

A MULTIPRODUCT DYNAMIC PRICING PROBLEM AND ITS APPLICATIONS TO NETWORK YIELD MANAGEMENT

GUILLERMO GALLEGO and GARRETT VAN RYZIN

Columbia University, New York

(Received May 1993; revisions received October 1993, November 1993; accepted September 1994)

A firm has inventories of a set of components that are used to produce a set of products. There is a finite horizon over which the firm can sell its products. Demand for each product is a stochastic point process with an intensity that is a function of the vector of prices for the products and the time at which these prices are offered. The problem is to price the finished products so as to maximize total expected revenue over the finite sales horizon. An upper bound on the optimal expected revenue is established by analyzing a deterministic version of the problem. The solution to the deterministic problem suggests two heuristics for the stochastic problem that are shown to be asymptotically optimal as the expected sales volume tends to infinity. Several applications of the model to network yield management are given. Numerical examples illustrate both the range of problems that can be modeled under this framework and the effectiveness of the proposed heuristics. The results provide several fundamental insights into the performance of yield management systems.

Yield management is the practice of using booking policies together with information systems data to increase revenues by intelligently matching capacity with demand. It is now widely practiced in capacity-constrained service industries such as the airlines, hotels, car rentals, and cruise-lines. Historically, yield management started as an operations function, focusing only on capacity allocation given *exogenous* demand estimates. However, there is a growing consensus among researchers and practitioners alike that the pricing decisions that induce demand cannot be separated from traditional, capacity-oriented yield management decisions; these two decisions are inextricably linked. Pricing decisions influence the demand statistics that form the basic inputs to any yield management system, and pricing decisions in turn must ultimately be based on knowledge of the underlying capacity constraints and yield management policies. Therefore, to realize the full benefits of yield management, *joint* pricing/allocation schemes are necessary.

While *what* must be done, namely coordinating pricing and allocation, is clear, exactly *how* it should be accomplished remains somewhat a mystery. The problem is complicated by the fact that in many applications, most notably the airlines, many markets (origin-destination pairs) compete for the same resources (leg capacities); therefore, the pricing/allocation problem must be solved jointly for a large number of markets and resources. Uncertainty in the demand process and the dynamic nature of the allocation decisions further compound the difficulty of the problem.

1. MODEL DESCRIPTION AND APPLICATIONS

We propose a dynamic, stochastic model of yield management problems that addresses the joint pricing/allocation

decision in a multiple-market, multiple-resource setting. Specifically, we define the problem in terms of *products* and *resources*. To provide a unit of product j , $j = 1, \dots, n$ requires a_{ij} units of resource i , $i = 1, \dots, m$. We define the *bill of materials* matrix $A = [a_{ij}]$. We assume A is an integer valued matrix. Demand for each product j is modeled as a stochastic point process with intensity λ^j . (All vectors are column vectors unless otherwise stated.) The vector of demand intensities $\lambda = (\lambda^1, \dots, \lambda^n)$ at time s , is a function of s and the vector of current prices $p = (p^1, \dots, p^n)$. This demand function need only satisfy some general regularity conditions. Demand for a given product j is allowed to depend on time, the price of product j , and the price of products other than j ; however, we do not allow the current demand to be a function of past or future prices. Thus, we assume that customers respond only to current prices and do not act strategically by adjusting their buying behavior in response to the firm's pricing policy. This is certainly a limitation of the model. To consider strategic customer behavior would require a game-theoretic formulation, which, though quite interesting, is beyond the scope of our present analysis.

Given initial quantities of resources at time zero and a time horizon t , the problem is to control prices, perhaps subject to price constraints, in order to maximize expected revenue over the interval $[0, t]$.

It is perhaps not immediately clear that this formulation is a reasonable model of a yield management problem. In particular, there appear to be no "allocation" decisions. However, one can view allocation, at least conceptually, as a pricing decision.

This is true for two reasons. First, if each product has only one fare class associated with it or each fare class is well differentiated in terms of payment requirements,

Subject classifications: Transportation: yield management, seat allocation, pricing. Dynamic programming: intensity control.

