

Single-Leg Air-Cargo Revenue Management

Kannapha Amaruchkul, William L. Cooper, Diwakar Gupta

Graduate Program in Industrial and Systems Engineering, Department of Mechanical Engineering,
University of Minnesota {amar0017@umn.edu, billcoop@me.umn.edu, guptad@me.umn.edu}

We consider a cargo booking problem on a single-leg flight with the goal of maximizing expected contribution. Each piece of cargo is endowed with a random volume and a random weight whose precise values are not known until just before the flight's departure. We formulate the problem as a Markov decision process (MDP). Exact solution of the formulation is impractical, because of its high-dimensional state space; therefore, we develop six heuristics. The first four heuristics are based on different value-function approximations derived from two computationally tractable MDPs, each with a one-dimensional state space. The remaining two heuristics are obtained from solving related mathematical programming problems. We also compare the heuristics with the first-come, first-served (FCFS) policy. Simulation experiments suggest that the value function approximation derived from separate "volume" and "weight" problems offers the best approach. Comparisons of the expected contribution under the heuristic to an upper bound show that the heuristic is typically close to optimal.

Key words: air-cargo operations; revenue management

History: Received: April 2005; revision received: March 2006; accepted: August 2006.

Introduction

Air-cargo operations are a significant source of revenue for passenger airlines, which currently carry a bulk of worldwide air cargo either in the belly of passenger planes or on freight-only planes. Until recently, the air-cargo industry grew 6%–7% annually, and forecasts suggest that it will continue to expand at an average annual rate of 6.4% over the next two decades (Boeing Company 2005b). Although most airlines have implemented highly sophisticated revenue management (RM) systems for controlling prices and availability of passenger tickets to maximize revenue, related methodologies for managing cargo bookings are relatively less sophisticated.

There are several key differences between cargo RM and passenger RM, which explain why passenger RM techniques cannot be easily transported to the cargo environment. Cargo consumes multidimensional capacity: Weight and volume are two such dimensions. In addition, customers often do not know the true volume and weight of a shipment at the time of advance booking. Instead, the volume and weight are learned only shortly before the cargo shipment is to be loaded onto the plane. Moreover, if the volume or weight requirements of booked shipments exceed available capacity, then some cargo needs to be off-loaded; determining what to off-load is a computationally hard problem. Yet another difference, which we do not model in this paper, is that there may be many different routes cargo can take between its origin and destination, and it is largely up to the carrier to choose a route. (For a review of vehicle routing problems, see Powell, Jaillet, and Odoni 1995

or Toth and Vigo 2002.) Finally, air-cargo space can be sold either as an allotment or on a free-sale basis (Slager and Kapteijns 2004). A prenegotiated amount of capacity on a specific flight or weekday is reserved for major shippers and forwarders through the allotment, whereas there are no capacity guarantees in the free-sale mode. We focus in this article on managing bookings against free sale capacity.

We present a Markov decision process (MDP) model of the free sale air-cargo booking process for a single flight with multiple shipment types. A shipment type is characterized by two properties: a volume-weight category and a freight-type class. Associated with each category is a joint distribution of volume and weight, whose realized values are known only at the end of the booking horizon. Freight type refers to the type of cargo, e.g., fresh meat, flowers, electronic components, which determines freight charges. The objective is to find an accept or reject policy that maximizes the expected net contribution margin from cargo bookings. Contribution earned from a shipment depends on weight, volume, and the type of each cargo shipment. Shipments that are declined may be accepted for delivery on another date or sent on alternate routes or flights, but we do not model such situations. We also do not penalize the carrier for declinations (other than lost margin), however, if we were to add the penalty for each rejected booking request, the formulation would be the same up to a constant.

Exact solution of the MDP is difficult because of its multidimensional state space. Therefore it is necessary

to develop easy-to-compute bounds and heuristics that are scalable to industry-sized problems. We propose a heuristic based on decomposing the MDP into separate “weight” and “volume” MDPs, and in addition, we obtain several easily computable upper bounds on the value of the original MDP. We test the performance of the heuristic in numerical experiments.

In these experiments, we find that the estimated expected contribution from the decomposition heuristic is relatively close to the tightest upper bound. In addition, we compare the recommended heuristic against other heuristics such as the first-come, first-served (FCFS) regime, a partitioned allocation scheme, a bid-price policy, and other value function approximations. Our heuristic outperforms other heuristics in almost all the problem settings considered. Conditions that affect the heuristic’s performance are also identified. We find that the difference between the expected contribution under the heuristic and the best upper bound increases when capacity is tight relative to expected demand or when uncertainty in the volume distribution is high, given that the weight of a shipment is deterministic.

As mentioned earlier, most carriers already use sophisticated passenger RM systems. Commensurately, there is a vast literature on passenger RM (see Talluri and van Ryzin 2004 for a survey). However, air-cargo RM methodologies are less sophisticated, and there is little published work on this topic aside from some papers that provide overviews of air-cargo operations. Examples include Kasilingam (1996) and Billings, Diener, and Yuen (2003), which describe the characteristics and complexities of air-cargo RM; Bodendorf and Reinheimer (1998), which presents a decision support system for a Web-based air-cargo market, and Slager and Kapteijns (2004), which describes the implementation of cargo yield management processes at KLM Royal Dutch Airlines. DeLain and O’Meara (2004) outline techniques for evaluating costs and benefits from cargo RM implementation. These papers, however, do not focus on mathematical models.

Pak and Dekker (2004) formulate the short-term booking control problem as a multidimensional dynamic knapsack problem. They assume that the actual weight and volume of a shipment do not differ from the booked quantities, whereas we assume random volume and weight, which are realized shortly prior to the flight departure. Moussawi and Çakanyildirim (2005) and Luo and Çakanyildirim (2005) develop two-dimensional overbooking models, but do not consider a dynamic booking control problem as we do. Mathematical models for network routing optimization can be found in Prior, Slavens, and Trimarco (2004) and Bartodziej and Derigs (2004). These papers

focus on optimizing aircraft scheduling and routing in the network, given a set of fixed accepted requests. This work also differs from ours, because we focus on deciding whether or not to accept booking requests for a single-leg flight.

Air-cargo RM problems bear some similarities to passenger RM problems with multiple seat bookings. When booking a shipment, the carrier cannot accept a partial request, just as a typical passenger group cannot be split. Lee and Hersh (1993) study a single-leg passenger RM problem that incorporates batch arrivals. Van Slyke and Young (2000) consider a single-leg passenger RM problem with group reservations as a special case of the finite-horizon stochastic knapsack problem, and propose a general algorithm to solve the problem. However, their algorithm is computationally impractical for solving large air-cargo problems. Kleywegt and Papastavrou (2001) study a variation of the stochastic knapsack problem, and provide sufficient conditions for monotonicity and convexity properties of the optimal value function. With the exception of Van Slyke and Young (2000), these models do not allow arrivals to have multidimensional resource requirements. In contrast, our model assumes that each cargo booking request has two dimensions: a volume and a weight.

In passenger RM problems, a passenger can cancel or upgrade a reservation; analogously, in air-cargo RM problems, a shipment’s size may differ from that stated in the original reservation. Subramanian, Stidham, and Lautenbacher (1999) provide a discrete-time MDP model of a single-leg passenger RM problem that incorporates overbooking, cancellations, and no-shows. Zhao and Zheng (2001) study a two-class dynamic seat allocation model, which includes both cancellations and passenger diversion. Karaesmen and van Ryzin (2004) consider a problem that includes multiple substitutable reservation classes. These passenger RM papers assume that should a booked passenger show up for the flight(s), the amount of resource (s)he will consume (one seat) is the amount specified in the original booking. Our model also includes the possibility of overbooking, however, the volume of a shipment that shows up is random. As described in §4, we can easily incorporate no-shows in our model as well.

The remainder of this paper is organized as follows. In §1, we present the MDP formulation and discuss computational difficulties associated with it. In §2, we provide a variety of bounds for our problem and scalable heuristics for making booking decisions. In §3, we report the results of numerical tests, and we conclude with some thoughts on future directions in §4. Proofs can be found in the appendix.

1. Formulation

We consider a single flight with volume capacity k_v and weight capacity k_w . The booking horizon is divided into n time periods such that at most one booking request arrives per period. Time periods are numbered in reverse chronological order so that the beginning of the booking horizon is time $t = n$, and the flight departs at time $t = 0$. We assume that each booking request belongs to one of m types and that arrivals are independent across time periods. Let p_{it} be the probability that a type- i booking request arrives in period t , and $p_{0t} = 1 - \sum_{i=1}^m p_{it}$ be the probability that no booking request arrives in period t . Associated with each type is a joint distribution of volume and weight. Upon receiving a booking request, the carrier must decide whether to accept or reject the request. At the time of making the accept or reject decision, the carrier knows the type of a shipment, and consequently knows the joint distribution for volume and weight of the shipment. The carrier, however, does not know its precise volume and weight.

Just before the scheduled flight departure time, the true volume and weight of each booked shipment are realized. If a type- i shipment has volume v and weight w , then its margin is $r_i(\max\{w, v/\gamma\})$, where γ is a constant defined by the International Air Transportation Association (IATA) volumetric standard. The quantity v/γ is called the *dimensional weight*, and $\max\{w, v/\gamma\}$ is called the *chargeable weight* (Thuermer 2005). The margin (also known as *contribution margin*) of a shipment is the revenue it generates less the associated operational costs including, e.g., handling costs and variable fuel costs (Slager and Kapteijns 2004). We implicitly assume that the revenue and the variable costs are functions of the chargeable weight. Throughout, in a slight abuse of terminology, we will use the term revenue interchangeably with margin. That is, we simply refer to $r_i(\cdot)$ as the type- i revenue function.

