there is insufficient capacity.

The control policy that we use for accepting or rejecting the booking requests on the spot market is similar to a bid price policy. In particular, we compute a revenue barrier for different types of booking requests, capturing the expected opportunity cost of the capacity used by the booking requests. There has been little work that uses bid prices in cargo booking control. Pak and Dekker (2005) propose a bid price policy for the booking control problem that takes place over a network of flights. Sandhu and Klabjan (2006) consider an integer programming formulation for the fleeting problem that includes a bid price-based booking control component for air cargo on a flight network. The bid prices in the paper are static, ignore the stochastic aspects of the booking problem and the performance of these controls is not evaluated. Lastly, Karaesmen (2001) uses infinite dimensional linear programs to compute bid prices and the proposed approach is applicable to cargo settings. Bid price policies are ubiquitous in passenger booking control. Simpson (1989) and Williamson (1992) were the first to compute bid prices by using a deterministic linear program. Talluri and van Ryzin (1998) show that the bid prices computed by this linear program are asymptotically optimal as the capacities and expected numbers of booking requests increase linearly with the same rate. One shortcoming of the deterministic linear program is that it uses only the expected numbers of the future booking requests. Talluri and van Ryzin (1999) try to make up for this shortcoming by proposing a randomized version that uses actual samples of the future booking requests. Bertsimas and Popescu (2003) provide extensions of the deterministic linear program to cover cancellations. Adelman (2007) and Topaloglu (2009) propose numerous strategies to approximate the value functions in the dynamic programming formulation of the booking control problem and their approximations provide bid price policies. The consumer choice model that we use for the spot market in our paper is similar to the approach developed by Talluri and van Ryzin (2004) for the passenger case.

There is little work done on allotment and spot market booking coordination. The existing literature focuses on the properties of an allotment contract, whereas we focus of operational decisions. Hellermann (2006) reviews the industry and market structure and provides a survey of related literature. The author also formulates a stylized static game theoretic model of a capacity contract between one forwarder and one carrier with one dimensional capacity. The contract consists of a reservation fee and an execution fee. The articles of Gupta (2008) and Amaruchkul et al. (2007a) are concerned with the form and properties of a contract between a carrier and a single forwarder under similar assumptions. Gupta (2008) considers two flexible schemes that allow the carrier to adjust contract parameters based on the realized demand and shows that this flexibility allows the carrier to achieve an efficient capacity allocation between the forwarder and the spot market. Amaruchkul et al. (2007a) study contracts that are described by a constant allotment, a lump sum payment for the season and a per-flight payment that depends on allotment utilization. They show when such a scheme can eliminate informational rents.

Other research areas related to our work include the analysis of dual-channel supply chains (since spot market can be viewed as a direct sales channel while the allotments is a form of wholesale distribution), analysis of advance sales in retail, and supply chain contracts. Chiang et al. (2003), Tsay and Agrawal (2004), Xie and Shugan (2001) and Cachon (2004) provide comprehensive reviews of these fields. The works in these fields usually focus on fundamental qualitative analysis of the marketplace and the supply chain. In relations to these bodies of literature, our contribution is an efficient operational approach to management of a large number of potential advance service contracts which balances supply and inventory risks in the presence of spot market.

3. Problem Formulation

We consider a problem faced by an airline that operates a collection of flights over a certain planning horizon to transport cargo between a particular origin destination pair. The flights are arranged into a repeated *schedule cycle* according to their departure time and the planning horizon includes multiple schedule cycles. A typical length of a cycle is one week, whereas the planning horizon is on the order of a few months. The airline sells cargo capacity on these flights either through allotment contracts or on the spot market. The allotment contracts are signed at the beginning of the planning horizon, whereas the booking requests on the spot market occur continuously. Both Hellermann (2006) and Billings et al. (2003) describe practices of major air cargo carriers which use allotment contracts of half a year or more in duration. The carriers sign the new contracts twice a year for winter and summer schedules based on the bids submitted by potential allotment customers. Therefore, the parameters of all potential allotment contracts result from an infrequent bidding and negotiation process. As the outcome of such a process, each potential allotment customer provides a list of acceptable contracts. For example, a lower rate may be acceptable for a larger allotment and vice versa. We assume that this list is available to the airline as a *bid* from this customer. The bid structure depends on specific practices of the airline and the type of the customer. There are two issues to consider for the allotment contracts. First, we need to describe the profits and capacity utilizations of allotment contracts. Second, we need to address the possibility that allotment customers may want to bid jointly for a combination of allotments on different flights.

Allotment profits and utilizations. We define an *allotment quote* as a combination of capacity utilization and revenue (payment by the allotment customer) for a specific flight in the schedule. The physical capacity utilized by an allotment contract on the flight is random, becomes known only at the departure time, and can be formally described as a random variable. Similarly, the revenue can be described as a random variable, whether it is a fixed fee or a function of the actual capacity utilization on the flight. Therefore, we treat the allotment quote as a random vector which captures the capacity utilization and revenue for a particular flight. This provides a homogeneous modeling description of various quote formats and information collection practices.

The allotment quote information is also sufficient to quantify the costs of providing capacity for the allotment on the flight. Generally, the allotments costs have a complex structure related to both the direct costs of providing capacity, such as fuel and handling, the penalty costs associated with a failure to fulfill shipment obligations and the indirect opportunity costs of displacing potential spot market sales. The airlines already have techniques for estimation of direct costs. The revenue less the variable costs per shipment is called a contribution margin of a shipment and is usually known in practice. Slager and Kapteijns (2003) give a detailed discussion of margin management principles. The penalty costs are associated with additional handling, alternative shipping arrangements and the loss of goodwill but can also be estimated given information about allotment utilization. The opportunity costs are perhaps the hardest to quantify and require an appropriate model of the spot market booking control process, which we explicitly do in our paper.

Multi-unit bids. The rationale for multi-unit bids is two-fold. The first one is that allotments usually cover a set of repeated flights, such as all scheduled 6pm flights between a particular OD pair on Tuesdays. Such sets of flights are convenient to describe in combination, since the same capacity utilization and payment structure applies. Moreover, potential customers expect to obtain all flights in such sets together. Therefore, to express this *indivisibility*, we employ a notion from the combinatorial auctions theory and define *atomic bids* as sets of allotment quotes which can only be granted in combination. The quotes comprising atomic bids in our model play the role similar to that of regular items in conventional combinatorial auctions.

The second rationale is the phenomenon of consumer choice which is quite important for modelling allotments. The payments considered acceptable by a prospective allotment customer may depend on a combination of allotments granted to this customer by the airline. For example, a customer may want to ship cargo between a pair of locations both on Tuesday and on Thursday. A single large allotment on either day may be good but two smaller allotments on *both* of these days may be better. The three options are mutually exclusive.

The rest of this section is separated into two subsections. In §3.1, we formally introduce the model elements. In §3.2, we give the objectives of the airline and formulate the optimization problem.

3.1. Model elements and notation

Allotment contracts. Decisions for the allotment contracts are made at the beginning of the planning horizon and they remain fixed throughout. We use $\mathcal{J}$ to denote the set of flights operated by the airline over the whole planning horizon. At the beginning of the planning horizon, the airline receives a finite set $\mathcal{I}$ of atomic bids. Each atomic bid i specifies a set of flights $\mathcal{J}_i \subseteq \mathcal{J}$ that is used by this bid if an allotment contract on the terms of the atomic bid i is granted. We use the set $\mathcal{D}$ to denote the combined set of dummy items of all customers. Each dummy item has a one-to-one correspondence with a particular subset of mutually exclusive atomic bids. For each atomic bid i, we let $\mathcal{D}_i \subseteq \mathcal{D}$ to be the set of its dummy items. We represent a decision to grant a contract on the terms of the atomic bid i by a binary decision variable x_i. Since each dummy item $d \in \mathcal{D}$ indicates a membership of atomic bids in a particular subset $\{i \in \mathcal{I} : d \in \mathcal{D}_i\}$ of bids which cannot be jointly granted, the decision variables must satisfy a constraint that at most one of the variables in this subset is equal to one, $\sum_{i \in \mathcal{I}: d \in \mathcal{D}_i} x_i \leq 1$ for all $d \in \mathcal{D}$. Our use of dummy items to represent mutually exclusive atomic bids corresponds to the OR* bidding language in combinatorial auctions; see Nisan (2000).

We use V^a_{ij} to denote the capacity utilization of the atomic bid i on flight $j \in \mathcal{J}_i$. Noting the discussion at the beginning of this section, V^a_{ij} is a random variable and its value becomes known at the departure time of flight j. The randomness in V^a_{ij} captures a common issue of uncertainty in the actual capacity utilization of allotment contracts. As we mention in the introduction, capacity utilization in cargo revenue management is usually measured in units of volume and weight and if this is the case, then we can visualize V^a_{ij} as a two dimensional vector, each component capturing the volume and weight utilization. With this interpretation, we can handle the multi-dimensional capacity utilization with no modification in our development. We use R^a_{ij} to denote the revenue generated by the atomic bid i on flight j. As indicated by Slager and Kapteijns (2003), the margin and capacity utilization of each allotment contract are collected on a regular basis. This provides enough information to construct appropriate models for V^a_{ij} and R^a_{ij}. Moreover, an appropriate model for R^a_{ij} is available regardless of whether customers pay per capacity booked or per actual utilization. Hellermann (2006) refers to these types of contracts respectively as capacity purchasing agreements and guaranteed capacity agreements. Thus, we assume that (i) the joint distribution of nonnegative random vector (V^a_{ij}, R^a_{ij}) for each $i \in \mathcal{I}$ and $j \in \mathcal{J}_i$ is known, depends only on i and j, and has a finite first moment. Moreover, (ii) random vectors (V^a_{ij}, R^a_{ij}), $i \in \mathcal{I}$, $j \in \mathcal{J}_i$ are mutually independent within this collection as well as with any other sources of uncertainty in the model. Furthermore, the distributions of these vectors do not depend on the decisions in the model. The independence assumption (i)-(ii) aids in parsimonious specification of model inputs since only the joint distributions for each atomic bid/flight combination are necessary. Moreover, in practice, the same distribution can describe flights which are repeated throughout the schedule (e.g. 6pm flight on Tuesday for a particular OD pair). Assumption (i) also implies that capacity utilization for each quote does not depend on which atomic bids are granted (i.e. decisions x_i, $i \in \mathcal{I}$). This is reasonable with respect to decision about bids submitted by different bidders. With respect to decisions which apply to atomic bids of the same bidder, this assumption is reasonable as long as submitted bids accurately capture the partition of the expected load generated by the bidder between the atomic bids. In Online Appendix EC.3, we indicate how one can relax assumption (ii).

Temporal structure. Although decisions for allotment contracts remain fixed throughout the planning horizon, the spot market booking requests occur continuously over the whole planning horizon. We divide the planning horizon into τ time periods indexed by $\{1, \dots, \tau\}$. We assume that each time period corresponds to a small interval of time that there is at most one spot market booking request at each time period. This assumption is not problematic since the complexity of our model is insensitive to the number of time periods. The flights depart at different time periods over the planning horizon $\{1, \ldots, \tau\}$. For simplicity of exposition of the base model, we assume that the departures of all flights and the corresponding cargo loading decisions are independent. In particular, the cargo that does not fit on a particular flight results in a penalty cost, but does not occupy capacity on subsequent flights. We point out possible relaxations of this assumption later in the paper. Under independent departures assumption combined with independence of allotment capacity utilizations on different flights, it is irrelevant when the departures actually occur during the planning horizon since loading decisions associated with each flight do not affect the system state. Thus, we assume without loss of generality that all flights depart at time period $\tau + 1$. Independent departures assumption is justified when the penalty for rescheduling cargo to a later flight operated by the airline is comparable to that of seeking alternative shipping arrangements. This is realistic when competition enforces high customer service standards. Additional discussion of the relation between flight independence and assumption (i)-(ii) is provided in Online Appendix EC.3.