Area of review: STOCHASTIC PROCESSES & THEIR APPLICATIONS.

travel restrictions, etc., then closing a fare class is mathematically equivalent to setting the price for the corresponding product at a sufficiently high price, so that the expected demand rate is zero, and the fare class is effectively closed. In practice, of course, one would simply tell customers that the fare is no longer available. By modeling the closing of fares as we do, with no additional demands occurring after the last sale, we intentionally blur the distinction between demand and sales, and view pricing and allocation as one decision. In Section 7.2.1, we discuss situations in which limitations on pricing flexibility require a separate allocation policy.

Second, if one offers multiple fare classes for a given product which is not well differentiated, closing a low fare is mathematically equivalent to setting it to the next highest fare. Notice that this may cause an *increase* in demand at the next highest fare because some customers who are willing to buy at the low fare are also willing to buy at the next highest fare when the former is not available. This phenomenon is known in the airline industry as *demand recapture*.

In both instances, closing fare classes is equivalent to changing the “menu” of prices offered to customers at any point in time. In these two ways, one can map a wide range of joint pricing/allocation schemes into equivalent pure-pricing schemes. The advantage of the pricing framework is that it allows for a unified analysis of a rich class of yield management problems, and it directly reveals the relationship between pricing and allocation decisions.

Multidimensional problems of the type suggested by our formulation arise in a wide variety of applications. For example, in the case of airlines, a “product” is an itinerary (path) from an origin (O) to a destination (D) in an airline network and a “resource” is a seat on a particular flight leg (edge) in the network. A customer traveling from O to D requires one unit of resource on each leg in a particular itinerary from O to D. In some cases, one might have two products that use precisely the same set of resources. Again, this occurs in airline applications when well differentiated coach and super-saver products exist for each itinerary. In this case a product is an O-D itinerary at a certain fare level/restriction combination.

Another example of such multidimensional problems is found in the hotel industry. Yield management problems for hotels are often viewed as quite different from airline problems since customers may occupy a room for several nights. Because of this overlapping effect, each day cannot be viewed as an independent instance. To model this problem under the above framework, one defines a product for each possible combination of days that a customer might request over a specific period of time, e.g., one month. Resources correspond to room capacities on each day of the planning period (perhaps net of any group reservations), and products correspond to particular subsets of days. For example, one product could be a Monday-to-Thursday stay during Week 1. If sold, this product would then consume one unit each of the resources correspond-

ing to the days Monday, Tuesday, Wednesday, and Thursday of Week 1. Our multiproduct, multiresource formulation naturally captures the inter-day dependencies found in hotel booking.

Our model also applies to situations quite distinct from those of traditional yield management. One example is the style or fashion goods industry. In the fashion industry, contracts for garment production typically are fixed well in advance of a given selling season. During a sales season, retailers face the problem of adjusting prices, usually through periodic sales and/or mark-downs, so as to maximize the total revenue from their initial stock. These problems are often multi-dimensional in nature for two reasons. First, demand for products may be dependent due to the fact that one product may substitute for or complement another. Second, both products and resources may correspond to specific garment types at particular locations (either retail outlets or warehouses). Thus, if transshipments are allowed, the same “product” (garment j at location x) may be provided using one of several “resources” (garment j at another location y).

Other possible applications of the model include production/inventory problems where resources are components that can be assembled into a certain number of available products in a product line. The problem then is to price the product line to maximize the revenue from a given initial stock of components. For example, assembling a line of personal computers which share common components but are subject to rapid obsolescence is one possible application in this category.

In the interest of brevity and to maintain focus, we shall concentrate on airline yield management applications of our model. However, both our model and analysis are indeed quite general and can be used in a wide variety of other applications.

2. LITERATURE REVIEW

The study of yield management problems in the airlines dates back to the work of Rothstein (1971) for an overbooking model and to Littlewood (1972) for a model of space allocation of a stochastic two-fare, single-leg problem. Belobaba (1987b and 1989) proposed and tested a multiple-fare-class extension of Littlewood’s rule, which he termed the *expected marginal seat revenue* (EMSR) heuristic. Extensions and refinements of the multiple-fare-class problem include recent papers by Brumelle and McGill (1993), Curry (1989), Robinson (1995), and Wollmer (1992). The above papers develop optimal rules or heuristics under the assumption that the demand for different fare classes are random variables instead of stochastic processes. Some of the rules developed for random variables have been applied dynamically. See Weatherford et al. (1993) for extensive simulations on the performance of some of these rules under various models of the customer arrival processes. Lee and Hersh (1993) use discrete time