A shipment type is determined by volume and weight distribution, type of packaging, type of cargo, time sensitivity, and possibly other characteristics. The notion of types is a modeling convenience that allows us to represent situations in which two air-cargo shipments with a common weight and volume distribution are charged differently. This could happen, for instance, if one contains perishable products, and the other contains nonperishable products whose delivery may be delayed.

In case of an oversold situation (that is, if the carrier's capacity (k_v, k_w) is not sufficient to accommodate all of the booked shipments), the carrier needs to decide which shipments to off-load. When examined up close, the real problem of choosing which shipments to off-load is complex. The penalty associated

with oversale may depend on a variety of situation-specific details, including the type of cargo (perishable or not), type of customer, availability or cost of space on other carriers' flights, storage charges, and contractual penalties. Even if overage costs were linear in the amount of chargeable weight off-loaded, the carrier would need to solve an NP-hard integer program to determine precisely which shipments to off-load. In this paper, we simplify the situation by assuming (as an approximation to the complicated reality) that the oversale penalty is a separable function of the total realized volume and weight of the bookings. Specifically, we assume that if the total booked volume and weight are v_T and w_T , then the oversale penalty is $h_v(v_T) + h_w(w_T)$, where $h_v(\cdot)$ and $h_w(\cdot)$ are nonnegative increasing convex functions. Implicit in this simplification is an assumption that, once booked, a shipment can be partially off-loaded (or split) if putting only a portion of it on the flight will prevent exceeding either of the dimensions of capacity. In our numerical examples, we use $h_v(v_T) = h_v(v_T - k_v)^+$ and $h_w(w_T) = h_w(w_T - k_w)^+$, which are linear in the volume and weight in excess of capacity. Note also that the penalty function can be calibrated (set) by the carrier to achieve a policy with desired properties. For instance, setting large h_v and h_w will induce a less aggressive overbooking policy.

1.1. The Optimality Equations

We use a boldface font to denote a vector, \mathbf{e}_i to denote the i th unit m -vector (i.e., a vector whose i th component is one, and all other components are zero), \mathbf{e}_0 to denote the zero m -vector, \mathbb{Z}_+^m to denote the m -fold cross product of the nonnegative integers, and \mathbb{R} to denote the set of real numbers. For a state $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{Z}_+^m$, the quantity x_i represents the number of type- i shipments that have been accepted. We define the single-period reward from accepting a type- i shipment to be $\rho_i = E[r_i(\max\{W_i, V_i/\gamma\})]$, where (V_i, W_i) is a random vector with the joint (volume, weight) distribution of a type- i booking request.

Under this accounting system, the expected revenue from a booking is received immediately, rather than at the departure time. This does not affect the booking policy, and it also does not affect expected revenue, because cash flows are not discounted on account of the fact that bookings occur over a short period.

Let $g_t: \mathbb{Z}_+^m \rightarrow \mathbb{R}$ be the value function at time t ; that is, $g_t(\mathbf{x})$ is the maximum expected revenue that can be obtained from time period t until the time of departure, given the state at the beginning of period t is \mathbf{x} . The objective is to maximize the expected revenue $g_n(\mathbf{e}_0)$ over the entire booking horizon. The argument \mathbf{e}_0 indicates that there are no bookings yet at the beginning of the horizon. A starting state of \mathbf{e}_0 is used

for concreteness. Our approach remains valid for any starting state \mathbf{x}_0 . The value functions can be computed recursively via the Bellman optimality equations

$$g_t(\mathbf{x}) = \sum_{i=1}^m p_{it} \max\{\rho_i + g_{t-1}(\mathbf{x} + \mathbf{e}_i), g_{t-1}(\mathbf{x})\} + p_{0t} g_{t-1}(\mathbf{x}),$$

$$t = 1, \dots, n. \quad (1)$$

To specify the terminal value function $g_0(\mathbf{x})$, suppose (V_{il}, W_{il}) is the volume and weight of the l th type- i shipment, defined such that (V_{il}, W_{il}) is equal in distribution to (V_i, W_i) for each i and l . Furthermore, we assume that $\{(V_{il}, W_{il})\}$ are independent random variable pairs, i.e., (V_{il}, W_{il}) is independent of $(V_{i'l'}, W_{i'l'})$ for all $(i', l') \neq (i, l)$. Then, the terminal value function is

$$g_0(\mathbf{x}) = -E \left[h_v \left(\sum_{i=1}^m \sum_{l=1}^{x_i} V_{il} \right) + h_w \left(\sum_{i=1}^m \sum_{l=1}^{x_i} W_{il} \right) \right]. \quad (2)$$

An optimal policy for the MDP accepts a type- i booking request in period t when the state is \mathbf{x} if and only if

$$\rho_i \geq g_{t-1}(\mathbf{x}) - g_{t-1}(\mathbf{x} + \mathbf{e}_i). \quad (3)$$

The right-hand side of (3) represents the expected drop in future revenue caused by accepting a type- i request in period t , whereas the left-hand side is the expected incremental revenue from accepting this request. Thus the inequality in (3) implies that an optimal policy accepts a type- i request if and only if the expected incremental revenue from the request is at least the expected drop in future revenue from accepting it.

Despite the simplicity of (3), the MDP suffers from the curse of dimensionality, because we still face the problem of storing actions (or values of $g_t(\mathbf{x})$) for each (\mathbf{x}, t) -pair. Except in some special cases, there do not appear to be any particular structural properties for optimal policies that reduce the computation and storage requirements to a practical level. One setting that will allow for a simplified state representation is when V_i and W_i are normally distributed for $i = 1, \dots, m$, in which case the m -dimensional state \mathbf{x} can be replaced by a four-dimensional vector comprised of the means and variances of booked volume and weight. Such a simplification is generally not possible when the family of volume or weight distributions is not closed under convolution.

1.2. Computational Challenges

In this section, we illustrate the difficulties associated with implementing an optimal policy. To use (3), one needs to store either all values of $g_t(\mathbf{x})$ or all actions specified by (3) (there is an action for each (\mathbf{x}, t) -pair that specifies which types are acceptable and which are not). This, however, requires too much computer

memory when the number of decision periods and/or the number of shipment types is large. Note that in period t , we have that $\sum_{i=1}^m x_i \leq t$ since there is at most one booking request within each time period. Hence the number of reachable states is $\sum_{\tau=0}^t \binom{\tau+m-1}{m-1}$, which is the cardinality of the set $\bigcup_{\tau=0}^t \{\mathbf{x} \in \mathbb{Z}_+^m : \sum_{i=1}^m x_i = \tau\}$. If $t = 200$ and $m = 10$, an array of size 3×10^{20} will be needed to store values for all possible states, which will require approximately 10^{11} gigabytes of computer memory. Hence, direct application of an optimal policy to problems of the size typically encountered in industry does not appear feasible.

Before we conclude that there is no practical method for implementing an optimal policy, we need to explore one more avenue. If the problem possesses some special structure, it may not be necessary to store value functions or actions for all possible states. For example, the value function of the MDP for certain single-leg passenger RM problems is concave and nonincreasing in the number of previously accepted reservations (see, e.g., Lautenbacher and Stidham 1999, Theorem 1); therefore, to specify an optimal policy, it is sufficient to store a booking limit (maximum number of seats to be sold) for each time and passenger class. Unfortunately, we will see that the structure of (3) is rather complicated, even for a simplified version of the booking control problem.

Suppose $(V_i, W_i) = (v_i, w_i)$ with probability 1, and there are two shipment types ($m = 2$) with $w_1 > w_2$, $v_1 > v_2$. The revenue function is defined as $\rho_i = r \max\{w_i, v_i/\gamma\}$ for both shipment types. Because (V_i, W_i) is deterministic, true volume and weight are observed at the time of the booking request. Suppose h_v, h_w are large enough that overbooking is not economical; that is, given high overbooking costs, an optimal policy *never* requires any off-loading. Let $x = \sum_{i=1}^2 x_i v_i$ be the total volume accepted, and $y = \sum_{i=1}^2 x_i w_i$ be the total weight accepted. Using these, we can reduce the state from \mathbf{x} to (x, y) . Now, the MDP optimality equations, with corresponding value functions denoted by $U_t(x, y)$, simplify to

$$U_t(x, y) = \sum_{i=1}^2 p_{it} \max\{\rho_i + U_{t-1}(x + v_i, y + w_i), U_{t-1}(x, y)\} + p_{0t} U_{t-1}(x, y),$$

$$U_0(x, y) = 0, \quad \text{if } x \leq k_v, y \leq k_w.$$

Observe that by choosing h_v and h_w large as described above, we do not need to define U_0 for states above capacity. An optimal policy accepts a type- i request at time t given the state is (x, y) if and only if $\rho_i \geq U_{t-1}(x, y) - U_{t-1}(x + v_i, y + w_i)$.