Spot market. We use $\mathcal{K}$ to denote the set of cargo types for which we can receive booking requests on the spot market. Following Amaruchkul et al. (2007b), we assume that cargo types differ in their capacity utilizations and profit margins. A spot market booking for cargo type k on flight j generates a revenue of R^s_{jk} and utilizes a capacity of V^s_{jk}. Similar to V^a_{ij}, the value of the random variable V^s_{jk} becomes known only at the departure time of flight j. Similarly to allotments, we assume that nonnegative random vectors (V^s_{jk}, R^s_{jk}) have a known joint distribution with a finite first moment which depends only on the combination of $j \in \mathcal{J}$ and $k \in \mathcal{K}$, and that these random vectors are mutually independent within the collection as well as with any other sources of uncertainty in the model. This assumption is reasonable if applied to capacity utilizations only. It is also reasonable if revenues are taken into account, as long as these revenues result from application of freight rates and rules which do not change throughout the planning horizon. For example, for two-dimensional V^s_{jk}, Amaruchkul et al. (2007b) consider R_{jk} which is computed via given type-dependent function of chargeable weight defined as the maximum of the actual cargo weight and its volume scaled by a given constant. An interesting practical concern is that the capacity utilization of a spot market booking for cargo type k may have different variability depending on when the booking is made. In particular, the uncertainty in the capacity utilization of the spot market bookings that are made later in the booking horizon may be smaller, since these spot market bookings are made with a shorter lead time for service. We can model different variability for spot market bookings by defining multiple cargo types depending on when the spot market booking is made. This approach amounts to working with the full set of cargo types $\{(k,t) : k \in \mathcal{K},\ t = 1, \dots, \tau\}$, where (k,t) corresponds to cargo type k booked on the spot market at time t. In this case, we simply need to adjust the random variables (V^s_{ij}, R^s_{ij}) as (V^s_{jkt}, R^s_{jkt}) so that V^s_{jkt} and R^s_{jkt} respectively correspond to the capacity utilization and revenue of a spot market booking of type k for flight j accepted at time period t. The random variables V^s_{jkt} and $V^s_{jkt'}$ may have different variances.

The spot market demands exhibit consumer choice behavior. In particular, the control exerted by the airline on the spot market is captured by the set $\mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}$, corresponding to the flight and cargo type combinations that are open for the spot market bookings. This is to say that $(j,k) \in \mathcal{S}$ if and only if flight j is open for cargo type k spot market bookings at the current time period. A spot market customer arriving into the system observes the set of open flight and cargo type combinations and makes a choice within this set. Given that the set of open flight and cargo type combinations at the time period t is $\mathcal{S}$, we use $P_{jkt}(\mathcal{S})$ to denote the probability that there is a spot market booking for flight j and cargo type k. The probabilities $\{P_{jkt}(\mathcal{S}) : \mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}\}$ are a part of the problem data and we naturally have $P_{jkt}(\mathcal{S}) = 0$ whenever $(j,k) \notin \mathcal{S}$ or time period t corresponds to a time period after the departure time of flight j. This model is a standard way of modeling consumer choice in revenue management; see Talluri and van Ryzin (2004). With respect to spot market demand process, we make an independence assumption similar to that on allotment and spot market capacity utilization and revenue uncertainty. In particular, probabilities

$P_{jkt}(\mathcal{S})$ depend only on flight $j \in \mathcal{J}$, cargo type $k \in \mathcal{K}$, time t and the set of open flight-cargo type combinations $\mathcal{S}$; and the booking events are mutually independent across time as well as with the other sources of uncertainty in the model.

3.2. Airline's objectives and optimization problem

The airline is interested in maximizing the total expected profit over the planning horizon by selecting the appropriate allotments from the available bids, controlling bookings on the spot market and managing the cargo loading process. We structure the optimization problem formulation according to these three decision areas.

Loading decisions. As mentioned before, the independent departures assumption allows us to assume that all flights depart at time period $\tau + 1$. Since the allotment contracts typically generate composite shipments, we assume that the airline may ship or fail to ship a part of the total cargo generated by an allotment contract. The regular cost of shipping fraction α of the allotment from atomic bid i on flight j is αC^a_{ij}. Likewise, the penalty cost of not accommodating fraction α of the allotment from atomic bid i on flight j is $\alpha \hat{C}^a_{ij}$. The costs $\hat{C}^a_{ij}$ and C^a_{ij} are treated in general as random variables dependent on capacity utilization V^a_{ij}. For spot market cargo, we assume that each unit is indivisible and let the regular and penalty costs for cargo type k on flight j be C^s_{jk} and $\hat{C}^s_{jk}$, respectively. These costs are also random variables dependent on capacity utilization V^s_{jk}. Different values of allotment and spot market penalties ($\hat{C}^a_{ij}$ and $\hat{C}^s_{jk}$) allow to balance the loading priorities of different allotment and spot market cargo booked for the same flight. If we set all $\hat{C}^s_{jk}$ to a sifficiently high constant and calibrate the parameter $\hat{C}^a_{ij}$ accordingly, then our model minimizes the expected number of rejected spot market bookings along with an accordingly calibrated cost for rejected allotment requests. The total available capacity on flight j is $\bar{V}_j$. Let n_{jk} be the total number of accepted spot market bookings for cargo type k on flight j. In this case, each spot market booking $l = 1, \dots, n_{jk}$ will have its own realization of costs and capacity requirements $(V^s_{jkl}, C^s_{jkl}, \hat{C}^s_{jkl})$. Using $n = \{n_{jk} : j \in \mathcal{J},\ k \in \mathcal{K}\}$ to denote the total numbers of accepted spot market bookings, $x = \{x_i : i \in \mathcal{I}\}$ to denote the accepted atomic bids, and $U = \{\{(V^s_{jkl}, C^s_{jkl}, \hat{C}^s_{jkl}) :$

$l=1,\dots,n_{jk},\ k\in\mathcal{K}\},\{(V^a_{ij},C^a_{ij},\hat{C}^a_{ij}):i\in\mathcal{I}\},j\in\mathcal{J}\}$ to denote a collection of realized costs and capacity requirements, we can minimize the total of all regular and penalty costs by solving the mixed integer program

$$\Gamma(x,n,U)=\max \quad -\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\,(C^a_{ij}\,z^a_{ij}+\hat{C}^a_{ij}\,[1-z^a_{ij}])-\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}(C^s_{jkl}\,z^s_{jkl}+\hat{C}^s_{jkl}\,[1-z^s_{jkl}]) \tag{1}$$

$$\text{subject to} \quad \sum_{i\in\mathcal{I}:j\in J_i} x_i\,V^a_{ij}\,z^a_{ij}+\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}V^s_{jkl}\,z^s_{jkl}\le\bar{V}_j \quad \text{for all } j\in\mathcal{J} \tag{2}$$

$$0\le z^a_{ij}\le 1 \quad \text{for all } i\in\mathcal{I},\ j\in\mathcal{J}_i \tag{3}$$

$$z^s_{jkl}\in\{0,1\} \quad \text{for all } j\in\mathcal{J},\ k\in\mathcal{K},\ l=1,\dots,n_{jk}. \tag{4}$$

The decision variable z^a_{ij} corresponds to the fraction of the allotment from atomic bid i that we load on flight j and the decision variable z^s_{jkl} takes value one if we load the l-th booked spot market request for flight j and cargo type k. The first set of constraints represent a requirement that the total capacity utilization from all the allotment and sport market bookings cannot exceed the available capacity. The variables and constraints of problem (1)-(4) can be grouped by flight so that decisions associated with each flight do not participate in the constraints corresponding to other flights. Therefore, the problem has a special separable structure: its feasible set is formed as a direct product of the feasible sets for the decisions associated with each flight. Because of the linearity of the objective, the problem decomposes by the flights, but this is simply due to the fact that the loading decisions for different flights are independent. We shortly describe a possible relaxation of this assumption after we elaborate on the booking control problem for the spot market.

Booking control for the spot market. We formulate the spot market booking control problem as a dynamic program over the time periods $\{1,\dots,\tau\}$. In this dynamic programming formulation, we use $n=\{n_{jk}:j\in\mathcal{J},\ k\in\mathcal{K}\}$ as the state variable, where n_{jk} is the total number of accepted spot market bookings for cargo type k on flight j up to the current time period. We use $\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}$ to capture the decisions, where $\mathcal{S}$ corresponds to the set of flight and cargo type combinations that are open for spot market bookings at the current time period. At time t, a new booking for cargo type k on flight j occurs with probability $P_{jkt}(\mathcal{S})$ resulting in a new state $n+e_{jk}$, where we use

e_{jk} to denote the $|\mathcal{J}| \times |\mathcal{K}|$ dimensional unit vector with a one in the element corresponding to $(j,k) \in \mathcal{J} \times \mathcal{K}$. In this case, we can find the optimal booking control policy for the spot market by solving the optimality equation

$$\begin{aligned} J_t(x,n) &= \max_{\mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}} \left\{ \sum_{(j,k) \in \mathcal{S}} P_{jkt}(\mathcal{S}) \big[\mathbb{E}\{R^s_{jk}\} + J_{t+1}(x, n + e_{jk}) \big] + \Big[1 - \sum_{(j,k) \in \mathcal{S}} P_{jkt}(\mathcal{S}) \Big] J_{t+1}(x,n) \right\} \\ &= \max_{\mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}} \left\{ \sum_{(j,k) \in \mathcal{S}} P_{jkt}(\mathcal{S}) \big[\mathbb{E}\{R^s_{jk}\} + J_{t+1}(x, n + e_{jk}) - J_{t+1}(x,n) \big] \right\} + J_{t+1}(x,n). \qquad (5) \end{aligned}$$

In the expression above, the revenue R^s_{jk} associated with a spot market booking is assumed to be random, since the revenue may depend on the realization of the size of the cargo at the departure time. Therefore, we have to charge the expected revenue when making the booking control decision. The boundary condition for the optimality equation above is $J_{\tau+1}(x,n) = \mathbb{E}\{\Gamma(x,n,U)\}$, accounting for the total cost incurred at the departure time. In this case, letting $\bar{0}$ be the $|\mathcal{J}| \times |\mathcal{K}|$ dimensional vector of zeros, $J_1(x,\bar{0})$ corresponds to the optimal total expected profit from the spot market bookings over the whole planning horizon and the loading decisions at the departure time, given that the allotment contracts granted on the terms of the different atomic bids are captured by the vector $x = \{x_i : i \in \mathcal{I}\}$.

Even when the loading problem in (1)-(4) decomposes by the flights under independent departures assumption, the optimality equation in (5) still does not decompose by the flights due to the fact that an arriving spot market customer makes a choice over all flights according to the choice probabilities $\{P_{jkt}(S) : j \in \mathcal{J},\ k \in \mathcal{K}\}$. Furthermore, we emphasize that if there are two flights j and j' such that the cargo for flight j can be also be carried on flight j', then we can extend the model by minimal modifications in problem (1)-(4). In particular, we can introduce decision variables into problem (1)-(4) to capture the portion of the cargo that is shifted from flight j to j'. If the departure time of flights j and j' are close enough in time that the spot market bookings in between the two departure times are negligible, then the dynamic programming formulation in (5) goes through with no modifications and we can still assume that all flight departures occur at the end of the planning horizon. These modifications partially relax independent departures assumption. We elaborate on this extension further in Online Appendix EC.4.

Allotment selection. Noting that $J_1(x, \bar{0})$ captures the total expected profit from the spot market bookings and the costs associated with loading decisions at the departure time, we can choose the allotment contracts by solving the nonlinear integer program

$$\max \quad \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} \mathbb{E}\{R^a_{ij}\}\, x_i + J_1(x, \bar{0}) \tag{6}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}: d \in \mathcal{D}_i} x_i \le 1 \qquad \text{for all } d \in \mathcal{D} \tag{7}$$

$$x_i \in \{0, 1\} \qquad \text{for all } i \in \mathcal{I}. \tag{8}$$

In the problem above, $\sum_{j \in \mathcal{J}_i} R^a_{ij}$ is the total revenue from an allotment contract granted on the terms of the atomic bid i. At the time of making the bid allocation decisions, we charge the expected value of this revenue. The costs associated with serving this allotment are already included into $J_1(x, \bar{0})$.

Problem (6)-(8) looks similar to the traditional value determination problem in combinatorial auctions, especially in the form of its constraints. However, the important difference is that the substance being auctioned, which is the capacity of flights, does not enter the constraints at all. Instead, the capacity utilization enters through the potentially nonlinear term $J_1(x, \bar{0})$ in the objective. In general, computing $J_1(x, \bar{0})$ for a fixed value of x is difficult since the state variable in the optimality equation in (5) is a high dimensional vector. This difficulty, coupled with a discrete nonlinear optimization over x, makes problem (6)-(8) computationally intractable. In the next section, we develop a method to approximate $J_1(x, \bar{0})$, which allows us to find good solutions to problem (6)-(8). An approximation to $J_1(x, \bar{0})$ is also important in itself as it allows us to construct booking control policies for the spot market.