dynamic programming to develop optimal rules when demands are modeled as stochastic processes. They also allow for group bookings. Feng and Gallego (1995) develop optimal threshold rules when demands are modeled as continuous time stochastic process. Kimes (1989) gives a general overview of yield management practice in the hotel industry. Bitran and Gilbert (1992), Liberman and Yechiali (1978) and Rothstein (1974) investigate analytical models of hotel problems. Ladany and Arbel (1991) investigate pricing policies for cabins in a cruise-line. Chatwin (1992) characterizes the form of the optimal, dynamic overbooking policy for a single market.

A recent review of research on yield management as well as a comprehensive taxonomy of perishable asset revenue management (PARM) problems is given by Weatherford and Bodily (1992). They identify 14 descriptors that can be used to categorize a range of PARM problems. Our general formulation can model a broad subset of these descriptors, including dynamic pricing and allocation decisions, group demand, diversion, displacement, a time-varying willingness to pay, and no-shows (see Weatherford and Bodily 1992 for a detailed discussion of this taxonomy). In Section 4.3, a variety of examples that illustrate these features are presented.

Glover et al. (1982) were, to our knowledge, the first to address a deterministic network flow model for the allocation of seats between passenger itineraries and fare classes. Their formulation, however, does have a restriction to networks with no undirected cycles. In addition to their formulation, Glover et al. give an interesting account of the effects of deregulation. They recognize two components to the airline problem, one tactical and one strategic, with the tactical dealing with monitoring reservations and the strategic consisting of the selection of prices and routes.

Wang (1983) provides an algorithm for the sequential allocations of seats on a plane to different uplift-discharge city pairs and fare classes within a flight segment when demands are random. Wang reports a 4% increase in revenue for high load factor flights. In addition, Wang provides a good explanation of the price/market differentiation scheme used by most airlines.

Dror et al. (1988) present a rolling horizon network flow formulation for the seat inventory control problem assuming deterministic demands. Their formulation is extremely complex to account for passengers switching planes at intermediate stops, but allows for arc losses to account for cancellations and no shows. No specification of how to obtain the parameters of the formulation or how to solve the network flow problem is given.

In two internal McDonnell Douglas reports, Wollmer (1986a and b) proposes a mathematical programming formulation incorporating random demands, where the objective is to maximize the total expected network revenue. The formulation is very large since there is a variable for each origin/destination, fare class, and seat on each flight leg. On the positive side, the formulation is very accurate.

In her Ph.D. thesis, Williamson (1992) proposes schemes that use well-known leg-based results for solving the network problem. In addition, she uses dual price information to develop a bid-price approach to the network problem. These schemes are extensively tested via simulation.

A traffic flow model of airline networks is proposed by Soumis and Nagurney (1993). They formulate and analyze a stochastic, multiclass airline network model for determining the equilibrium between the realized demand for different travel routes and the supply of seats. The traffic flows predicted by the model are tested against data from Air Canada.

The most recent reference known to us is due to Talluri (1993), who formulates a deterministic network model based on origin-destination pairs. His key idea is that passengers may be indifferent between different routes for an origin-destination pair provided they have similar departure and arrival times. Airlines can take advantage of this indifference by intelligently routing passengers through origin-destination pairs. The decision variables of Talluri's model correspond to the allocation of demand for origin-destination pairs to routes for which customers are indifferent. When demand is assumed to be stochastic, Talluri provides a deterministic integer-programming formulation based on the idea of maximizing expected marginal revenue. In addition to his formulation, Talluri offers an interesting account of the state of the art of computer reservation systems used to support network models.

Our heuristics are based on deterministic network flow models that are similar in spirit to those of Dror et al., Glover et al., and Talluri. However, the exact relationship of these models to the underlying adaptive stochastic control problem has not been previously addressed. Our results allow us to rigorously analyze this relationship.