In time period 1, the optimal policy accepts a booking request of type- i for $i = 1, 2$ if there is enough capacity left, i.e., if $(x, y) \in [0, k_v - v_i] \times [0, k_w - w_i]$. In

time period 2, the decision whether to accept or deny a type-1 request remains the same. However, the decision for a type-2 request in time period 2 depends on parameters of the problem. Figure 1 shows the optimal decisions for a type-2 request in time period 2. Note that there are two cases to consider, as reflected in parts 1(a) and 1(b) of the figure. Here, the decision for a type-2 request requires up to six critical numbers to specify; namely, $k_v - v_1$, $k_w - w_1$, $k_v - v_2$, $k_w - w_2$, $k_v - v_1 - v_2$, and $k_w - w_1 - w_2$. In general, if there are m shipment types, then up to $2(2^m - 1)$ critical numbers are required to specify an optimal decision at time period 2. Furthermore, in a problem instance where (volume, weight) is random at the time of booking the request, an optimal policy may include overbooking. The resulting decision rules will be even more complicated, and the associated computational burden even higher.

2. Bounds and Heuristics

As we saw in the previous section, optimal policies are typically too complex to compute or store. In this section, we focus on obtaining bounds and heuristics. In §2.1, we derive upper bounds for the value function of the MDP formulation. Some of these bounds are obtained from computationally tractable recursive functions; these functions are, in fact, value functions of certain lower dimensional MDPs. Also, in §2.2, we develop a computationally feasible heuristic that avoids the curse of dimensionality.

2.1. Bounds

In this section, we use the notation $\bar{v}_i = EV_i$ and $\bar{w}_i = EW_i$ to represent, respectively, the mean volume and mean weight of a type- i shipment.

LEMMA 1. For $\mathbf{x} \in \mathbb{Z}_+^m$, let $x = \sum_{i=1}^m x_i \bar{v}_i$ and $y = \sum_{i=1}^m x_i \bar{w}_i$. Then,

$$g_0(\mathbf{x}) \leq -h_v(x) - h_w(y). \quad (4)$$

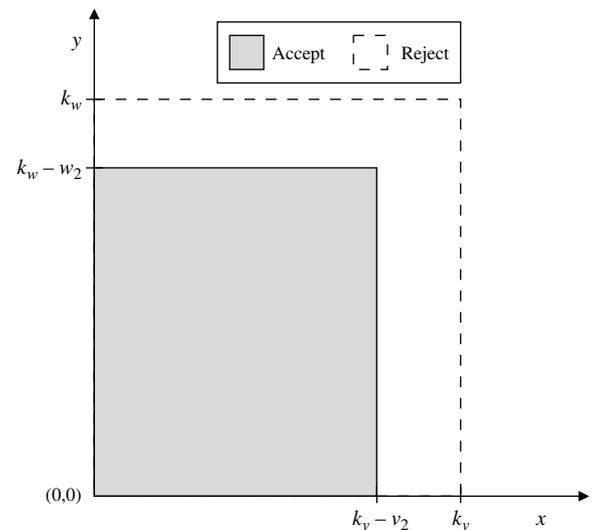
The lemma states that terminal cost $g_0(\mathbf{x})$ is bounded above by the sum of two functions. The first function, $h_v(x)$, depends only on the total expected volume of the \mathbf{x} accepted bookings, and the second, $h_w(y)$, depends only on the total expected weight of the \mathbf{x} accepted bookings.

Suppose $a, b: \mathbb{R}_+ \rightarrow \mathbb{R}$ are functions such that $a(x) \leq h_v(x)$ and $b(y) \leq h_w(y)$ for all $x, y \in \mathbb{R}_+$. Let $\alpha = (\alpha_1, \dots, \alpha_m)$ and $\beta = (\beta_1, \dots, \beta_m)$ be m -vectors such that $\alpha_i + \beta_i = \rho_i$, for $i = 1, \dots, m$. Recursively, define the value function of the *volume problem* as

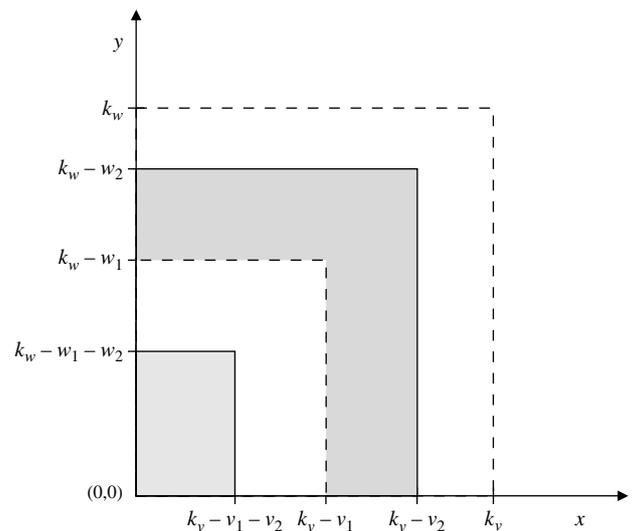
$$\varphi_t^\alpha(x) = \sum_{i=1}^m p_{it} \max\{\alpha_i + \varphi_{t-1}^\alpha(x + \bar{v}_i), \varphi_{t-1}^\alpha(x)\} + p_{0t} \varphi_{t-1}^\alpha(x), \quad (5)$$

$$t = 1, \dots, n,$$

$$\varphi_0^\alpha(x) = -a(x). \quad (6)$$



(a) $\max\{w_2, v_2/\gamma\} - p_{11} \max\{w_1, v_1/\gamma\} \geq 0$



(b) $\max\{w_2, v_2/\gamma\} - p_{11} \max\{w_1, v_1/\gamma\} < 0$

Figure 1 Decisions for a Type-2 Request in Period 2

Note. For (x, y) pairs in the shaded regions, type-2 requests should be accepted in period 2.

Similarly, define the value function of *weight problem* as

$$\psi_t^\beta(y) = \sum_{i=1}^m p_{it} \max\{\beta_i + \psi_{t-1}^\beta(y + \bar{w}_i), \psi_{t-1}^\beta(y)\} + p_{0t} \psi_{t-1}^\beta(y), \quad t = 1, \dots, n, \quad (7)$$

$$\psi_0^\beta(y) = -b(y). \quad (8)$$

The terminology *volume* (respectively, *weight*) problem refers to the fact that the state of the recursion (5)–(6) (resp. (7)–(8)) is a one-dimensional quantity corresponding to volume (resp. weight).

PROPOSITION 1. For $\mathbf{x} \in \mathbb{Z}_+^m$, let $x = \sum_{i=1}^m x_i \bar{v}_i$ and $y = \sum_{i=1}^m x_i \bar{w}_i$. Then,

$$g_t(\mathbf{x}) \leq \varphi_t^\alpha(x) + \psi_t^\beta(y) \quad \text{for each } t = 0, 1, \dots, n.$$

Through different choices of $a(\cdot)$, $b(\cdot)$, α , and β , Proposition 1 allows us to generate a family of upper bounds for the value function $g_t(\mathbf{x})$ of the high-dimensional MDP. Of particular importance is the fact that computation of any such upper bound requires only the solutions of two separate one-dimensional MDPs. The idea of constructing an upper bound for a high-dimensional MDP by using separate sub-problems can be found in other contexts (e.g., White and Schlüssel 1981; Lovejoy 1986; Zhang and Cooper 2005). Below, we consider three particular instances of the bound in the proposition.

For the first of these, observe that the expected revenue from accepting a type- i request can be written as $\rho_i = f_i^w + f_i^v$, where $f_i^w = E[r_i(W_i)]$ and

$$f_i^v = E([r_i(V_i/\gamma) - r_i(W_i)]1\{V_i \geq \gamma W_i\}).$$

(We use $1\{A\}$ to denote an indicator of event A , i.e., $1\{A\} = 1$ if A occurs, and 0 otherwise.) Hence we can apply the proposition with $\alpha_i = f_i^v$, $a(x) = h_v(x)$ and $\beta_i = f_i^w$, $b(y) = h_w(y)$. For these particular choices of α , $a(\cdot)$, β , and $b(\cdot)$, we use $u_t^v(x)$ (resp. $u_t^w(y)$) to denote the solution to (5)–(6) (resp. (7)–(8)). That is, $u_t^v(x) = \varphi_t^\alpha(x)$ with $\alpha_i = f_i^v$ and $a(x) = h_v(x)$, and $u_t^w(y) = \psi_t^\beta(y)$ with $\beta_i = f_i^w$ and $b(y) = h_w(y)$. For clarity, $u_t^v(x)$ satisfies

$$u_t^v(x) = \sum_{i=1}^m p_{it} \max\{f_i^v + u_{t-1}^v(x + \bar{v}_i), u_{t-1}^v(x)\} + p_{0t} u_{t-1}^v(x),$$

$$t = 1, \dots, n, \quad (9)$$

$$u_0^v(x) = -h_v(x), \quad (10)$$

and $u_t^w(y)$ is defined similarly.

Formally, we have the following result, which is an immediate consequence of Proposition 1 and the discussion above.