4. Approximating the Booking Problem

In this section, we develop a method to approximate the value functions $\{J_t(x, \cdot) : t = 1, \ldots, \tau\}$ in the spot market booking control problem. We note that a major complicating factor in problem (1)-(4) is the presence of the capacity constraints. This suggests relaxing these constraints by associating the positive Lagrange multipliers $\lambda = \{\lambda_j : j \in \mathcal{J}\}$ with them. In this case, the relaxed version of problem (1)-(4) takes the form

$$\begin{aligned}\tilde{\Gamma}(x,n,\lambda,U) = \max \quad & -\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i \Big\{ C^a_{ij}\, z^a_{ij} + \hat{C}^a_{ij}\,[1 - z^a_{ij}] + \lambda_j\, V^a_{ij}\, z^a_{ij} \Big\} \\ & \qquad - \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}} \Big\{ C^s_{jkl}\, z^s_{jkl} + \hat{C}^s_{jkl}\,[1 - z^s_{jkl}] + \lambda_j\, V^s_{jkl}\, z^s_{jkl} \Big\} + \sum_{j\in\mathcal{J}} \lambda_j\, \bar{V}_j \qquad (9) \\ \text{subject to} \quad & (3),\ (4). \qquad (10)\end{aligned}$$

The shift in our notation from $\Gamma(x,n,U)$ to $\tilde{\Gamma}(x,n,\lambda,U)$ emphasizes that the problem above provides only an approximation to the penalty cost that we incur at the departure time and its optimal objective value depends on the Lagrange multipliers. If the capacity is measured in multiple dimensions, such as weight and volume, then our development goes through as long as λ_j is a multi-dimensional vector and $\lambda_j\, V^a_{ij}$, $\lambda_j\, V^s_{jkl}$ and $\lambda_j\, \bar{V}_j$ are understood as scalar products.

In this case, we can replace the boundary condition of the optimality equation in (5) with the approximate boundary condition $\tilde{J}_{\tau+1}(x,n,\lambda) = \mathbb{E}\{\tilde{\Gamma}(x,n,\lambda,U)\}$ and obtain the value function approximations $\{\tilde{J}_t(x,\cdot,\lambda) : t = 1,\ldots,\tau\}$ by solving the optimality equation

$$\tilde{J}_t(x,n,\lambda) = \max_{\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}} \left\{ \sum_{(j,k)\in\mathcal{S}} P_{jkt}(\mathcal{S}) \big[\mathbb{E}\{R^s_{jk}\} + \tilde{J}_{t+1}(x, n + e_{jk}, \lambda) - \tilde{J}_{t+1}(x,n,\lambda) \big] \right\} + \tilde{J}_{t+1}(x,n,\lambda). \quad (11)$$

Once more, the shift in our notation from $J_t(x,\cdot)$ to $\tilde{J}_t(x.\cdot,\lambda)$ emphasizes that $\tilde{J}_t(x,\cdot,\lambda)$ is only an approximation to $J_t(x,\cdot)$, and the solution to the optimality equation in (11) with the approximate boundary condition $\tilde{J}_{\tau+1}(x,n,\lambda) = \mathbb{E}\{\tilde{\Gamma}(x,n,\lambda,U)\}$ depends on the Lagrange multipliers. We shortly dwell on how we should choose the Lagrange multipliers so that $\{\tilde{J}(x,\cdot,\lambda) : t = 1,\ldots,\tau\}$ are good approximations to the value functions $\{J(x,\cdot) : t = 1,\ldots,\tau\}$.

There are two important properties of the optimality equation in (11). First, this optimality equation can be solved very efficiently. As a matter of fact, Proposition 1 below shows that there is a closed form solution to this optimality equation. To motivate this result, we begin by defining some notation. We let

$$B^a_{ij}(\lambda_j) = \mathbb{E}\{\min(C^a_{ij} + \lambda_j\, V^a_{ij}, \hat{C}^a_{ij})\}$$

for all $i \in \mathcal{I}$ and $j \in \mathcal{J}_i$. Intuitively speaking, the term $B^a_{ij}(\lambda_j)$ captures *the expected total cost* of serving the atomic bid i on flight leg j, when this cost is viewed from the beginning of the planning horizon. In particular, the atomic bid i utilizes V^a_{ij} units of capacity on flight j. Noting that the Lagrange multiplier λ_j measures the opportunity cost of the capacity on flight j, the total opportunity cost of the capacity consumed by the atomic bid i on flight j is $\lambda_j V^a_{ij}$. By adding a direct regular cost C^a_{ij} we get the total cost of shipping as $C^a_{ij} + \lambda_j V^a_{ij}$. However, we also have the option of not loading this cargo on the flight and the cost associated with this option is $\hat{C}^a_{ij}$. It is sensible to follow the option with the smallest cost, in which case, the cost of the atomic bid i on flight j is $\min(C^a_{ij} + \lambda_j V^a_{ij}, \hat{C}^a_{ij})$. Since we do not know the capacity utilization of the atomic bid at the beginning of the planning horizon, we take an expectation to capture the expected cost of serving the atomic bid i on flight j, when this cost is viewed from the beginning of the planning horizon. Similar to $B^a_{ij}(\lambda_j)$, we also define

$$B^s_{jk}(\lambda_j) = \mathbb{E}\{\min(C^s_{jk} + \lambda_j V^s_{jk}, \hat{C}^s_{jk})\}$$

for all $j \in \mathcal{J}$ and $k \in \mathcal{K}$. A similar intuitive reasoning indicates that the term $B^s_{jk}(\lambda_j)$ captures the expected total cost of serving a spot market booking request of cargo type k on flight j. We note that both $B^a_{ij}(\lambda_j)$ and $B^s_{jk}(\lambda_j)$ are straightforward functions of the Lagrange multipliers. We are now ready to show that there is a closed form solution to the optimality equation in (11). All of our proofs are deferred to Online Appendix EC.1.

Proposition 1. *Letting* $\{\tilde{J}_t(x, \cdot, \lambda) : t = 1, \ldots, \tau\}$ *be the solution to the optimality equation in* (11)*, we have the identity*

$$\tilde{J}_t(x, n, \lambda) = \sum_{j \in \mathcal{J}} \lambda_j \bar{V}_j - \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} x_i B^a_{ij}(\lambda_j) - \sum_{j \in \mathcal{J}} \sum_{k \in \mathcal{K}} B^s_{jk}(\lambda_j)\, n_{jk} + \sum_{t'=t}^{\tau} \Phi_{t'}(\lambda), \tag{12}$$

where

$$\Phi_t(\lambda) = \max_{\mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}} \left\{ \sum_{(j,k) \in \mathcal{S}} P_{jkt}(\mathcal{S}) \left[\mathbb{E}\{R^s_{jk}\} - B^s_{jk}(\lambda_j) \right] \right\}. \tag{13}$$

The term $\tilde{J}_t(x,n,\lambda)$ on the left side of (12) is an approximation to the optimal total expected profit from the spot market bookings over the time periods $\{t,\ldots,\tau\}$ and the loading decisions at the departure time. The right side of (12) separates this expected profit into four components with interesting insights. The first component captures the total value of the capacity available on all of the flights. Noting the intuitive interpretation of $B^a_{ij}(\lambda_j)$ as expected total costs associated with serving a flight j part of the atomic bid i, the second component corresponds to the expected total costs associated with serving the accepted bids on all of the flight legs. Similarly, the third component represents the expected total costs associated with serving all of the accepted spot market bookings. Finally, the last component estimates the maximum total expected future profit from the spot market. In particular, noting the intuitive interpretation for $B^s_{jk}(\lambda_j)$ as expected total cost, the difference $\mathbb{E}\{R^s_{jk}\} - B^s_{jk}(\lambda_j)$ is the expected profit from a spot market booking for cargo type k on flight j. The weighted average represented by the summation over $(j,k) \in \mathcal{S}$ in expression (13) for $\Phi_t(\lambda)$ corresponds to the expected value over all possible spot market bookings at time t. Therefore, the optimization problem in (13) finds a set of flight and cargo type combinations to open at time period t so as to maximize the expected profit $\Phi_t(\lambda)$ at time t. In Online Appendix EC.5, we discuss important choice models that make the solution of this optimization problem particularly easy. An interesting aspect of our solution approach is that the computation of $\tilde{J}_t(x,n,\lambda)$ does not get much more difficult when capacity is measured in multiple dimensions, such as weight or volume. Furthermore, our Lagrangian relaxation approach gives a natural way to assess the value of a unit of capacity, irrespective of how many different units are used to measure capacity.

The second important property of the optimality equation in (11) is that it provides an upper bound on the optimal expected spot market profits. In particular, the next proposition shows that $\tilde{J}_1(x,\bar{0},\lambda)$ is an upper bound on the $J_1(x,\bar{0})$ as long as the Lagrange multipliers are nonnegative.

Proposition 2. *If $\lambda \geq 0$, then we have $\tilde{J}_1(x,\bar{0},\lambda) \geq J_1(x,\bar{0})$.*

The implication of this result is that we can obtain the tightest possible upper bound on the optimal total expected profit from the spot market bookings and the loading decisions by solving the problem

$$\min_{\lambda \geq 0} \; \tilde{J}_1(x, \bar{0}, \lambda). \tag{14}$$

Letting $\lambda^*(x, \bar{0})$ be an optimal solution to problem (14), we can use $\tilde{J}_1(x, \bar{0}, \lambda^*(x, \bar{0}))$ as an approximation to $\tilde{J}_1(x, \bar{0})$. It is important to emphasize that problem (14) provides a concrete method for choosing the Lagrange multipliers.

Closing this section, we note that $B^a_{ij}(\lambda_j)$ and $B^s_{jk}(\lambda_j)$ are concave functions of λ_j as minimum of a linear function and a constant is concave. Noting that the maximum of convex functions is also convex, (13) implies that $\Phi_t(\lambda)$ and, thus, $\tilde{J}_t(x, \bar{0}, \lambda)$ are convex functions of the Lagrange multipliers. Therefore, problem (14) is a convex optimization problem and it can be solved by using standard convex optimization tools. This observation becomes useful in the next section.

5. Approximating the Allotment Problem

Our development in the previous section suggests approximating $J_1(x, \bar{0})$ by $\min_{\lambda \geq 0} \tilde{J}_1(x, \bar{0}, \lambda)$. In this case, we can replace $J_1(x, \bar{0})$ in problem (6)-(8) with $\min_{\lambda \geq 0} \tilde{J}_1(x, \bar{0}, \lambda)$ and choose the allotment contracts by solving the problem

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} \mathbb{E}\{R^a_{ij}\}\, x_i + \min_{\lambda \geq 0} \tilde{J}_1(x, \bar{0}, \lambda) && (15) \\ \text{subject to} \quad & (7)\text{-}(8). && (16) \end{aligned}$$

Our approach for solving the problem above is based on the idea of Benders decomposition or constraint generation. In a nutshell, we reformulate problem (15)-(16) as a mixed integer programming problem with an infinite number of constraints and devise a method to sample a finite subset of constraints resulting in an optimal solution to the original problem. To this end, we replace $\min_{\lambda \geq 0} \tilde{J}_1(x, \bar{0}, \lambda)$ in the problem above with a single decision variable y and impose the constraint that $y \leq \tilde{J}_1(x, \bar{0}, \lambda)$ for all $\lambda \geq 0$ on this decision variable. Using the closed form expression for $\tilde{J}_1(x, \bar{0}, \lambda)$ given in Proposition 1, this idea yields the problem

$$\max \quad \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} \mathbb{E}\{R^a_{ij}\}\, x_i + y \tag{17}$$

$$\text{subject to} \quad (7)\text{-}(8) \tag{18}$$

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} B^a_{ij}(\lambda_j)\, x_i + y \leq \sum_{j \in \mathcal{J}} \lambda_j \bar{V}_j + \sum_{t=1}^{\tau} \Phi_t(\lambda) \quad \text{for all } \lambda \geq 0. \tag{19}$$

Problem (17)-(19) is equivalent to problem (15)-(16). Furthermore, the objective function and constraints of this problem are linear in all of the decision variables $x = \{x_i : i \in \mathcal{I}\}$ and y. However, problem (17)-(19) has an infinite number of constraints and cannot be solved directly.

Throughout this section, we refer to problem (17)-(19) as problem (P). We consider a relaxed version of problem (P), where the infinite number of constraints (19) are replaced by the finite number of constraints

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} B^a_{ij}(\lambda_j)\, x_i + y \leq \sum_{j \in \mathcal{J}} \lambda_j \bar{V}_j + \sum_{t=1}^{\tau} \Phi_t(\lambda) \quad \text{for all } \lambda \in \{\lambda^1, \ldots, \lambda^S\}, \tag{20}$$

where $\{\lambda^1, \ldots, \lambda^S\}$ is a finite set of Lagrange multipliers. We refer to this relaxed version of problem (P) as (R-$\{\lambda^1, \ldots, \lambda^S\}$). In this case, we can solve problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) for a finite set of Lagrange multipliers to obtain the optimal solution $(\hat{x}, \hat{y})$. Problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) is a mixed integer program with a finite number of constraints. Following this, we can solve the problem $\min_{\lambda \geq 0} \tilde{J}_1(\hat{x}, \bar{0}, \lambda)$ to obtain an optimal solution $\lambda^*(\hat{x}, \bar{0})$. If the optimal objective value of the last problem matches $\hat{y}$, then $(\hat{x}, \hat{y})$ provide an optimal solution to the full problem (P) and we stop. Otherwise, we add $\lambda^*(\hat{x}, \bar{0})$ to the set of Lagrange multipliers $\{\lambda^1, \ldots, \lambda^S\}$ and solve problem (R-$\{\lambda^1, \ldots, \lambda^S, \lambda^*(\hat{x}, \bar{0})\}$). This idea yields the following algorithm for solving problem (P).

1. Start with some initial set of Lagrange multipliers $\{\lambda^1, \ldots, \lambda^S\}$ of size S. The set can be empty, in which case $S = 0$.
2. Solve problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) to within the accuracy $\frac{\epsilon}{3}$ of the optimum. Let the resulting feasible solution be (x^S, y^S).
3. Solve the problem $\min_{\lambda \geq 0} \tilde{J}_1(x^S, \bar{0}, \lambda)$ to within the accuracy $\frac{\epsilon}{3}$ of the optimum. Let the resulting feasible solution be λ^{S+1}.
4. If $\tilde{J}_1(x^S, \bar{0}, \lambda^{S+1}) \geq y^S - \frac{\epsilon}{3}$, then stop. Otherwise, increment S by 1 and go back to Step 2.

We have the following result for the proposed algorithm.

PROPOSITION 3. *The proposed algorithm terminates in a finite number of steps. At termination,* $(x^S, y^S - 2\epsilon/3)$ *is an* ϵ*-optimal solution to problem* (P).

This result implies that for every fixed ϵ, we can find ϵ-optimal solution to problem (P) in a finite number of steps even though optimization algorithms used in practice for mixed integer programming and convex minimization are approximate.

6. Practical Implementation of Spot Market Booking Control

Problem (15)-(16) provides a method to choose the allotment contracts. However, we also need to make control decisions for the booking requests on the spot market. A practical implementation of the booking control policy for the spot market has to be computationally efficient as well as sufficiently robust to perform well under a variety of types of consumer behavior and demand patterns. In this section, we focus on setting up a booking control policy for the spot market.

Letting x^* be an optimal solution to problem (15)-(16), we choose the allotment contracts as indicated by the solution x^*. In this case, we can solve the problem $\min_{\lambda \geq 0} \tilde{J}_1(x^*, \bar{0}, \lambda)$ to choose a good set of Lagrange multipliers $\lambda^{**} = \lambda^*(x^*, \bar{0})$. This allows us to use $\{\tilde{J}_t(x^*, \cdot, \lambda^{**}) : t = 1, \dots, \tau\}$ as approximations to the value functions $\{J_t(x^*, \cdot) : t = 1, \dots, \tau\}$. In other words, we can replace $\{J_t(x^*, \cdot) : t = 1, \dots, \tau\}$ in problem (5) with $\{\tilde{J}_t(x^*, \cdot, \lambda^{**}) : t = 1, \dots, \tau\}$ and solve this problem to make the booking control decisions for the spot market at different time periods. Proposition 1 shows that $\tilde{J}_t(x, n, \lambda^{**})$ is a linear function of n and the difference $\tilde{J}_{t+1}(x^*, n + e_{kt}, \lambda^{**}) - \tilde{J}_{t+1}(x^*, n, \lambda^{**})$ in the right side of (5) is equal to $-B^s_{jk}(\lambda^{**}_j)$. Therefore, we can solve the problem

$$\max_{\mathcal{S} \subseteq \mathcal{J} \times \mathcal{K}} \left\{ \sum_{(j,k) \in \mathcal{S}} P_{jkt}(\mathcal{S}) \left[\mathbb{E}\{R^s_{jk}\} - B^s_{jk}(\lambda^{**}_j) \right] \right\} + \tilde{J}_{t+1}(x^*, n, \lambda^{**}) \tag{21}$$

to make the booking control decisions for the spot market at time period t. This is to say that if $\mathcal{S}^*$ is the optimal solution to the optimization problem above, then set of flight and cargo type combinations that the airline opens for sale at time period t is given by $\mathcal{S}^*$. We note that the term $\tilde{J}_{t+1}(x^*, n, \lambda^{**})$ in (21) does not affect the booking control decisions at time period t, and the first term is equal to $\Phi_t(\lambda^{**})$ defined by (13).

The chief drawback of the booking control policy obtained from problem (21) is that the optimal solution to this problem does not depend on the numbers of accepted spot market bookings. In particular, irrespective of whether we reach time period t with too many accepted spot market bookings or with too few, we always open the same set of flight and cargo type combinations for sale. One way to remedy this shortcoming is to refresh the Lagrange multipliers periodically over the planning horizon. In other words, if the numbers of accepted spot market bookings by time period t are given by n, then we can solve the problem $\min_{\lambda \geq 0} \tilde{J}_t(x^*, n, \lambda)$ to obtain a new set of Lagrange multipliers at the current time period and use these Lagrange multipliers in problem (21) to make the spot market booking control decisions. In this way, since the optimal solution to the problem $\min_{\lambda \geq 0} \tilde{J}_t(x^*, n, \lambda)$ depends on the numbers of accepted booking requests n, the booking control decisions from problem (21) also depend on the numbers of accepted bookings. An interesting question is how problem (21) changes when we reach the same time period with different numbers of accepted spot market bookings. The next proposition attempts to give an answer to this question.

PROPOSITION 4. *Fix cargo allotment decisions x, flight j and cargo type k. Let $\lambda^*(x,n)$ be an optimal solution to the problem $\min_{\lambda \geq 0} \tilde{J}_t(x, n, \lambda)$. If there are multiple optimal solutions to this problem, then choose $\lambda^*(x,n)$ as the one that yields the largest value for $B^s_{jk}(\lambda^*_j(x,n))$. In this case, we have $B^s_{jk}(\lambda^*_j(x, n + e_{jk})) \geq B^s_{jk}(\lambda^*_j(x,n))$.*

This result indicates that if we reach time period t with more accepted spot market bookings for flight j and cargo type k and we solve the problem $\min_{\lambda \geq 0} \tilde{J}_t(x^*, n, \lambda)$ to choose a set of Lagrange multipliers at this time period, then the adjustment $B^s_{jk}(\lambda_j)$ to the revenue R^s_{jk} in problem (21) becomes larger. This makes flight and cargo type pair (j,k) less attractive to open for sale. Therefore, if we reach time period t with larger number of accepted spot market bookings for flight j and cargo type k, then we are less willing to open this flight and cargo type combination for sale.

A remaining issue of efficient calculation of the optimal set $\mathcal{S}$ in equation (21) or its dynamic variant is addressed in Online Appendix EC.5.

7. Computational Experiments

In this section, we describe numerical experiments with the model. We lay out the setup of these experiments and provide practical motivation for the settings in §7.1. The simulated performance of the proposed allotment/booking control policy is discussed in §7.2. Under the proposed control policy, the average simulated profit is at least within 13% of the theoretical upper bound in all experimental configurations and is within 6% of the bound in the best case. We also find that profit and cost structures exhibit intuitive dependence on experimental parameters. For example, all indicators suggest that the airline creates a higher capacity buffer for the spot market when uncertainty in spot market cargo volume is high. Moreover, this buffer reduces the contribution of penalties to the total costs. In §7.3, we generate additional insight into priorities of the decision maker by analyzing allotment and spot market policies suggested by the model. Among other findings, we see how the optimization process handles additional spot market risk by accepting a larger number of smaller spot market bookings when uncertainty in their volume is high and by increasing the shipping rates charged on oversized cargo. We conclude this section with a brief discussion of time performance and scalability to industry-sized instances in §7.4.

7.1. Setup

The experimental setup consists of the following elements: 1) the planning horizon, the flight schedule and capacities of flights; 2) classification of spot market bookings into cargo types and their characteristics (volumes, weights, revenues, penalties, and costs); 3) allotment bids and their characteristics; 4) the arrival process and the consumer choice model on the spot market; and, finally, 5) options for allotment selection and spot market optimization procedures. In the main body of the paper, we highlight only the most important elements of this setup, with more technical details relegated to Online Appendix EC.2.

The planning horizon approximately corresponds to a typical half-a-year allotment contract. We assume that *the flight schedule* follows a weekly pattern, and consider jointly 104 flight departures arranged into 26 schedule cycles with 4 flights each (equally spaced during the cycle). Our earlier development in the paper assumed that the capacity of flight j is measured in terms of a single quantity $\bar{V}_j$, but we also pointed out how to extend our model to the case where the capacity is measured in terms of multiple units, such as volume and weight. In our computational experiments, we indeed assume that the capacity of a flight is measured in terms of both volume and weight. The volume $\bar{V}$ and weight $\bar{W}$ *capacities* of each flight are the same, but we consider 4 different combinations of capacity settings. The maximum total cargo volume $\bar{V}$ can be 75m^3 or 150m^3, and the maximum structural payload (weight) $\bar{W}$ can be 10 ton or 20 ton. These values are in the range of the commonly used air cargo airplanes (for example, Boeing 727-100C aircraft has the maximum cargo volume of 118m^3 and the maximum structural payload from 14.7 to 18.6 ton depending on its operational mode). To make the scale of capacity dimensions easier to compare, we convert all volume values into *dimensional weight* by dividing them with the factor of 6000cm^3/kg=6m^3/metric ton (the value commonly used in industry). This results in the volume capacity levels expressed as 12.5 and 25 tons of dimensional weight.

Cargo types and their characteristics. The variety of cargo encountered in practice is quite large, and each type of cargo can, in general, have its own weight and volume distribution, shipping rate, as well as cost/penalty characteristics. To represent this variety in our experiments while keeping the number of different settings under control, we randomly generate 1000 cargo types according to the rules described in Online Appendix EC.2. The generation procedure is repeated 8 times and each set of classes is processed under each combination of the controlled experimental settings. The form of the distribution of cargo weight and volume is the same as in numerical tests of Amaruchkul et al. (2007b), and reflects an observation of Slager and Kapteijns (2003) that volume measurement and registration are problematic, so there is significantly more uncertainty in volume. In particular, since it is easy for shippers to provide accurate weight information at the time of booking, we assume that the weight for each cargo type is known with certainty. On the other hand, the volume of each cargo type is uncertain and follows a lognormal distribution. We test two levels of the coefficient of variation of the volume distribution $\theta^s = 0.2$ (low uncertainty) and $\theta^s = 0.8$ (high uncertainty). The rates of different types are randomly sampled from the interval $[1500, 3000]$ (in

\$/ton). Corresponding to industry practice, the actual amount charged to the customer is the rate times the maximum of the shipment weight and its dimensional weight. The magnitude of shipping costs and penalties for spot market cargo is proportional to its shipping rate.