COROLLARY 1. For $\mathbf{x} \in \mathbb{Z}_+^m$, let $x = \sum_{i=1}^m x_i \bar{v}_i$ and $y = \sum_{i=1}^m x_i \bar{w}_i$. Then,

$$g_t(\mathbf{x}) \leq u_t^v(x) + u_t^w(y) \quad \text{for each } t = 0, 1, \dots, n.$$

It is also of interest to consider two other decompositions. One of these corresponds to solving a fictitious problem in which the weight capacity of the flight is infinite and volume is deterministic ($P(V_i = \bar{v}_i) = 1$, for all i). The other corresponds to solving a fictitious problem in which the volume capacity of the flight is infinite and the weights are deterministic ($P(W_i = \bar{w}_i) = 1$, for all i). For the problem with infinite weight, let $u_t^1(x) = \varphi_t^\alpha(x)$ in (5)–(6) with $\alpha_i = \rho_i$ and $a(x) = h_v(x)$. Similarly, for the problem with infinite volume, let $u_t^2(y) = \psi_t^\beta(y)$ in (7)–(8) with $\beta_i = \rho_i$ and $a(y) = h_w(y)$. Then, we obtain the following corollary.

COROLLARY 2. For $\mathbf{x} \in \mathbb{Z}_+^m$, let $x = \sum_{i=1}^m x_i \bar{v}_i$ and $y = \sum_{i=1}^m x_i \bar{w}_i$. Then,

$$g_t(\mathbf{x}) \leq u_t^1(x) \quad \text{and} \quad g_t(\mathbf{x}) \leq u_t^2(y)$$

for each $t = 0, 1, \dots, n$.

The next upper bound is motivated by the linear programming formulation in the passenger RM context (see, e.g., Cooper 2002; Bertsimas and Popescu 2003). For any fixed sample path, let d_i be the total number of type- i booking requests during the booking horizon. Consider a situation where we need to decide how many of the d_i requests to accept to maximize revenue. Note that we are assuming knowledge of d_i ; $i = 1, \dots, m$, prior to making this decision. Let x_i be the decision variable representing the number of type- i requests to be accepted. The revenue can be approximated by

$$\phi(\mathbf{x}) = \sum_{i=1}^m \rho_i x_i - h_v\left(\sum_{i=1}^m \bar{v}_i x_i\right) - h_w\left(\sum_{i=1}^m \bar{w}_i x_i\right).$$

Let $\zeta: \mathbb{Z}_+^m \rightarrow \mathbb{R}$ defined by

$$\zeta(\mathbf{d}) = \max\{\phi(\mathbf{x}): \mathbf{x} \leq \mathbf{d}, \mathbf{x} \in \mathbb{Z}_+^m\} \quad (11)$$

denote the maximum approximate revenue. Relaxing the requirement that \mathbf{x} and \mathbf{d} be integer vectors in (11), we define $\xi(\mathbf{y}): \mathbb{R}_+^m \rightarrow \mathbb{R}$, where

$$\xi(\mathbf{y}) = \max\{\phi(\mathbf{x}): \mathbf{x} \leq \mathbf{y}, \mathbf{x} \in \mathbb{R}_+^m\}. \quad (12)$$

Let $E[\mathbf{D}] = (E[D_1], \dots, E[D_m])$, where D_i is the random number of type- i booking requests for the entire booking horizon. The following proposition relates the value function of the MDP and math programs (11) and (12). The computationally tractable upper bound we propose is $\xi(E[\mathbf{D}])$.

PROPOSITION 2. $g_n(\mathbf{e}_0) \leq E[\zeta(\mathbf{D})] \leq \xi(E[\mathbf{D}])$.

We will refer to the minimum, denoted by \bar{u} , of these four upper bounds as the best upper bound, i.e.,

$$\bar{u} = \min\{u_n^1(0), u_n^2(0), u_n^v(0) + u_n^w(0), \xi(E[\mathbf{D}])\}.$$

2.2. Heuristics

In this section, we describe the various heuristic booking policies that we will subsequently test in §3. We begin by introducing heuristics obtained by substituting the upper bounds of the previous section in place of $g_t(\cdot)$ in (3).

For instance, we can construct a heuristic based on the upper bound $u_t^v(x) + u_t^w(y)$ as follows: accept a type- i booking request in period t when the state is \mathbf{x} if and only if

$$\rho_i \geq [u_{t-1}^v(x) + u_{t-1}^w(y)] - [u_{t-1}^v(x + \bar{v}_i) + u_{t-1}^w(y + \bar{w}_i)], \quad (13)$$

where $x = \sum_{i=1}^m x_i \bar{v}_i$ is the expected volume consumed by the \mathbf{x} bookings, and $y = \sum_{i=1}^m x_i \bar{w}_i$ is the expected weight consumed by the same bookings. We will use the shorthand *HD* for this heuristic. The “*D*” in *HD* relates to the fact that the rule in (13) approximates the opportunity cost in (3) by decoupling the problem into two separate one-dimensional problems.

Similarly, we can specify heuristics based on upper bounds $u_i^k(\cdot)$; $k = 1, 2$. The heuristic booking policy *H1* (resp. *H2*) is as follows: accept a type- i booking request in period t when the state is \mathbf{x} if and only if $\rho_i \geq u_{i-1}^1(x) - u_{i-1}^1(x + \bar{v}_i)$ (resp. $\rho_i \geq u_{i-1}^2(y) - u_{i-1}^2(y + \bar{w}_i)$). The “1” in *H1* relates to the fact that it approximates the opportunity cost by using upper bound u_i^1 . There is a similar explanation of the “2” in *H2*. We also consider the heuristic booking policy *HM*—“*M*” is for minimum—that accepts a type- i booking request in period t when the state is \mathbf{x} if and only if

$$\begin{aligned} x + \bar{v}_i &\leq k_v, & y + \bar{w}_i &\leq k_w, & \text{and} \\ \rho_i &\geq \min\{u_{i-1}^1(x), u_{i-1}^2(y)\} \\ &\quad - \min\{u_{i-1}^1(x + \bar{v}_i), u_{i-1}^2(y + \bar{w}_i)\}. \end{aligned} \quad (14)$$

In (14), one need not impose the condition involving capacities, but we found that adding the condition improved performance.

It is important to note that the state of the MDP $g_t(\mathbf{x})$ is an m -dimensional vector; whereas each of the recursive functions needed to implement the above booking policies has a scalar domain. In u^v and u^1 (or u^w and u^2), the argument is of the same magnitude as the maximum volume (or weight) available to book. Thus the computer memory required to store $u_i^j(\cdot)$; $j \in \{v, w, 1, 2\}$ is manageable for a realistic problem size. For example, a Boeing 747-400 ER freighter has a maximum cargo capacity of 158.6 cubic meters and a maximum payload of 412,770 kilograms (Boeing Company 2005a). If volume is discretized into cubic centimeters and weight is discretized into grams, then storing two arrays of size 1.6×10^8 and 4.1×10^8 requires approximately 2 gigabytes. This much space is typically available on personal computers. More generally, the storage space and computational effort required for $u_i^j(\cdot)$ are not affected by the number of shipment types, whereas the state space and the associated storage or computational needs for $g_t(\mathbf{x})$ are exponential in the number of types (m).

We also consider heuristics based on the mathematical programming problem $\xi(E[\mathbf{D}])$. Let \mathbf{z}^* be an optimal solution of $\xi(E[\mathbf{D}])$. The partitioned allocation policy *PA* will accept a type- i booking request in period t when the state is \mathbf{x} if and only if

$$\begin{aligned} x + \bar{v}_i &\leq k_v, & y + \bar{w}_i &\leq k_w, & \text{and} & & x_i < \lceil z_i^* \rceil \\ & & & & & & \text{for all } i = 1, \dots, m, \end{aligned} \quad (15)$$

where $\lceil a \rceil$ denotes the ceiling of a , i.e., the least integer that is greater than or equal to a . We interpret $\lceil z_i^* \rceil$ as the booking limit for a type- i shipment.

For another approach based on a math program, consider the linear program

$$\max \left\{ \sum_{i=1}^m \rho_i x_i : 0 \leq \mathbf{x} \leq E[\mathbf{D}], \sum_i \bar{v}_i x_i \leq k_v, \sum_i \bar{w}_i x_i \leq k_w \right\}.$$

Let λ_v (λ_w) be a dual variable corresponding to the volume (weight) constraint. The dual variable can be interpreted as the monetary value per unit of resource or that resource’s shadow price. Similar to Equation (3) in Pak and Dekker (2004), the bid-price policy *BP* accepts a type- i booking request in period t when the state is \mathbf{x} if and only if

$$\begin{aligned} x + \bar{v}_i &\leq k_v, & y + \bar{w}_i &\leq k_w, & \text{and} & & \rho_i \geq \bar{v}_i \lambda_v + \bar{w}_i \lambda_w \\ & & & & & & \text{for all } i = 1, \dots, m. \end{aligned} \quad (16)$$

Finally, the *FCFS* policy will accept a type- i booking request in period t when the state is \mathbf{x} if and only if

$$x + \bar{v}_i \leq k_v \quad \text{and} \quad y + \bar{w}_i \leq k_w. \quad (17)$$

The *FCFS* heuristic does not require computations or demand forecasting (other than mean volumes and weights), and is easy to implement. Regardless of the problem size, the amount of information needed to implement this policy is limited to the total previously committed expected volume and weight. These features make *FCFS* desirable from a practical viewpoint. Note that *FCFS* can perform quite poorly if the unit overage penalties are large. This problem could be mitigated, to a certain extent, by using carefully chosen booking limits, instead of k_v and k_w in (17). However, the calculation of such limits would require more computational effort and forecasting of aggregate volume and weight distributions.