Allotment bids and their characteristics. For each set of randomly chosen spot market cargo types we randomly generate 20 allotment bids. Since we consider a particular origin-destination pair, such number of bidders is reasonable. The bids generate the same capacity utilization pattern for every schedule cycle and are equally likely 1) to require one flight *or* two flights in a cycle conditional on each other, and 2) to be inflexible (only one flight/pair of flights in a cycle is acceptable) *or* inflexible (two different flight/pairs are acceptable). The requested rate and size of a bid are the same for all flights indicated in the bid. The rates are sampled from the interval $[1000, 2000]$ (in \$/ton) and the sizes (the maximum weight) from the normal distribution with mean 5 tons and standard deviation 2 tons. The allotment utilization (volume and weight) distribution is selected to be the same for all bids so that the weight utilization reflects a practical rule of "80% utilization 80% of the time" while the volume distribution is lognormal conditional on the weight. We specifically examine two levels of the coefficient of variation of the allotment volume distribution: $\theta^a = 0.2$ (low uncertainty) and $\theta^a = 0.4$ (high uncertainty). Thus, there are a total of four combinations in the level of uncertainty in volume: low/high uncertainty for the spot market combined with low/high uncertainty for the allotments. These four combinations together with four combinations for flight capacity result in a total of 16 experimental scenarios. Each of these scenarios is tested on 8 different instances of allotment bids and cargo types.

The remaining elements are *the arrival process and the consumer choice model.* The time between consecutive flights is partitioned into 60 periods. There is a 0.5 probability of a spot market request arrival in each period with each request equally likely to be any of the 1000 cargo types. Arriving customers consider flights up to 2 cycles in advance (a total of 8 flights corresponding to a two-week booking horizon common for cargo) as well as a no-purchase option. The choice probabilities for each cargo type follow a multinomial logit model with randomly selected parameters.

Finally, among *the options for optimization procedures*, we highlight the following practical techniques to reduce dimensions of λ vectors used in upper bound calculations and associated computational overhead:

- For allotment selection, the bound $\tilde{J}_1(x, \bar{0}, \lambda)$ is optimized under the restriction that components of λ are the same for the corresponding flights in different schedule cycles. This results in an 8-dimensional convex minimization problem.

- For spot market booking control, the bound $\tilde{J}_t(x, \cdot, \lambda)$ is reoptimized (as suggested in §6), resulting in new λ, twice between consecutive flights (that is, every 30 decision periods – eight times per schedule cycle and 208 times during the entire planning horizon). The components of λ are allowed to vary individually for flights within the first two schedule cycles around t but are constrained to be the same for all corresponding flights of the subsequent schedule cycles. This results in a 24-dimensional convex minimization problem.

The problem $\min_{\lambda \geq 0} \tilde{J}_t(x, n, \lambda)$, in all cases, is solved by a bundle method for nondifferentiable convex minimization.

7.2. Performance and structure of profit, cost and penalty

We start this subsection by discussing simulated performance of the allotment selection method and spot market booking control compared to the theoretical upper bound for the optimal expected profit. The performance is measured as the average percentage ratio of simulated profit of the 104 flights to the upper bound on the profit value (the average is over different realizations of allotment bids and cargo types). Table 1 shows the performance for each of the 16 experimental settings. We observe that the simulated profit is in the range $[87.1\%, 94.8\%]$ of the upper bound, and there are some intuitive patterns. For example, the relative performance deteriorates as the level of uncertainty increases in three out of 4 capacity configurations. This is particularly clear in the case of low volume capacity of 12.5. Uncertainty in spot market cargo volume compounds with uncertainty in allotment volume. It is interesting to note that the level of volume uncertainty has the highest impact when volume capacity is low but weight capacity is high. On the other

		$\theta^s = 0.2$		$\theta^s = 0.8$	
$\bar{V}$	$\bar{W}$	$\theta^a = 0.2$	$\theta^a = 0.4$	$\theta^a = 0.2$	$\theta^a = 0.4$
12.5	10.0	92.5%	90.0%	89.8%	88.9%
12.5	20.0	89.9%	87.2%	89.2%	87.1%
25.0	10.0	93.6%	94.9%	93.5%	93.9%
25.0	20.0	94.8%	93.7%	94.7%	93.4%

Table 1 Performance as the average simulated profit relative to the theoretical upper bound

hand, the level of uncertainty has almost no effect on relative performance in the opposite situation of high volume and low weight capacity. This is natural because, in the first case, the uncertain capacity dimension (volume) is the most restrictive whereas, in the second case, the certain capacity dimension (weight) is the most restrictive. Among all scenarios, the relative performance is the worst when volume capacity is low but weight capacity is high.

Since the profit structure highlights the priorities of the decision maker, we examine the percentage of simulated profits obtained from allotments in Table 2. The percentage of profit from allotments significantly decreases when the level of uncertainty in spot market cargo volume increases. There can be several explanations for this phenomenon but our subsequent analysis suggests that the airline is taking fewer allotments and creating a larger capacity buffer for spot market bookings. Moreover, the airline is taking advantage of the profit potential from the spot market bookings with a high dimensional weight. (In this example, the revenue rate of the spot market cargo is higher than that of allotments. Moreover, higher volume uncertainty, on average, results in a higher dimensional weight, since the latter's downside variability is restricted by known cargo weight.) The percentage of profit from allotments increases when the level of uncertainty in allotment volume increases. A possible explanation is that the airline is then forced to leave more space for the allotments. This effect is confirmed when we look at the number of accepted allotments and spot market bookings in the next subsection. Finally, we observe an increase in the percentage of profit from allotments when weight capacity increases. This is especially clear in situations when the level of uncertainty in spot market cargo volume is high and the level of uncertainty in allotment volume is low.

More light can be shed on the role of uncertainty by considering a percentage of penalty costs in

		$\theta^s = 0.2$		$\theta^s = 0.8$	
$\bar{V}$	$\bar{W}$	$\theta^a = 0.2$	$\theta^a = 0.4$	$\theta^a = 0.2$	$\theta^a = 0.4$
12.5	10.0	83.9%	87.6%	58.3%	67.1%
12.5	20.0	87.9%	91.3%	67.9%	72.8%
25.0	10.0	84.0%	88.1%	58.5%	67.8%
25.0	20.0	84.9%	88.8%	69.5%	73.3%

Table 2 Percentage of profit derived from allotments

		$\theta^s = 0.2$		$\theta^s = 0.8$	
$\bar{V}$	$\bar{W}$	$\theta^a = 0.2$	$\theta^a = 0.4$	$\theta^a = 0.2$	$\theta^a = 0.4$
12.5	10.0	13.0%	19.9%	14.2%	17.3%
12.5	20.0	17.5%	25.3%	15.8%	19.9%
25.0	10.0	11.5%	12.3%	9.6%	11.0%
25.0	20.0	7.3%	9.6%	7.6%	9.3%

Table 3 Percentage of penalties in the total cost

the total cost given in Table 3. Overall, the relative contribution of penalties into total costs is the lowest when both volume and weight capacity is high, and the highest when weight capacity is high but volume capacity is low. This points to uncertainty in volume as the main driver of penalties. Interestingly, an increase in the level of spot market volume uncertainty decreases the relative contribution of penalties. This agrees with our earlier observation that the airline is likely to leave more space for spot market bookings by creating a larger capacity buffer when spot market volume uncertainty is high. On the other hand, a higher level of uncertainty in allotment volume increases the relative contribution of penalties, especially when volume capacity is tight. This suggests that a good strategy for reduction of penalties is to decrease the level of uncertainty in allotment volume.

7.3. Decision structure and capacity utilization

In this subsection, we examine the structure of the airline decisions for the allotments and the spot market. Table 4 shows the average characteristics of accepted allotments for each experimental setting in terms of volume and weight capacity and the level of uncertainty in spot market cargo and allotment volume. The percentage of accepted allotment bids is higher when weight capacity increases. The volume capacity increase has no effect on allotment decisions if weight capacity remains low but its effects become significant at the high level of weight capacity. Generally, the

$\bar{V}$	$\bar{W}$	θ^s	θ^a	%Accepted	%Single-flight	%Flexible	Δrate	Δsize
12.5	10.0	0.2	0.2	30%	46%	53%	293	0.031
12.5	10.0	0.2	0.4	32%	49%	56%	283	-0.045
12.5	10.0	0.8	0.2	22%	48%	52%	345	-0.070
12.5	10.0	0.8	0.4	25%	48%	55%	324	-0.004
12.5	20.0	0.2	0.2	35%	48%	47%	280	0.243
12.5	20.0	0.2	0.4	39%	47%	49%	271	-0.037
12.5	20.0	0.8	0.2	28%	40%	41%	320	0.229
12.5	20.0	0.8	0.4	32%	43%	46%	308	-0.010
25.0	10.0	0.2	0.2	30%	46%	53%	293	0.031
25.0	10.0	0.2	0.4	32%	49%	56%	283	-0.045
25.0	10.0	0.8	0.2	22%	48%	52%	345	-0.070
25.0	10.0	0.8	0.4	25%	48%	55%	324	-0.004
25.0	20.0	0.2	0.2	55%	47%	52%	226	0.029
25.0	20.0	0.2	0.4	59%	47%	51%	203	-0.035
25.0	20.0	0.8	0.2	48%	45%	48%	262	-0.001
25.0	20.0	0.8	0.4	51%	47%	52%	248	-0.091

Table 4 Average characteristics of accepted allotments for each scenario

percentage of accepted allotments is somewhat higher when the level of uncertainty in allotment volume is high. This is accompanied by a slightly stronger preference for flexible bids. In contrast, a higher level of spot market volume uncertainty reduces the fraction of accepted bids. This indicates that more capacity is allocated to spot market bookings. The last two columns of the table show the average difference (Δrate) in the rate of accepted allotment bids with the average bid rate (per ton), and the average difference (Δsize) of the size of accepted allotment bids with the average bid size. When the airline accepts a smaller number of bids, the difference in the accepted rate with the average rate increases (scenarios with low uncertainty in allotment volume or high uncertainty in spot market volume). The sizes of accepted allotments remain close to the average except for one scenario (low volume, high weight capacities, and low levels of uncertainty in volume of allotment and spot market cargo).

Finally, we examine the preferences of the airline with respect to spot market cargo. Table 5 shows the average characteristics of the accepted spot market cargo for each scenario: The number accepted in each capacity/uncertainty scenario averaged over 8 data instances, the average volume and weight of bookings, as well as the linear regression model parameters expressing a relation of the rate of accepted bookings to their volume and weight. The volume in all cases is expressed as

				Number	Average		Rate	Sensitivity to	
$\bar{V}$	$\bar{W}$	θ^s	θ^a	Accepted	Volume	Weight	Intercept	Volume	Weight
12.5	10.0	0.2	0.2	135	0.409	0.296	2550	-524	750
12.5	10.0	0.2	0.4	99	0.447	0.300	2541	-504	807
12.5	10.0	0.8	0.2	334	0.346	0.296	2548	-505	615
12.5	10.0	0.8	0.4	266	0.362	0.298	2573	-495	625
12.5	20.0	0.2	0.2	142	0.306	0.296	2611	138	-133
12.5	20.0	0.2	0.4	105	0.304	0.293	2626	136	-133
12.5	20.0	0.8	0.2	321	0.291	0.296	2572	402	-405
12.5	20.0	0.8	0.4	274	0.293	0.301	2608	416	-417
25.0	10.0	0.2	0.2	137	0.408	0.297	2551	-522	738
25.0	10.0	0.2	0.4	100	0.450	0.301	2537	-520	819
25.0	10.0	0.8	0.2	329	0.348	0.298	2552	-482	580
25.0	10.0	0.8	0.4	267	0.362	0.298	2570	-523	665
25.0	20.0	0.2	0.2	284	0.353	0.297	2495	-446	565
25.0	20.0	0.2	0.4	215	0.372	0.294	2517	-459	623
25.0	20.0	0.8	0.2	488	0.330	0.297	2403	-297	358
25.0	20.0	0.8	0.4	437	0.333	0.297	2436	-407	476

Table 5 Average characteristics of accepted spot market cargo for each scenario

dimensional weight and the unit of measurement is one ton. The table shows that the number of accepted bookings increases when the level of uncertainty in spot market cargo volume increases or when the level of uncertainty in allotment volume decreases. Simultaneously with this increase in number, the average volume of bookings goes down. This suggests that the proposed booking control approach implements an implicit risk pooling by taking a larger number of, on average, smaller volume bookings. The average weight (a deterministic capacity dimension) remains almost the same throughout all scenarios. The regression model parameters describing the rate of accepted bookings as a linear function of their volume and weight are interesting. The rate intercepts are the highest when volume capacity is low but weight capacity is high (a scenario difficult for optimization), but they are the lowest when both kinds of capacity are plentiful. According to the rate sensitivity parameters, in most settings, the proportional increase in the volume and weight of a package increases its shipping rate. This indicates that the airline prefers to charge more for oversized cargo since it entails more uncertainty. The exception is the case of low volume capacity combined with high weight capacity. In this case, the airline prefers to reduce the volume of accepted bookings.

7.4. Time performance and scalability to industry-sized instances

The time performance of the proposed solution method in these experiments indicates that it is scalable to industry-sized instances. All simulations in the experiments are implemented in C++ on an open-source system (Linux), and both the initial allotment selection problem and the terminal off-loading problems are solved with a general open-source mixed-integer programming solver CBC of the COIN-OR project. The bundle method used for solving $\min_\lambda \tilde{J}_t(x,n,\lambda)$ utilizes a linear/quadratic programming solver CLP of COIN-OR. The jobs are executed on a SHARCNET serial throughput cluster, and a typical simulation run takes, on average, about 7.5 hours with allotment optimization taking approximately 3 minutes. We point out that a relatively long 7.5 hour total execution time represents simulation of a process that takes 26 weeks to complete in reality. The main source of complexity is the reoptimization of λ and, within it, the repeated calculation of $\tilde{J}_t(x,\cdot,\lambda)$. However, because of the special structure of $\tilde{J}_t(x,\cdot,\lambda)$, this task can be efficiently parallelized with existing technology. In practice, because of the short booking horizon for cargo, and a limited number of flights for a particular origin-destination pair, the dimensions of λ would not exceed an order of 100 (if its components are restricted as suggested in our experiments). With a high-quality quadratic programming solver, the bundle method for optimizing $\tilde{J}_t(x,\cdot,\lambda)$ is quite efficient for problems of this dimension. (In our experiments, a typical reoptimization of λ takes approximately 10-30 seconds at the allotment selection stage, and 1-3 minutes at the booking control stage.) The remaining concern is a solution of the allotment selection problem. However, this problem needs to be solved only twice a year, and the proposed algorithm for the allotment selection problem employs an integer programming problem similar to the ones arising in combinatorial auctions. There exist efficient parallel methods for mixed-integer programming problems, as well as considerable practical experience in computing the outcomes of combinatorial auctions of a very large size. Therefore, scalability challenges posed by the proposed solution method are purely technical in nature and can be resolved in practice.

8. Conclusions

In this article, we propose a method to coordinate allotment and spot market cargo capacity allocations for collections of parallel flights. Allotment contracts are chosen from multiple bids specifying the combinations of flights. The form of the contract and associated capacity utilization patterns considered in the model are general. Moreover, the combinations included in bids can be related to each other by means of a general OR* bidding language developed in the combinatorial auctions literature.

The coordination method employs a direct estimate of the expected profit from the spot market, which is a challenging task in the case of cargo capacity management. This estimate is useful even in the absence of the accurate margin data for the allotments. The proposed numerical procedure is efficient, performs well in the numerical experiments, and results in allotment assignments that have intuitive properties. Moreover, the approach also suggests a bid-price based booking policy which compares well against the theoretical upper bound in numerical experiments. Among other findings, numerical experiments indicate that the proposed model accepts more of a lower size spot market cargo bookings when the level of uncertainty in the spot market cargo volume increases. In effect, this indicates that risk pooling among capacity requirements of different cargo is captured by the proposed capacity allocation method.

In future work, it would be interesting to consider the problem of allotment and spot market coordination across a network of flights. Moreover, the proposed general method facilitates a comparative study of specific forms of allotment contracts in detailed operational settings.

Acknowledgement

We acknowledge the comments of the associate editor and anonymous referees that substantially improved our model. This research was supported, in part, by Natural Sciences and Engineering Research Council of Canada (grant numbers 261512-04 and 341412-07) and Natural Science Foundation of USA (grant number CMMI 0825004). We would also like to thank Adam Dudar of Air Canada for his valuable input regarding current air cargo industry practices, as well as John

Forrest of IBM and Stefan Vigerske of Humboldt University, Berlin for their helpful suggestions on the use of the COIN-OR software.

References

Adelman, D. 2007. Dynamic bid-prices in revenue management. *Operations Research* **55**(4) 647–661.

Amaruchkul, K., W. Cooper, D. Gupta. 2007a. Air-cargo capacity contracts under asymmetric information. *Working paper, University of Minnesota* .

Amaruchkul, K., W. Cooper, D. Gupta. 2007b. Single-leg air-cargo revenue management. *Transportation Science* **41**(4) 457–469.

Anderson, S., A. de Palma, J. Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.

Becker, B., N. Dill. 2007. Managing the complexity of air cargo revenue management. *Journal of Revenue and Pricing Management* **6**(3) 175–187.

Bertsimas, D., I. Popescu. 2003. Revenue management in a dynamic network environment. *Transportation Science* **37** 257–277.

Billings, J., A. Diener, B. Yuen. 2003. Cargo revenue optimization. *Journal of Revenue and Pricing Management* **2**(1) 69–79.

Blomeyer, J. 2006. Air cargo revenue management. Ph.D. thesis, Tilburg University, Netherlands.

Cachon, G. 2004. The allocation of inventory risk in a supply chain: Push, pull, and advance-purchase discount contracts. *Man. Sci.* **50**(2) 222–238.

Chiang, W., D. Chhajed, J. Hess. 2003. Direct marketing, indirect profits: A strategic analysis of dual-channel supply-chain design. *Man. Sci.* **49**(1) 1–20.

Gupta, D. 2008. Flexible carrier-forwarder contracts for air cargo business. *Journal of Pricing and Revenue Management* **7** 341–356.

Hellermann, R. 2006. *Capacity Options for Revenue Management: Theory and Applications in the Air Cargo Industry*. Springer.

Huang, K., W. Hsu. 2005. Revenue management for air cargo space with supply uncertainty. *Proceedings of the Eastern Asia Society for Transportation Studies* **5** 570–580.

IATA. 2009. Fact sheet: Industry statistics Http://www.iata.org/pressroom/facts_figures/fact_sheets/.

Karaesmen, I. Z. 2001. Three essays on revenue management. Ph.D. thesis, Columbia University.

Kasilingam, R. 1996. Air cargo revenue management: characteristics and complexities. *European Journal of Operational Research* **96**(1) 36–44.

Luo, S., M. Cakanyildirim, R. Kasilingam. 2009. Two-dimensional cargo overbooking models. *EJOR* **197**(3) 862–883.

Moussawi, L., M. Cakanyildirim. 2005. Profit maximization in air cargo overbooking. Tech. rep., University of Texas at Dallas.

Nisan, N. 2000. Bidding and allocation in combinatorial auctions. *Proceedings of the 2nd ACM Conference on Electronic Commerce*. 1–12.

Pak, K., R. Dekker. 2005. Cargo revenue management: Bid-prices for a 0-1 multi knapsack problem. *Working paper, Erasmus University Rotterdam* .

Sandhu, R., D. Klabjan. 2006. Fleeting with passenger and cargo origin-destination booking control. *Transportation Science* **40**(4) 517–528.

Simpson, R. W. 1989. Using network flow techniques to find shadow prices for market and seat inventory control. Tech. rep., MIT Flight Transportation Laboratory Memorandum M89-1.

Slager, B., L. Kapteijns. 2003. Implementation of cargo revenue management at KLM. *Journal of Revenue and Pricing Management* **3**(1) 80–90.

Song, J., P. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Oper. Res.* **41**(2) 351–370.

Talluri, K., G. van Ryzin. 1998. An analysis of bid-price controls for network revenue management. *Management Science* **44**(11) 1577–1593.

Talluri, K., G. van Ryzin. 1999. A randomized linear programming method for computing network bid prices. *Transportation Science* **33**(2) 207–216.

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

Topaloglu, H. 2009. Using Lagrangian relaxation to compute capacity-dependent bid-prices in network revenue management. *Oper. Res.* **57**(3) 637–649.

Tsay, A., N. Agrawal. 2004. Modeling conflict and coordination in multi-channel distribution systems: A review. D. Simchi-Levi, D. Wu, M. Shen, eds., *Handbook of Quantitative Supply Chain Analysis in the e-Business Era*, chap. 13. Springer, 557–606.

van Ryzin, G., Q. Liu. 2008. On the choice-based linear programming model for network revenue mangement. *Manuf. Serv. Oper. Man.* **10**(2) 288–310.

Williamson, E. L. 1992. Airline network seat control. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Xiao, B., W. Yang. 2010. A revenue management model for products with two capacity dimensions. *EJOR* **205** 412–421.

Xie, J., S. Shugan. 2001. Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Science* **20**(3) 219–243.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix

EC.1. Proofs

Proof of Proposition 1

We show the result by using induction over the time periods. We begin by showing the result for time period $\tau+1$. Rearranging the objective function of problem (9)-(10), we observe that this problem is equivalent to maximizing

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\Big\{[\hat{C}^a_{ij}-C^a_{ij}-\lambda_j\,V^a_{ij}]\,z^a_{ij}-\hat{C}^a_{ij}]\Big\}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}\Big\{[\hat{C}^s_{jkl}-C^s_{jkl}-\lambda_j\,V^s_{jkl}]\,z^s_{jkl}-\hat{C}^s_{jkl}\Big\}+\sum_{j\in\mathcal{J}}\lambda_j\,\bar{V}_j$$

subject to (3)-(4). By constraints (3), the optimal value of the decision variable z^a_{ij} is one if $\hat{C}^a_{ij}\geq C^a_{ij}+\lambda_j\,V^a_{ij}$ and zero otherwise. By constraints (4), the optimal value of the decision variable z^s_{jkl} is one if $\hat{C}^s_{jk}\geq C^s_{jk}+\lambda_j\,V^s_{jkl}$ and zero otherwise. Thus, problem (9)-(10) has the optimal objective value

$$\begin{aligned}&\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\Big\{[\hat{C}^a_{ij}-C^a_{ij}-\lambda_j\,V^a_{ij}]^+-\hat{C}^a_{ij}]\Big\}+\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}\Big\{[\hat{C}^s_{jkl}-C^s_{jkl}-\lambda_j\,V^s_{jkl}]^+-\hat{C}^s_{jkl}\Big\}+\sum_{j\in\mathcal{J}}\lambda_j\,\bar{V}_j\\
&\quad=-\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\,\min(C^a_{ij}+\lambda_j\,V^a_{ij},\hat{C}^a_{ij})-\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}\min(C^s_{jkl}+\lambda_j\,V^s_{jkl},\hat{C}^s_{jkl})+\sum_{j\in\mathcal{J}}\lambda_j\,\bar{V}_j.\end{aligned}$$

Taking expectations in the expression above, noting the definitions of $B^a_{ij}(\lambda_j)$ and $B^s_{jk}(\lambda_j)$ and the fact that $\mathbb{E}\{\min(C^s_{jkl}+\lambda_j\,V^s_{jkl},\hat{C}^s_{jkl})\}=\mathbb{E}\{\min(C^s_{jk}+\lambda_j\,V^s_{jk},\hat{C}^s_{jk})\}=B^s_{jk}(\lambda_j)$, we obtain

$$\mathbb{E}\{\tilde{\Gamma}(x,n,\lambda,U)\}=-\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\,B^a_{ij}(\lambda_j)-\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} n_{jk}\,B^s_{jk}(\lambda_j)+\sum_{j\in\mathcal{J}}\lambda_j\,\bar{V}_j$$

and the result holds for time period $\tau+1$.

Assuming that the result holds for time period $t+1$, noting that $\tilde{J}_{t+1}(x,n+e_{jk},\lambda)-\tilde{J}_{t+1}(x,n,\lambda)=-B^s_{jk}(\lambda_j)$ and plugging this expression in (11), we obtain

$$\begin{aligned}\tilde{J}_t(x,n,\lambda)=\max_{\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}}&\Bigg\{\sum_{(j,k)\in\mathcal{S}} P_{jkt}(\mathcal{S})\big[\mathbb{E}\{R^s_{jk}\}-B^s_{jk}(\lambda_j)]\Bigg\}\\
&+\sum_{j\in\mathcal{J}}\lambda_j\,\bar{V}_j-\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\,B^a_{ij}(\lambda_j)-\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}} B^s_{jk}(\lambda_j)\,n_{jk}\\
&+\sum_{t'=t+1}^{\tau}\max_{\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}}\Bigg\{\sum_{(j,k)\in\mathcal{S}} P_{jkt'}(\mathcal{S})\,\big[\mathbb{E}\{R^s_{jk}\}-B^s_{jk}(\lambda_j)\big]\Bigg\}\end{aligned}$$

and the result holds at time period t. This concludes the proof.