As we will see in the next section, *HD* performed the best among the above heuristics in our numerical examples. On an intuitive level, *HD* provides a good decision rule, because the expected drop in future revenue caused by accepting a type- i shipment is approximated reasonably well by the difference,

$$[u_{i-1}^v(x) + u_{i-1}^w(y)] - [u_{i-1}^v(x + \bar{v}_i) + u_{i-1}^w(y + \bar{w}_i)].$$

3. Computational Results and Insights

In this section, we describe a series of simulation experiments to test the heuristics from §2. We describe the choice of parameters for the numerical experiments in §3.1. We illustrate the performance of heuristics for an industry-sized problem in §3.2 and study conditions that affect their performance in §3.3.

Table 1 Weight and Expected Volume for Shipment Category

Category	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Weight (kg)	50	50	50	50	100	100	100	100	200	200	200	250	250	300	400	500	1,000	1,500	2,500	3,500	70	70	210	210
Mean volume ($\times 10^4$ cm ³)	30	29	27	25	59	58	55	52	125	119	100	147	138	179	235	277	598	898	1,488	2,083	233	17	700	52

3.1. Parameters for Numerical Examples

In all of the numerical examples, we assume that volume is random, but weight is deterministic. In practice, it is common for volume to be unknown at the time of booking requests, because obtaining a precise volume measurement is time consuming and may require an expensive CubiScan device, which is often not available to the shipper (Slager and Kapteijns 2004; Beidermand 2002). In contrast, weighing devices are more readily available. We assume that the terminal value function is given by (2) with $h_v(v_T) = h_v \cdot (v_T - k_v)^+$ and $h_w(w_T) = h_w \cdot (w_T - k_w)^+$.

We use the following scheme to determine the shipment type. Suppose there are g different (weight, mean volume) pairs. Each of these pairs is called a *category*. In addition, each shipment belongs to one of s *classes*. A class- l shipment generates revenue according to a revenue function $r_l(\hat{w})$ that maps its chargeable weight $\hat{w} = \max\{w, v/\gamma\}$ to revenue. Different revenue functions represent shipments that contain different types of goods or that arrive at different points in time. We define a shipment *type* by a pair of class and category. Then, the number of shipment types is $m = gs$. Note that under this scheme, two otherwise identical shipments that contain different types of cargo or that are booked at different times may generate different revenues.

We assume that the volume of a category- k request follows a lognormal distribution with mean μ_k and variance $(\theta\mu_k)^2$, where θ is the coefficient of variation for every shipment category. In all examples, there are $g = 24$ shipment categories; the corresponding weight, mean volume pairs are shown in Table 1. The weights range from 50 to 3,500 kilograms. Categories 1–20 have densities that vary from 160 kilograms per cubic meter to 200 kilograms per cubic meter, reflecting normal cargo (IATA standard is 167 kilograms per cubic meter). However, we also allow some extremes in categories 21–24. These have densities as low as 30 kilograms per cubic meter and as high as 400 kilograms per cubic meter. Assume that a shipment belongs to category k with probability 0.072 for $k = 1, \dots, 10$; 0.04 for $k = 11, \dots, 16$; 0.009 for $k = 17, \dots, 20$; and 0.001 for $k = 21, \dots, 24$. With this choice of parameters, the differences between the dimensional and the physical weights can vary significantly, and the percentage of shipments that are charged by dimensional weights can be large. For example, when the coefficient of variation is 0.2,

about 37% of all shipments have dimensional weights greater than physical weights, and among these shipments, dimensional weights are typically larger by 3%–50% of actual weights.

We assume that there are $s = 10$ shipment classes, and that the revenue function is piecewise linear (with three kinks) in chargeable weight \hat{w} . Table 2 shows a revenue per unit chargeable weight for each shipment class, e.g., a class-1 shipment whose chargeable weight is 50 kilograms brings a revenue of $50 \times 1.12 = \$56$. These revenues vary quite significantly among different classes; the range can be more than 100%. Also, note that the revenue per unit chargeable weight is nonincreasing in the chargeable weight.

There are 60 decision periods, and Table 3 shows probabilities of a booking request for each shipment class. We index the $m = 240$ ($gs = 24 \times 10$) types, so that a shipment of class l and category k is labeled as type $i = g(l - 1) + k$. The probability of a request for a particular shipment type is the product of the probability of a booking request for the corresponding class and the probability of a request for the corresponding category. For example, the probability of a type-1 booking request in each time period $t = 1, \dots, 10$ is $(0.02)(0.072) = 0.00144$, and $(0.03)(0.072) = 0.00216$ in each time period $t = 11, \dots, 20$. These parameters are such that the expected total number of booking requests is 22.61. The total expected weight and volume demand are 5,088.61 kilograms, and 2.97×10^7 cubic centimeters, values which represent 26% of structural limit payload or 21% of total cargo volume (main deck and lower hold bulk volume) of a Boeing 727-100 airplane (Boeing Company 2005). We do not model the entire plane capacity, because a large proportion of cargo space may be prepurchased at the beginning of the season by freight forwarders, and some may be reserved for mail.

Other parameters; namely, the coefficient of variation for the volume distribution, the volume and weight capacity, the penalty for overage, and the number of decision periods, are specified for each

Table 2 Revenue Function for Shipment Class

Class	1	2	3	4	5	6	7	8	9	10
$0 < \hat{w} \leq 90$	1.12	1.04	0.92	0.82	0.8	0.87	0.99	0.72	0.7	0.55
$90 < \hat{w} \leq 990$	1.11	1.03	0.91	0.81	0.79	0.86	0.98	0.71	0.69	0.54
$990 < \hat{w} \leq 1,990$	1.09	1.01	0.89	0.79	0.77	0.84	0.96	0.69	0.67	0.52
$1,990 < \hat{w}$	1.08	1.0	0.88	0.78	0.76	0.83	0.95	0.68	0.66	0.51

Table 3 Request Probabilities for Shipment Class

Periods	1-10	11-20	21-30	31-40	41-50	51-60
Class 1	0.02	0.03	0.04	0.04	0.05	0.05
Class 2	0.006	0.007	0.01	0.01	0.015	0.02
Class 3	0.005	0.005	0.05	0.07	0.065	0.08
Class 4	0.02	0.02	0.02	0.045	0.045	0.07
Class 5	0.025	0.025	0.025	0.025	0.025	0.03
Class 6	0.03	0.02	0.03	0.02	0.03	0.04
Class 7	0.05	0.05	0.05	0.05	0.05	0.06
Class 8	0.078	0.06	0.07	0.06	0.07	0.09
Class 9	0.03	0.035	0.04	0.045	0.05	0.055
Class 10	0.001	0.045	0.002	0.002	0.05	0.05

example. The volume and weight capacity are specified as fractions of total expected demand. The total expected volume demand for the entire booking horizon is $d_v = \sum_{i=1}^n \sum_{j=1}^m p_{ij} \mu_j$, and the total expected weight demand is $d_w = \sum_{i=1}^n \sum_{j=1}^m p_{ij} w_j$. The capacity-demand ratio $(k_v/d_v, k_w/d_w) = (1.0, 1.0)$ represents a situation when the demand and the capacity are well matched. To specify the penalty cost rates (h_v, h_w) , let (η_v, η_w) denote a benchmark penalty cost defined by $\sum_{i=1}^n \sum_{j=1}^m p_{ij} \rho_j = \eta_v d_v = \eta_w d_w$. The benchmark volume penalty cost rate is the total expected revenue per unit expected volume demand from all booking requests. The penalty cost rates at which $(h_v/\eta_v, h_w/\eta_w) = (1.0, 1.0)$ represent a situation where both penalty costs are at the benchmark.

Simulation is required to estimate the expected revenue of each heuristic, because exact calculation of the expected revenue is computationally impractical owing to the size of the state space. To estimate the expected revenue under each heuristic policy, we simulated arrivals, which were then accepted or rejected based on the heuristic's decision rule. We kept track of a vector of the number of each type accepted, and added the terminal cost at the end of the booking horizon. For each setting of parameters, we ran multiple replications until all the lengths of the 95% confidence intervals for the expected revenues under all heuristics are within 1% of their estimated values. Specifically, let $Z_\pi^{(i)}$ denote the revenue under policy π in the i -th simulation replication. Let \bar{Z}_π^q denote the sample mean of the set $\{Z_\pi^{(1)}, \dots, Z_\pi^{(q)}\}$, i.e., $\bar{Z}_\pi^q = \sum_{i=1}^q Z_\pi^{(i)} / q$, and let S_π^q denote the sample standard deviation, i.e., $S_\pi^q = (\sum_{i=1}^q (Z_\pi^{(i)} - \bar{Z}_\pi^q)^2 / (q - 1))^{1/2}$. The number of simulation replications is

$$q = \min \left\{ d \in \mathbb{Z}_+ : 2 \left(1.96 \frac{S_\pi^d}{\sqrt{d}} \right) \leq 10^{-2} \bar{Z}_\pi^d \text{ for all } \pi \in \{HD, H1, H2, HM, PA, BP, FCFS\} \right\}.$$

In addition to conducting the simulations, we also computed the best upper bound. The functions u^v and u^w as well as u^1 and u^2 were computed and

stored prior to the simulation. (Consequently, the accept or reject decisions could be made in real time during the simulation.) In the numerical example, volume is discretized into units of 10^4 cubic centimeters, and weight is discretized into units of 10 kilograms. The domain of the accepted volume (in 10^4 cubic centimeters) is $\{0, 1, \dots, 5940\}$, and that of the accepted weight (in 10 kilograms) is $\{0, 1, \dots, 1020\}$. The offline computational times on an Intel Pentium III 1.2 GHz processor (0.5 GB RAM, Linux 8.0) for $u^v + u^w$, u^1 , and u^2 were always less than 30 seconds.