Proof of Proposition 2

The proof is immediate by the relaxation argument since $\tilde{\Gamma}(x,n,\lambda,U)$ is an upper bound for $\Gamma(x,n,U)$ as long as $\lambda \geq 0$. The optimality equations in (11) is identical to the one in (5), except that its boundary condition is an upper bound on the boundary condition of the optimality equation in (5). Therefore, the value functions computed through the optimality equation in (11) are upper bounds on those computed through the optimality equation in (5).

Proof of Proposition 3

Suppose the algorithm terminated. Consider (x^S, y^S) and λ^{S+1} at termination. Since λ^{S+1} is within $\frac{\epsilon}{3}$ of the optimum, that is

$$\tilde{J}_1(x^S, \bar{0}, \lambda^{S+1}) \leq \min_{\lambda \geq 0} \; \tilde{J}_1(x^S, \bar{0}, \lambda) + \frac{\epsilon}{3},$$

and $\tilde{J}_1(x^S, \bar{0}, \lambda^{S+1}) \geq y^S - \frac{\epsilon}{3}$ by the termination criterion, it follows that

$$y^S - \frac{2\epsilon}{3} \leq \min_{\lambda \geq 0} \; \tilde{J}_1(x^S, \bar{0}, \lambda),$$

which can be written as

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}_i} x_i^S \, B_{ij}^a(\lambda) + y^S - \frac{2\epsilon}{3} \leq \sum_{j \in \mathcal{J}} \lambda_j \, \bar{V}_j + \sum_{t=1}^{\tau} \Phi_t(\lambda) \qquad \text{for all } \lambda \geq 0.$$

That is, $(x^S, y^S - 2\epsilon/3)$ is feasible for problem (P). Since the optimal objective value of problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) is within $\epsilon/3$ of the objective value provided by the solution (x^S, y^S), the optimal objective value of problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) is within ϵ of the objective value provided by the solution $(x^S, y^S - 2\epsilon/3)$. Since problem (R-$\{\lambda^1, \ldots, \lambda^S\}$) is a relaxation of problem (P), this implies that $(x^S, y^S - 2\epsilon/3)$ has a value within ϵ of the optimal solution to problem (P).

To prove finite termination, suppose on the contrary that the algorithm does not terminate in a finite number of steps. We observe that $y^S = \min_{s=1,\ldots,S} \tilde{J}_1(x^S, \bar{0}, \lambda^s)$ at any iteration of the algorithm. Moreover, since there is only a finite number of feasible values for the binary x-variables, at least one of them must be visited more than once. That is $x^S = x^{s'}$ for some $s' < S$. Since the algorithm does not terminate at step S, it must be that

$$y^S > \tilde{J}_1(x^S, \bar{0}, \lambda^{S+1}) + \frac{\epsilon}{3} \geq \min_{\lambda \geq 0} \; \tilde{J}_1(x^S, \bar{0}, \lambda) + \frac{\epsilon}{3}.$$

On the other hand, since $\lambda^{s'}$ is $\epsilon/3$-optimal at $x^{s'} = x^S$, the fact that $y^S = \min_{s=1,\ldots,S} \tilde{J}_1(x^S, \bar{0}, \lambda^s)$ implies

$$y^S \leq \tilde{J}_1(x^{s'}, \bar{0}, \lambda^{s'}) \leq \min_{\lambda \geq 0} \tilde{J}_1(x^{s'}, \bar{0}, \lambda) + \frac{\epsilon}{3},$$

which is a contradiction.

Proof of Proposition 4

We begin by defining the partial ordering $\succeq_{jk}$ to write $\lambda^+ \succeq_{jk} \lambda^-$ when $B^s_{jk}(\lambda^+) \geq B^s_{jk}(\lambda^-)$. To prove the proposition, we equivalently show that $\lambda^*(x, n + e_{jk}) \succeq_{jk} \lambda^*(x, n)$. For notational brevity, we let $\lambda^+ = \lambda^*(x, n + e_{jk})$ and $\lambda^0 = \lambda^*(x, n)$. We choose λ^- such that $\lambda^0 \succeq_{jk} \lambda^-$. By (12), we have

$$\begin{aligned}\tilde{J}_t(x, n, \lambda^0) - \tilde{J}_t(x, n + e_{jk}, \lambda^0) &= B^s_{jk}(\lambda^0) \\ &\geq B^s_{jk}(\lambda^-) = \tilde{J}_t(x, n, \lambda^-) - \tilde{J}_t(x, n + e_{jk}, \lambda^-),\end{aligned}$$

where the inequality follows from the fact that $\lambda^0 \succeq_{jk} \lambda^-$. Arranging the terms, the expression above can be written as

$$\tilde{J}_t(x, n + e_{jk}, \lambda^0) \leq \tilde{J}_t(x, n + e_{jk}, \lambda^-) + \tilde{J}_t(x, n, \lambda^0) - \tilde{J}_t(x, n, \lambda^-). \tag{EC.1}$$

Since λ^0 is an optimal solution to problem $\min_{\lambda \geq 0} \tilde{J}_t(x, n, \lambda)$, we have $\tilde{J}_t(n, x, \lambda^0) \leq \tilde{J}_t(n, x, \lambda^-)$. In this case, by (EC.1), we have

$$\tilde{J}_t(x, n + e_{jk}, \lambda^0) \leq \tilde{J}_t(x, n + e_{jk}, \lambda^-) \tag{EC.2}$$

for all λ^- such that $\lambda^0 \succeq_{jk} \lambda^-$. This result immediately implies that $\lambda^+ \succeq_{jk} \lambda^0$. In particular, if, on the contrary, $\lambda^+ \not\succeq_{jk} \lambda^0$, then we have $B^s_{jk}(\lambda^0_j) > B^s_{jk}(\lambda^+_j)$. This implies that $\lambda^0 \succeq_{jk} \lambda^+$, in which case, we can use (EC.2) with $\lambda^- = \lambda^+$ to obtain $\tilde{J}_t(x, n + e_{jk}, \lambda^0) \leq \tilde{J}_t(x, n + e_{jk}, \lambda^+)$. This inequality and the fact that $B^s_{jk}(\lambda^0_j) > B^s_{jk}(\lambda^+_j)$ contradict the fact that (λ^+) is an optimal solution to the problem $\min_{\lambda \geq 0} \tilde{J}_t(x, n + e_{jk}, \lambda)$ that is largest according to the partial ordering $\succeq_{jk}$.

EC.2. Elements of Experimental Setup

In this section, we describe further elements of experimental setup, including the distributions used in the random sampling of data, the volume and weight distributions of allotments and spot market cargo, as well as the details of the revenue, cost, and penalty models for the cargo.

We start by describing the spot market cargo types and the model for their volumes, weights, revenues, costs and penalties. As already mentioned, the weight of each cargo type is assumed to be constant in the experiments. The value of constant weight V_{2k}^s of each type k is randomly sampled from the exponential distribution with the mean of 0.25 truncated to the interval $[0.05, 3.5]$ (all values in tons). The reference density δ_k is sampled from the normal distribution with the mean 0.167 and the standard deviation 0.04 truncated to the interval $[0.03, 0.4]$ (all values in ton/m^3). The volume V_{1k}^s of cargo type k is then a lognormal random variable with the mean V_{2k}^s/δ_k and the coefficient of variation θ^s (common for all types and varied in a controlled fashion). The revenue of type k cargo (payment by a customer to the airline) is computed as $R_k = \rho_k \max\{V_{2k}^s, V_{1k}^s/\gamma\}$, where $\gamma = 6$ m^3/ton is an industry-standard constant and ρ_k is the shipping rate randomly sampled from the interval $[1500, 3000]$ (in \$/ton). The regular shipping cost and penalty of type k are proportional to the shipping rate and are computed in the experiments as $C_k^s = \rho_k(0.04V_{1k}^s + 0.4V_{2k}^s)$ and $\hat{C}_k^s = \rho_k(0.15V_{1k}^s + 1.5V_{2k}^s)$ (a lower scale in volume is selected primarily to adjust for the units of measurement).

The demand for each cargo type on the spot market is subject to consumer choice among the flights departing within the next two weeks. There are a total of 8 flights within this time span. The choice between flights is described by a multinomial logit model (EC.3) in which the preference weights v_{jkt} depend on the time of the flight j within a week as well as on the number of weeks before the week of flight departure (this number can be 0, 1 or 2). The required $3 \times 4 = 12$ preference weights are generated as $0.9e^{0.1Z}$ where Z is a standard normal random variable. The no-purchase option has a constant preference weight $v_{0kt} = 0.1$.

Finally, we sample the maximum allotted weight W_i of each bid i from the normal distribution with mean 5 and standard deviation 2 tons (the same for all flights included in the bid). Utilization of capacity by allotments is described by a hierarchical model. The utilized fraction F_{ij} of the allotted weight of bid i on flight j is distributed as normal with mean 0.9 and standard deviation 1/9.7 truncated to the interval [0,1] (the parameters are selected so that utilized weight follows a practical rule of thumb – at least 80% weight utilization at least 80% of the time). Conditionally on utilized weight $V^a_{2ij} = F_{ij}W_i$, utilized volume V^a_{1ij} is distributed as a lognormal random variable with the mean V^a_{2ij}/γ and the coefficient of variation θ^a (controlled in the experiments). Allotment i revenues from flight j are $R_{ij} = \rho_i \max\{V^a_{2ij}, V^a_{1ij}/\gamma\}$. The cost and penalty models are similar to those of spot market cargo (although the costs are potentially lower and penalties are potentially higher, both unrelated to the rates). In particular, we let $C^a_{ij} = 25V^a_{1ij} + 250V^a_{2ij}$ and $\hat{C}^a_{ij} = 150V^a_{1ij} + 1500V^a_{2ij}$.

EC.3. Uncertainty Structure, Its Implications for Independence Between Flights and Possible Generalizations

In this section, we provide a detailed discussions of the uncertainty structure in our model, and, in particular, its independence assumptions. There are three source of uncertainty in the model: (1) capacity utilizations, payments and costs associated with allotments, (2) capacity utilizations, payments and costs associated with accepted spot market bookings, and (3) spot market demand resulting from consumer choice behavior. In each case, we choose to keep specifications of distributions general as long as the same modelling and analytic approach can be used. Moreover, the general form of distributions, particularly for payment and cost structures, has the advantage of capturing the industry practices which vary across different cargo carriers. The essence of independence assumptions can be summarized by two aspects: independence of distributional specifications and statistical independence.

By independence of specifications, we mean that distributions depend only on the indices of the object they describe: allotment quotes indexed by atomic bid/flight combinations, spot market cargo bookings by flight and cargo type, and spot market choice probabilities by flight, type and time. This excludes any dependence of input distributions on the allotment decisions, the current number of spot market bookings or cargo loading decisions.

The statistical independence is used in its usual sense: allotment capacity utilizations/revenues for each atomic bid/flight combination, spot market cargo capacity utilizations/revenues for each flight/cargo type combination, and spot market demand realizations for each decision period are mutually independent. The immediate implication of statistical independence is that probability measure describing future behavior of the system does not depend on any past realizations of uncertain quantities. Moreover, since all bookings which cannot be accommodated on their scheduled flight are outsourced, this probability measure also does not depend on loading decisions made at departure times of the flights.

The most important implication of independence assumption is that the state of the system can be completely described by the allotment decisions and the current numbers of spot market bookings by flight/type. Since the state is not affected by loading decisions, and the objectives/feasible sets of cargo loading problems for each flight are independent, it is irrelevant when the departures occur. This explains why it is appropriate to assume that all departures occur at the end and all loading decisions are made at the same time. In practice, the loading problems for each flight are solved on on-going basis but without affecting the system state. This does not affect the allotment selection problem since all of the flights occur in the future. For the spot market booking control, the state components which apply to departed flights are irrelevant and can be ignored in the dynamic programming formulation. In the practical approximation of booking control (described in §6 and §EC.5), we can ignore the flights which have already departed.