3.2. Quality of the Heuristics

EXAMPLE 1 (QUALITY OF THE HEURISTICS). The volume and weight capacity, the coefficient of variation for the volume distribution, and the penalty for overage were varied systematically according to a $7 \times 2 \times 5$ factorial experiment. The first factor is the capacity-demand ratio $(k_v/d_v, k_w/d_w)$ and has 7 levels $C_1 = \{(1.0, 1.0), (1.1, 1.0), (1.0, 1.1), (1.1, 1.1), (0.9, 1.0), (1.0, 0.9), (0.9, 0.9)\}$. The second factor in the experiment is the coefficient of variation θ and has two levels $C_2 = \{0.2, 0.8\}$. The third factor is the penalty cost rate ratio $(h_v/\eta_v, h_w/\eta_w)$ and has five levels $C_3 = \{(0.8, 0.8), (1.2, 0.8), (0.8, 1.2), (1.2, 1.2), (1.0, 1.0)\}$. Let $C = C_1 \times C_2 \times C_3$ denote the experimental region.

Over the 70 problems, the average, minimum, and maximum percent differences between the estimated expected revenue from each heuristic and the best upper bound are shown in Table 4. Specifically, let \bar{u}_c be the best upper bound when the factor-level combination is $c \in C$. Let $Z_{\pi,c}^{(i)}$ denote the revenue under policy π in the i th simulation replication when the factor-level combination is c , and $\bar{Z}_{\pi,c}^{q_c}$ denote the sample mean of $\{Z_{\pi,c}^{(1)}, \dots, Z_{\pi,c}^{(q_c)}\}$, where q_c is the number of simulation replications. We use $\Delta_{\pi,c}$ to denote the difference between the estimated expected revenue under policy π and the best upper bound as a percentage of the bound when the factor-level combination is c . That is,

$$\Delta_{\pi,c} = \frac{\bar{u}_c - \bar{Z}_{\pi,c}^{q_c}}{\bar{u}_c} \times 100.$$

Table 4 shows, for each policy π , the average, minimum, and maximum of $\{\Delta_{\pi,c} : c \in C\}$. Observe that the average, minimum, and maximum of HD are the smallest among those of all the heuristics.

The average, minimum, and maximum of the estimated coefficient of variation of the expected revenue from each heuristic are shown in Table 5. Specifically, let $S_{\pi,c}^{q_c}$ denote the sample standard deviation of $\{Z_{\pi,c}^{(1)}, \dots, Z_{\pi,c}^{(q_c)}\}$. We use $\Gamma_{\pi,c}$ to denote the estimated coefficient of variation of the expected revenue under policy π when the factor-level combination is c , i.e.,

Table 4 Percent Difference from \bar{u} for Example 1

	<i>HD</i>	<i>H1</i>	<i>H2</i>	<i>HM</i>	<i>PA</i>	<i>BP</i>	<i>FCFS</i>
Average	6.04	11.25	11.02	8.08	12.63	9.36	8.89
Minimum	1.63	2.89	2.61	3.76	8.34	4.49	4.49
Maximum	11.47	23.01	24.41	13.06	17.22	14.52	13.66

$\Gamma_{\pi,c} = S_{\pi,c}^{q_c} / \bar{Z}_{\pi,c}^{q_c}$. Table 5 shows, for each policy π , the average, minimum, and maximum of $\{\Gamma_{\pi,c}: c \in C\}$. From Table 5, the average, minimum, and maximum from *HD* and *HM* are smaller than those of the other heuristics. This suggests that *HD* and *HM* are more robust than the others.

Table 6 shows the average off-loaded volume and weight as percentages of the available volume and weight capacity, when the capacity-demand ratio is $(k_v/d_v, k_w/d_w) = (1.0, 1.0)$, the penalty cost rate ratio is $(h_v/\eta_v) = (h_w/\eta_w) = (h/\eta)$, and the coefficient of variation of the volume distribution is θ . Specifically, in the policy- π column of the table, v^0 represents the average over all replications (for the given parameter settings) of $100 \times (v_{\pi}^{(i)} - k_v)^+ / k_v$, where $v_{\pi}^{(i)}$ is the total booked volume on replication i for policy π ; w^0 is defined similarly for weight. Note that there is no off-loaded weight for *HM*, *PA*, *BP*, and *FCFS*, because weights are assumed be deterministic (see (14)–(17)).

The highest values of v^0 and w^0 correspond to *H1* or *H2*, which assume either infinite payload or infinite volume capacity. This suggests that *H1* or *H2* are the most aggressive booking policies among those considered. On the other hand, *HD* has moderate values of v^0 and w^0 , suggesting that *HD*, which is the closest to the best upper bound (see Table 4), is neither too aggressive nor too conservative. Also, note that for all heuristics except *BP* and *FCFS*, the average percent off-loaded volume when the penalty cost ratio is 0.8 is at least that when the penalty cost ratio is 1.0. This suggests that a higher penalty cost induces a more conservative booking policy, as previously mentioned in §1. On the other hand, the penalty cost ratio does not affect *BP* and *FCFS* (see (16) and (17), respectively). Finally, note that on average, a higher percentage of volume is off-loaded when $\theta = 0.8$ than when $\theta = 0.2$, suggesting perhaps not surprisingly that higher volume variability leads to more volume off-loading.

Table 5 Estimated Coefficient of Variation for Example 1

	<i>HD</i>	<i>H1</i>	<i>H2</i>	<i>HM</i>	<i>PA</i>	<i>BP</i>	<i>FCFS</i>
Average	0.213	0.242	0.246	0.212	0.237	0.226	0.221
Minimum	0.174	0.185	0.198	0.176	0.204	0.195	0.194
Maximum	0.256	0.380	0.379	0.253	0.276	0.258	0.258

3.3. Insights

In Example 2 below, we expand the analysis for the data set in Example 1 to show the effect of the capacity-demand ratio on the performance of heuristics. We consider the effect of the volume uncertainty in Example 3, and the effect of the number of decision periods in Example 4.

EXAMPLE 2 (EFFECT OF CAPACITY-DEMAND RATIO). We varied capacity-demand ratios within the set $\{0.1, 0.2, \dots, 1.0, 1.1\}$, and fixed $\theta = 0.1$ and $(h_v/\eta_v, h_w/\eta_w) = (1.0, 1.0)$. We performed two tests. In the first test, we fixed $k_w/d_w = 1.0$ and varied k_v/d_v ; and in the second test, we fixed $k_v/d_v = 1.0$ and varied k_w/d_w . The percent difference between the best upper bound and the estimated expected revenue for each policy is shown in Figures 2(a) and 2(b).

The figures show that when capacity is at least expected total demand (that is, when $(k_v/d_v, k_w/d_w) \in \{(1.0, 1.1), (1.1, 1.0)\}$), all the heuristics perform quite well. In such situations, the decision to accept or deny a booking request is relatively easy: accepting almost all booking requests will usually incur little off-loading penalty when the volume and weight capacities are larger than the expected total demand.

Heuristic *HD* does better than the other heuristics in almost all cases, although it is outperformed by *H1* when the volume capacity-demand ratio is very small (namely, $k_v/d_v \in [0.1, 0.5]$). When the volume capacity is extremely tight, the volume constraint is often reached before the payload constraint. In these cases, *H1*, which assumes infinite weight capacity, yields higher revenue because weight (being deterministic in our numerical example) rarely becomes a bottleneck. However, when the volume capacity-demand ratio is moderate (namely, $k_v/d_v \in [0.6, 1.1]$), *HD* outperforms *H1*.

Note that the percent difference between the best upper bound and the estimated expected revenue under *HD* is smaller when the volume capacity is balanced with demand ($k_v/d_v = 1.0$ Figure 2(b)), compared to the case when the weight capacity is balanced with demand ($k_w/d_w = 1.0$ Figure 2(a)). Recall that the weight of each shipment type is known at the time of the booking request, whereas the volume is not realized until shortly prior to the scheduled departure time. This suggests a capacity shortage and uncertain resource requirements may interact to cause relatively worse performance for *HD* in relation to the best upper bound.

When $k_w/d_w = 1.0$ and $k_v/d_v \in [0.1, 0.8]$, the best three heuristics are *HD*, *H1*, and *HM*; whereas when $k_v/d_v = 1.0$ and $k_w/d_w \in [0.1, 0.8]$, the best three are *HD*, *H2*, and *HM*. *H2* performs worse than *H1* when the volume capacity is low. This seems to follow from the fact that *H2* is based on u^2 , which assumes infinite volume capacity. Similarly, *H1* (based on u^1 , which

Table 6 Average Percentage Volume and Weight Off-Loaded for Example 1

(h/η)	θ	HD		H1		H2		HM		PA		BP		FCFS	
		v^0	w^0												
0.8	0.2	2.5	1.4	13.6	13.3	13.5	13.4	1.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
1.0	0.2	1.8	0.8	7.5	7.0	7.6	7.0	0.9	0.0	1.0	0.0	1.0	0.0	1.0	0.0
0.8	0.8	6.9	1.4	15.5	13.3	15.2	13.2	4.7	0.0	4.8	0.0	4.5	0.0	4.5	0.0
1.0	0.8	5.7	0.7	11.1	7.0	10.6	6.9	4.5	0.0	4.2	0.0	4.8	0.0	4.9	0.0

assumes infinite payload) performs worse than *H2* when weight capacity is low.

EXAMPLE 3 (EFFECT OF VARIABILITY). To further understand the effect of volume variability, we varied the coefficient of variation θ among the values $\{0.1, 0.2, \dots, 0.9\}$, and we fixed $(h_v/\eta_v, h_w/\eta_w) = (1.0, 1.0)$. Figure 3 shows the percent difference between the best upper bound and the estimated expected revenue from each heuristic for balanced capacity $(k_v/d_v, k_w/d_w) = (1.0, 1.0)$. Figures for tight volume $(k_v/d_v, k_w/d_w) = (0.9, 1.0)$ and tight payload $(k_v/d_v, k_w/d_w) = (1.0, 0.9)$ are not shown, because they are similar to that for balanced capacity.

The percent difference between the best upper bound and the estimated expected revenue from each heuristic increases as the coefficient of variation increases. Also, consistent with the results in Example 1, *HD* outperforms the other heuristics. Our results suggest that airlines could improve their revenue by reducing volume uncertainty, for example, by introducing standard shipment sizes. To decide whether to carry out this recommendation, a carrier would also need to study the broader effect of shipment size standardization. For example, a sender might switch to other carriers offering flexible shipment sizes, offsetting the benefits of reduced volume uncertainty.

EXAMPLE 4 (NUMBER OF DECISION PERIODS). In this example, we fix $(h_v/\eta_v, h_w/\eta_w) = (1.0, 1.0)$, $(k_v/d_v, k_w/d_w) = (1.0, 1.0)$, and $\theta = 0.2$, and vary the number of decision periods among $\{24, 36, \dots, 72\}$. The probabilities of a booking request for each shipment class are shown in Table 3 except that the decision periods associated with the a th column from the left are numbered as $(an/6 + 1) - (a + 1)n/6$, where n is the number of decision periods, for $a = 0, 1, \dots, 5$. For example, the decision periods when $n = 72$ are $(1-12, 13-24, \dots, 61-72)$.

Figure 4 shows the percent difference between the best upper bound and the estimated expected revenue from each heuristic. The percent difference decreases when the problem size increases. In fact, when the problem size increases, the upper bound on the expected revenue tends to increase by approximately the same rate as the problem size (not shown). On the other hand, the uncertainty of aggregated

demand tends to increase at a slower rate. For discussion of these ideas in the passenger RM context, see §3.6 of Talluri and van Ryzin (2004). From a practical standpoint, this suggests that the heuristic generally performs quite well for larger problem instances. Also, consistent with the results in Example 1, here we again see that *HD* outperforms the other heuristics.

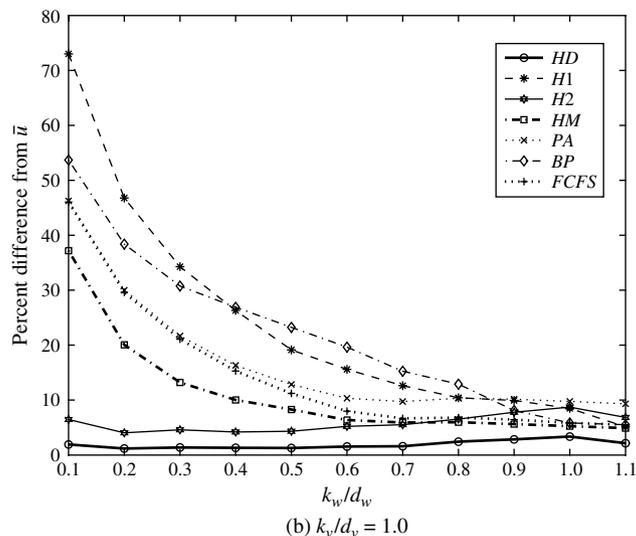
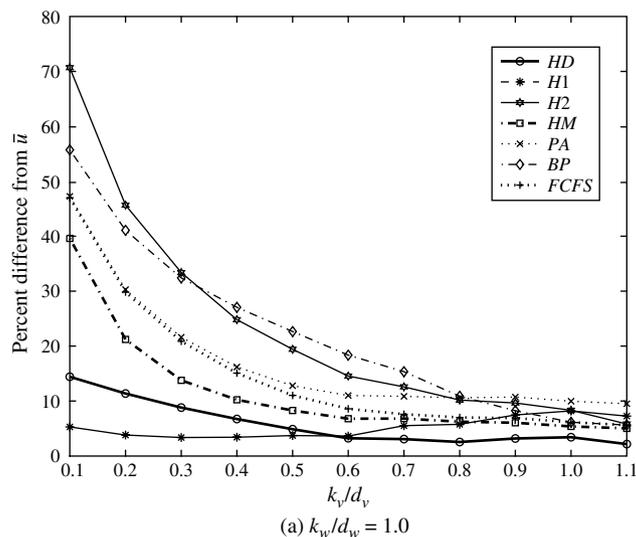


Figure 2 Percent Difference from \bar{u} vs. Capacity-Demand Ratio for Example 2

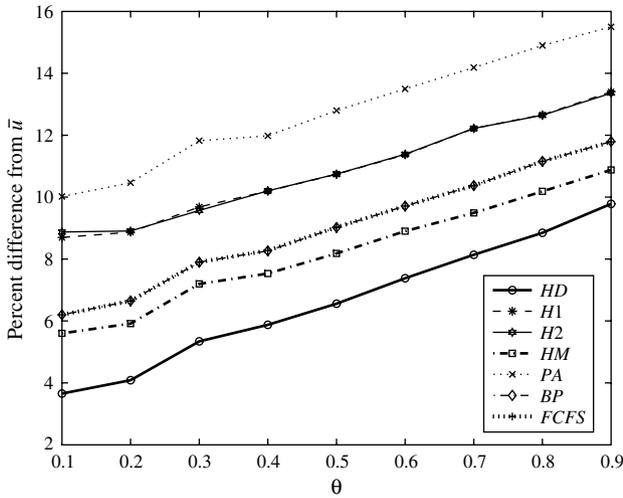


Figure 3 Percent Difference from \bar{u} vs. θ for Example 3

4. Conclusions and Extensions

In this paper, we presented an MDP model for single-leg air-cargo revenue management. Because the MDP has a high-dimensional state space, it is not practical to compute an optimal policy, and so it is important to develop implementable heuristics. We achieve this by making booking decisions based on an approximation of the value function that involves the sum of value functions of two related computationally tractable MDPs, each with a one-dimensional state space. In a series of simulation experiments, we tested these and other heuristics. We found that a particular one—which we call *HD*—performed consistently better than the others, and also yielded expected revenue close to an upper bound on the optimal expected revenue.

Our MDP formulation can be extended easily to include the possibility of no-show shipments. We

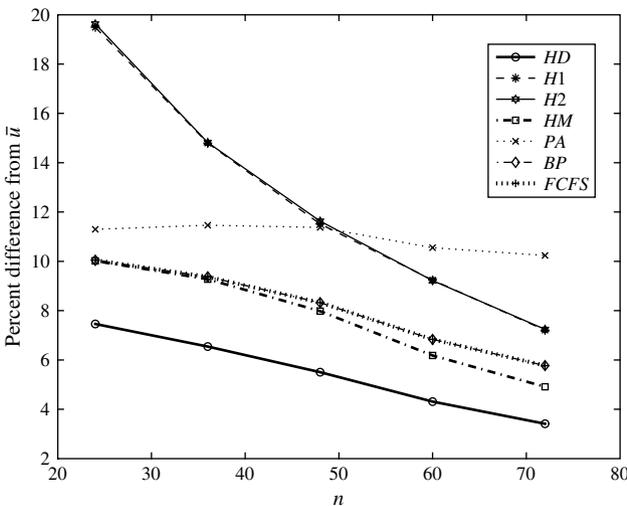


Figure 4 Percent Difference from \bar{u} vs. n for Example 4

need only modify the terminal value function by specifying a different joint distribution of weight and volume for each type of booked shipment. For instance, no-shows can be included by assigning positive probability to (volume, weight) = (0, 0). Another interesting feature of some air-cargo RM problems is that cargo capacity may itself be random, owing to the random weight and volume consumed by passengers and their baggage. Again, this can be integrated into our formulation by changing the terminal value function.

More generally, there are a host of other interesting and unexplored problems in air-cargo RM. For instance, how can a carrier most profitably enter into a contract with freight forwarders, or how can a carrier integrate cargo-booking decisions with cargo routing considerations? We hope to explore these and other related questions in the future.

Acknowledgments

This material is based on work supported by the National Science Foundation under Grant DMI-0450359. The authors are grateful to two anonymous referees and the associate editor for constructive comments on an earlier version of this paper.