Since the ability to treat flight departures in independent fashion is clearly very attractive from an analytic point of view and important in our approach, it is also interesting how independence assumptions can be relaxed without affecting the independence of the flights. One concern is that allotment utilization/revenues and spot market demand usually depend on macroeconomic environment characterized by such factors as interest rates, consumer confidence, and others. It is possible to extend our approach by introducing an uncertain but observable discrete *states of the environment* with Markovian dynamics. (The notion of a "fluctuating demand environment" of this form is discussed, for example, by Song and Zipkin (1993).) All steps in the analysis and computational approach go through with minimal changes. Another concern is a likely dependence of spot market demand on allotment decisions. In terms of our model, this means dependence of the booking probabilities $P_{jkt}(\cdot)$ on allotment decisions. While handling general dependence is difficult, we can extend our approach to the case where the dependence of $P_{jkt}(\cdot)$ on x_i's is *linear*. This is reasonable, for example, if each accepted spot market booking is estimated to cause a given reduction in booking probabilities:

$$P_{jkt}(x,\mathcal{S}) = P^0_{jkt}(\mathcal{S}) - \sum_{i\in\mathcal{I}} x_i\,\pi^i_{jkt}(\mathcal{S}),$$

where $P^0_{jkt}(\mathcal{S})$ is a booking probability when no bids are granted, and $\pi^i_{jkt}(\mathcal{S})$ is a reduction in the booking probability if atomic bid $i\in\mathcal{I}$ is granted. With the booking probabilities of this form, our approach is still applicable, because the Lagrangian bound remains convex in λ for fixed x and linear in x for fixed λ.

EC.4. Perfect Hindsight Upper Bound for the Case of Related Departures

In this section, we relax the assumption of independent loading decisions for different flights and allow the possibility to rebook a shipment, at a cost, for a later flight. The cost model for the shipments is generalized as follows. Let the shipping cost of a type k package originally booked for flight j but shipped on flight j' be described by a random variable $\tilde{C}^s_{jj'k}$. If it is not possible to shift cargo from flight j to j', then we can assume that $\tilde{C}^s_{jj'k}$ takes a prohibitively large value. Similarly, let the shipping cost of allotment i allocated to flight j but shipped on flight j' be $\tilde{C}^a_{ijj'}$. Finally, let flight ϕ be a dummy flight such that shipping on this flight represents outsourcing and outsourcing costs are represented by $\tilde{C}^s_{j\phi k}$ and $\tilde{C}^a_{ij\phi}$. The main difficulty in the exact handling of the loading decisions in this case is that if there are booking requests in between an earlier and a later flight, then the load on the later flight is not known at the time of departure of the early flight. Thus, loading problems of the earlier flight are subject to significant additional uncertainty. However, one can construct an upper bound to the combined loading problem of all flights by finding the loading decisions in hindsight, that is, under the assumption that all loading decisions are made simultaneously and with perfect knowledge of the number of shippings and their characteristics for all flights. Thus, the combined loading problem becomes

$$\begin{aligned}
\max \quad & -\sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}}\sum_{j'\in\mathcal{J}\cup\{\phi\}} \tilde{C}^s_{jj'kl}\, z^s_{jj'kl} - \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i}\sum_{j'\in\mathcal{J}\cup\{\phi\}} x_i\, \tilde{C}^a_{ijj'}\, z^a_{ijj'} \\
\text{subject to} \quad & \sum_{j\in\mathcal{J}}\sum_{k\in\mathcal{K}}\sum_{l=1}^{n_{jk}} V^s_{jkl}\, z^s_{jj'kl} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}_i} x_i\, V^a_{ij}\, z^a_{ijj'} \le \bar{V}_{j'} \qquad \text{for all } j'\in\mathcal{J} \\
& \sum_{j'\in\mathcal{J}\cup\{\phi\}} z^s_{jj'kl} = 1 \qquad \text{for all } l=1,\dots,n_{jk},\ k\in\mathcal{K},\ j\in\mathcal{J}, \\
& \sum_{j'\in\mathcal{J}\cup\{\phi\}} z^a_{ijj'} = 1 \qquad \text{for all } i\in\mathcal{I},\ j\in\mathcal{J}_i, \\
& z^s_{jj'kl}\in\{0,1\} \qquad \text{for all } l=1,\dots,n_{jk},\ k\in\mathcal{K},\ j\in\mathcal{J},\ j'\in\mathcal{J}\cup\{\phi\} \\
& 0\le z^a_{ijj'}\le 1 \qquad \text{for all } i\in\mathcal{I},\ j\in\mathcal{J}_i,\ j'\in\mathcal{J}\cup\{\phi\}.
\end{aligned}$$

In the problem above, the decision variable $z^s_{jj'kl}$ takes value one if the l-th booked spot market request for cargo type k on flight j is served through flight j'. Similarly, $z^a_{ijj'}$ corresponds to the portion of the cargo generated by atomic bid i on flight j that is served through flight j'. Using the problem in the boundary condition of the optimality equation in (5), all of our development goes through without any modifications to the methodology. In the resulting expressions (such as equations (12) and (13) of Proposition 1), the functions $B^a_{ij}(\lambda_j)$ and $B^s_{jk}(\lambda_j)$ are replaced by

$$\begin{aligned}
\tilde{B}^a_{ij}(\lambda) &= \mathbb{E}\Big[\min_{j\in\mathcal{J}\cup\{\phi\}} (\tilde{C}^a_{ijj'} + \lambda_j V^a_{ij})\Big], \text{ and} \\
\tilde{B}^s_{jk}(\lambda) &= \mathbb{E}\Big[\min_{j\in\mathcal{J}\cup\{\phi\}} (\tilde{C}^s_{jj'kl} + \lambda_j V^s_{jkl})\Big],
\end{aligned}$$

respectively (with a convention that $\lambda_\phi \equiv 0$).

EC.5. Consumer choice model for efficient implementation

Whether we compute the Lagrange multipliers once at the beginning of the planning horizon by solving the problem $\min_{\lambda\ge 0} \tilde{J}_1(x^*, \bar{0}, \lambda)$ or recompute them at every time period by solving the problem $\min_{\lambda\ge 0} \tilde{J}_t(x^*, n, \lambda)$, the spot market booking control decisions require solving the optimization problem in (21). This is a challenging combinatorial optimization problem as it involves choosing a subset of $\mathcal{J}\times\mathcal{K}$ and there are potentially $2^{|\mathcal{J}|\times|\mathcal{K}|}$ such subsets. Enumerating over all possible subsets is clearly not an option. However, it turns out that there is an important class of consumer choice models that make the optimization problem in (21) tractable. In this section, we briefly review this class of consumer choice models.

We recall that $P_{jkt}(\mathcal{S})$ is the probability that there is a spot market booking for flight j and cargo type k at time period t given that the set of open flight and cargo type combinations at time period t is $\mathcal{S}$. To construct a tractable model for $P_{jkt}(\mathcal{S})$, we assume that a customer that is interested in making a spot market booking for cargo type k arrives into the system at time period t with probability π_{kt}. We naturally have $\sum_{k\in\mathcal{K}} \pi_{kt} \leq 1$ and the difference between the two sides of the inequality gives the probability that there is no customer arrival at time period t. Once a customer that is interested in cargo type k arrives into the system, this customer makes a choice among the flights according to the multinomial logit choice model. The multinomial logit choice model stipulates that if a customer that arrives into the system is interested in cargo type k, then this customer associates the preference weights $\{v_{jkt} : j \in \mathcal{J}\}$ with the different flights. These preference weights characterize the attractiveness of different flights to the customer. The customer also associates the preference weight v_{0kt} with the no booking option. In this case, if the set of open flight and cargo type combinations at time period t is given by $\mathcal{S}$, then the customer chooses flight j with probability $v_{jkt}/[\sum_{j':(j',k)\in\mathcal{S}} v_{j'kt} + v_{0kt}]$, which is the preference weight of flight j relative to the preference weights of all available options. Therefore, the probability that there is a spot market booking for flight j and cargo type k at time period t is given by

$$P_{jkt}(\mathcal{S}) = \pi_{kt} \frac{v_{jkt}}{\sum_{j':(j',k)\in\mathcal{S}} v_{j'kt} + v_{0kt}}. \tag{EC.3}$$

Multinomial logit choice model is a widely used tool in marketing and economics [see Anderson et al. (1992)] as well as, more recently, in revenue management [see Talluri and van Ryzin (2004) and van Ryzin and Liu (2008)].

Plugging the expression above for $P_{jkt}(\mathcal{S})$ into the optimization problem in (21) and dropping the term $\tilde{J}_{t+1}(x^*, n, \lambda^{**})$ that does not affect the spot market booking control decisions, we need to solve the problem

$$\max_{\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}}\left\{\sum_{(j,k)\in\mathcal{S}}\frac{\pi_{kt}\,v_{jkt}\left[\mathbb{E}\{R^s_{jk}\}-B^s_{jk}(\lambda^{**}_j)\right]}{\sum_{j':(j',k)\in\mathcal{S}}v_{j'kt}+v_{0kt}}\right\}=\max_{\mathcal{S}\subseteq\mathcal{J}\times\mathcal{K}}\left\{\sum_{k\in\mathcal{K}}\left[\frac{\sum_{j:(j,k)\in\mathcal{S}}\pi_{kt}\,v_{jkt}\left[\mathbb{E}\{R^s_{jk}\}-B^s_{jk}(\lambda^{**}_j)\right]}{\sum_{j':(j',k)\in\mathcal{S}}v_{j'kt}+v_{0kt}}\right]\right\}$$

to make the booking control decisions at time period t. In the expression above, the equality follows from the fact that a sum over all $(j,k)\in\mathcal{S}$ can be written as one sum over all $k\in\mathcal{K}$ and another sum over all $j\in\mathcal{J}$ satisfying $(j,k)\in\mathcal{S}$. The expression in the square brackets corresponds to the contribution from cargo type k and the contributions from different cargo types do not interact with each other. This implies that we can solve the problem on the right side above by maximizing the contribution from each cargo type individually. In particular, if we let $\mathcal{J}_k$ be the set of flights that are open for spot market bookings for cargo type k, then the optimal objective value of the problem on the right side above is given by $\sum_{k\in\mathcal{K}}\psi_{kt}(\lambda^{**})$, where we have

$$\psi_{kt}(\lambda^{**})=\max_{\mathcal{J}_k\subseteq\mathcal{J}}\left\{\frac{\sum_{j\in\mathcal{J}_k}\pi_{kt}\,v_{jkt}\left[\mathbb{E}\{R^s_{jk}\}-B^s_{jk}(\lambda^{**}_j)\right]}{\sum_{j'\in\mathcal{J}_k}v_{j'kt}+v_{0kt}}\right\}. \tag{EC.4}$$

Problem (EC.4) is slightly more tractable than problem (21) as this problem involves the subsets of $\mathcal{J}$, which number on the order of $2^{|\mathcal{J}|}$, whereas problem (21) involves the subsets of $\mathcal{J}\times\mathcal{K}$, which number on the order of $2^{|\mathcal{J}|\times|\mathcal{K}|}$. However, for practical applications, $\mathcal{J}$ may still have too many subsets to enumerate explicitly.

It turns out that problem (EC.4) has a very special structure that makes it solvable in a tractable fashion. To illustrate this property, we assume without loss of generality that the set of flights is indexed by $\mathcal{J}=\{1,\ldots,|\mathcal{J}|\}$ and the flights are ordered such that

$$\mathbb{E}\{R^s_{1k}\}-B^s_{1k}(\lambda^{**}_1)\geq\mathbb{E}\{R^s_{2k}\}-B^s_{2k}(\lambda^{**}_2)\geq\ldots\geq\mathbb{E}\{R^s_{|\mathcal{J}|,k}\}-B^s_{|\mathcal{J}|,k}(\lambda^{**}_{|\mathcal{J}|}).$$

In this case, Talluri and van Ryzin (2004) show that there is an optimal solution to problem (EC.4) of the form $\{1,2,\ldots,j\}\subseteq\mathcal{J}$ for some $j\in\mathcal{J}$. Therefore, the possible candidates for an optimal solution to problem (EC.4) are $\varnothing,\{1\},\{1,2\},\ldots,\{1,2,\ldots,|\mathcal{J}|\}$. Since there are only $|\mathcal{J}|+1$ possible candidates for an optimal solution to (EC.4), we can check the objective function value provided by each one of these candidates and choose the best one. This result eliminates the need to enumerate over all subsets of $\mathcal{J}$ to solve problem (EC.4).