Appendix

PROOF OF LEMMA 1. Observe that $h_v(z)$ is convex. From Jensen’s inequality and the fact that $EV_{il} = \bar{v}_i$, it follows that

$$E \left[h_v \left(\sum_{i=1}^m \sum_{l=1}^{x_i} V_{il} \right) \right] \geq h_v \left(E \left[\sum_{i=1}^m \sum_{l=1}^{x_i} V_{il} \right] \right) = h_v \left(\sum_{i=1}^m x_i \bar{v}_i \right).$$

Similarly, we have

$$E \left[h_w \left(\sum_{i=1}^m \sum_{l=1}^{x_i} W_{il} \right) \right] \geq h_w \left(\sum_{i=1}^m x_i \bar{w}_i \right).$$

Hence

$$\begin{aligned} g_0(\mathbf{x}) &= -E \left[h_v \left(\sum_{i=1}^m \sum_{l=1}^{x_i} V_{il} \right) + h_w \left(\sum_{i=1}^m \sum_{l=1}^{x_i} W_{il} \right) \right] \\ &\leq -h_v \left(\sum_{i=1}^m x_i \bar{v}_i \right) - h_w \left(\sum_{i=1}^m x_i \bar{w}_i \right) = -h_v(\mathbf{x}) - h_w(\mathbf{y}). \quad \square \end{aligned}$$

PROOF OF PROPOSITION 1. The proof is based on induction. By Lemma 1 and the definitions of $\varphi_0^\alpha(x)$, $\psi_0^\beta(y)$, we have

$$g_0(\mathbf{x}) \leq -h_v(\mathbf{x}) - h_w(\mathbf{y}) \leq -a(\mathbf{x}) - b(\mathbf{x}) = \varphi_0^\alpha(\mathbf{x}) + \psi_0^\beta(\mathbf{y}).$$

Assume the proposition holds for $k = 1, \dots, t - 1$. Then, we have

$$\begin{aligned} g_t(\mathbf{x}) &= \sum_{i=1}^m p_{it} \max\{\alpha_i + \beta_i + g_{t-1}(\mathbf{x} + \mathbf{e}_i), g_{t-1}(\mathbf{x})\} + p_{0t} g_{t-1}(\mathbf{x}) \\ &\leq \sum_{i=1}^m p_{it} \max\{\alpha_i + \beta_i + \varphi_{t-1}^\alpha(x + \bar{v}_i) + \psi_{t-1}^\beta(y + \bar{w}_i), \\ &\quad \varphi_{t-1}^\alpha(x) + \psi_{t-1}^\beta(y)\} + p_{0t} [\varphi_{t-1}^\alpha(x) + \psi_{t-1}^\beta(y)] \end{aligned} \quad (18)$$

$$\begin{aligned} &\leq \sum_{i=1}^m p_{it} [\max\{\alpha_i + \varphi_{t-1}^\alpha(x + \bar{v}_i), \varphi_{t-1}^\alpha(x)\} \\ &\quad + \max\{\beta_i + \psi_{t-1}^\beta(y + \bar{w}_i), \psi_{t-1}^\beta(y)\}] \\ &\quad + p_{0t} [\varphi_{t-1}^\alpha(x) + \psi_{t-1}^\beta(y)]. \end{aligned} \quad (19)$$

Inequality (18) follows from the induction hypothesis, and inequality (19) follows from the fact that a maximum is a subadditive function, i.e., $\max\{a + b, c + d\} \leq \max\{a, c\} + \max\{b, d\}$ for any real numbers a, b, c, d . \square

PROOF OF COROLLARY 2. For $k = 1$, let $\alpha_i = \rho_i$, $a(x) = h_v(x)$, and $\beta_i = 0$, $b(y) = 0$. The result follows immediately from Proposition 1.

Similarly, for $k = 2$, take $\alpha_i = 0$, $a(x) = 0$, and $\beta_i = \rho_i$, $b(y) = h_w(y)$. Again, the result follows from Proposition 1. \square

PROOF OF PROPOSITION 2. For each sample path, the vector of numbers of accepted bookings under an optimal policy for (1)–(2) is a feasible solution of the maximization in (11); therefore, upon taking expectations, we have $u_n(\mathbf{e}_0) \leq E[\zeta(\mathbf{D})]$. Since $\xi(\mathbf{d})$ is a linear relaxation of $\zeta(\mathbf{d})$, we have $\zeta(\mathbf{d}) \leq \xi(\mathbf{d})$ for any \mathbf{d} . Then, $E[\zeta(\mathbf{D})] \leq E[\xi(\mathbf{D})]$. Furthermore, observe that in (12), the set $C = \{\mathbf{x}, \mathbf{y}: \mathbf{x} \leq \mathbf{y}, \mathbf{x} \in \mathbb{R}_+^m, \mathbf{y} \in \mathbb{R}_+^m\}$ is a convex set, and the objective function $\phi(\mathbf{x})$ is a concave function on C . Then, $\xi(\mathbf{y})$ is a concave function (see Heyman and Sobel 1984, p. 525). By Jensen's inequality, we have $E[\xi(\mathbf{D})] \leq \xi(E[\mathbf{D}])$. \square

References

- Bartodziej, P., U. Derigs. 2004. On an experimental algorithm for revenue management for cargo airlines. *Experimental and Efficient Algorithms: Lecture Notes in Computer Science*, Vol. 3059. Springer-Verlag, Berlin, Germany, 57–71.
- Beidermand, D. 2002. New freight dimensions: For shippers, finding the real cost of a shipment shouldn't be a matter of weight and see. *Air Cargo World* 92(8) 34–40.
- Bertsimas, D. J., I. Popescu. 2003. Revenue management in a dynamic network environment. *Transportation Sci.* 37 257–277.
- Billings, J., A. Diener, B. Yuen. 2003. Cargo revenue optimisation. *J. Revenue and Pricing Management* 2 69–79.
- Bodendorf, F., S. Reinheimer. 1998. Negotiation support for a Web-based air cargo market. INFORMS National Meeting, Montreal, Canada.
- Boeing Company. 2005. 2003 *Freighter Reference Guide*. <http://www.tiaca.org/cm/G2/I138/S8/>.
- Boeing Company. 2005a. Boeing 747-400 freighter. <http://www.boeing.com/commercial/747family/freighter/index.html>.
- Boeing Company. 2005b. World Air Cargo Forecast 2002–2003. http://www.boeing.com/commercial/cargo/exec_summary.html.
- Cooper, W. L. 2002. Asymptotic behavior of an allocation policy for revenue management. *Oper. Res.* 50 720–727.
- DeLain, L., E. O'Meara. 2004. Building a business case for revenue management. *J. Revenue Pricing Management* 2 368–377.
- Heyman, D. P., M. J. Sobel. 1984. *Stochastic Models in Operations Research: Volume II*. McGraw-Hill Book Company, New York.
- Karaesmen, I., G. van Ryzin. 2004. Overbooking with substitutable inventory classes. *Oper. Res.* 52 83–104.
- Kasilingam, R. 1996. Air cargo revenue management: Characteristics and complexities. *Eur. J. Oper. Res.* 96 36–44.
- Kleywegt, A. J., J. D. Papatavrou. 2001. The dynamic and stochastic knapsack problem with random sized items. *Oper. Res.* 49 26–41.
- Lautenbacher, C. J., S. Stidham. 1999. The underlying Markov decision process in the single-leg airline yield management problem. *Transportation Sci.* 33 136–146.
- Lee, T. C., M. Hersh. 1993. A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Sci.* 27 252–265.
- Lovejoy, W. S. 1986. Policy bounds for Markov decision processes. *Oper. Res.* 34 630–637.
- Luo, S., M. Çakanyildirim. 2005. Two-dimensional cargo overbooking models. Working paper, School of Management, University of Texas at Dallas, Richardson, TX.
- Moussawi, L., M. Çakanyildirim. 2005. Profit maximization in air cargo overbooking. Working paper, School of Management, University of Texas at Dallas, Richardson, TX.
- Pak, K., R. Dekker. 2004. Cargo revenue management: Bid-prices for a 0-1 multi knapsack problem. *ERIM Report Series Research in Management*. Rotterdam School of Management, Erasmus Universiteit Rotterdam, The Netherlands.
- Powell, W. B., P. Jaillet, A. Odoni. 1995. Stochastic and dynamic networks and routing. M. O. Ball, T. L. Magnanti, C. L. Monma, G. Nemhauser, eds. *Handbook in Operations Research and Management Science, Volume on Networks*. North-Holland, Amsterdam, The Netherlands.
- Prior, R. C., R. L. Slavens, J. Trimarco. 2004. Menlo worldwide forwarding optimizes its network routing. *Interfaces* 34 26–38.
- Slager, B., L. Kapteijns. 2004. Implementation of cargo revenue management at KLM. *J. Revenue Pricing Management* 3 80–90.
- Subramanian, J., S. Stidham, C. Lautenbacher. 1999. Airline yield management with overbooking, cancellations, and no-shows. *Transportation Sci.* 33 147–167.
- Talluri, K., G. van Ryzin. 2004. *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, MA.
- Thuermer, K. E. 2005. Air cargo pricing: No simple matter. *Logist. Management* 44 36–42.
- Toth, P., D. Vigo. 2002. *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, PA.
- Van Slyke, R., Y. Young. 2000. Finite horizon stochastic knapsacks with applications to yield management. *Oper. Res.* 48 155–172.
- White, C. C., III, K. Schlüssel. 1981. Suboptimal design for large scale, multimodule systems. *Oper. Res.* 29 865–875.
- Zhang, D., W. L. Cooper. 2005. Revenue management for parallel flights with customer choice behavior. *Oper. Res.* 53 415–431.
- Zhao, W., Y. Zheng. 2001. A dynamic model for airline seat allocation with passenger diversion and no-shows. *Transportation Sci.* 35 80–98.