3. OUTLINE AND OVERVIEW OF MAIN RESULTS

We begin by formally stating the formulation and assumptions of our general model in Section 4. We then prove that a deterministic version of the problem provides an upper bound on the revenue of the original stochastic problem. The deterministic solution suggests two possible heuristics, which we call the make-to-stock (MTS) and make-to-order (MTO) heuristic. We show in Section 5 that both heuristics are asymptotically optimal as the volume of expected sales tends to infinity. We extend these results to the case where overbooking and cancellations are allowed in Section 6.

Next, in Section 7 we examine several applications of our results to specific airline yield management problems. These include network, product differentiation, and overbooking problems and problems with time varying demand. We also examine how the lack of price flexibility and misspricing affect revenues and the performance of the various heuristics. Our conclusions and thoughts for further research are presented in Section 8.

4. ASSUMPTIONS AND FORMULATIONS

4.1. Modeling Assumptions

We assume a market with imperfect competition in which demand for each product $j = 1, \dots, n$ varies with its price p^j . Demand for products at time s is a multivariate, stochastic point process with Markovian intensities. At any time s , the vector of intensities $\lambda = (\lambda^1, \dots, \lambda^n)$ is determined by s and the current price vector $p = (p^1, \dots, p^n)$ through a demand function $\lambda(p, s)$, $\lambda: R^{n+1} \rightarrow R^n$. Thus, demand is a controlled Poisson process.

We assume the demand function $\lambda(p, s)$ is known, and further, that it satisfies the following regularity conditions.

1. For every s , the demand function $\lambda(p, s)$ has an inverse, denoted $p(\lambda, s)$, $p(\cdot, s): R_+^n \rightarrow R_+^n$. Thus, we can and shall view the vector of intensities $\lambda = (\lambda^1, \dots, \lambda^n)$ as the firm's decision variables. In this case, one can imagine the firm setting the output intensities λ and the market determining the prices p based on these output intensities.

2. The *revenue rate* at time s , defined by

$$r(\lambda, s) \doteq \lambda' p(\lambda, s), \quad (1)$$

is continuous, bounded and concave. (For a vector λ , we let λ' denote its transpose.) Further, we assume it satisfies the conditions $\lim_{\lambda, \lambda' \rightarrow 0} \lambda^j p^j(\lambda, s) = 0$ for all finite λ and all $j = 1, \dots, n$, and its maximizer $\lambda_0^j(s)$ is uniformly bounded, i.e., $\lambda_0^j(s) \leq \bar{\lambda} < +\infty$ for all j and s . Informally, the first condition says that no revenue is earned at a sales rate of zero, while the second says that sales rates are always bounded.

3. For each product j and all times s there exists a *null price* $p_\infty^j(s)$ (possibly $+\infty$) such that if $\{p_k\}$ is any sequence of price vectors satisfying $p_k^j \rightarrow p_\infty^j(s)$, then $\lim_{k \rightarrow \infty} \lambda^j(p_k, s) = 0$. Essentially, this condition says that there always exists some price that we can use to effectively "turn off" the demand for product j . This allows us to model the out-of-stock condition, i.e., not being able to provide a final product j due to the lack of one or more resources, as an implicit constraint that forces the firm to price at $p^j = p_\infty^j(s)$ for all times s during which the inventory of any resource i is less than a_{ij} . (Cohen 1977 and Karlin and Carr 1962 make a similar null price assumption for single-resource problems.)

A function $\lambda(p, s)$ that satisfies all these assumptions is called a *regular* demand function.

4.2. Formulation of the Stochastic Problem

The n final products are produced from m types of resources. There is an initial stock of each resource i , denoted x^i , and we define the vector $x = (x^1, \dots, x^m)' \in \mathcal{X}_+^m$. (Here \mathcal{X}_+ denotes the set of positive integers.) The firm has a deadline $t > 0$ after which selling must stop, and no additional resources can be obtained over $[0, t]$. Product j requires a_{ij} units of resource type i and $A = [a_{ij}]$. We assume A is integer valued and has no zero columns; that is, each product uses at least one of the m resource types. Let the counting process N_s^j denote the number of type j

products sold up to time s and $N_s = (N_s^1, \dots, N_s^m)$. Let $p_s = (p_s^1, \dots, p_s^n)$ be the vector of prices at time s . The intensity of the process N_s is controlled by p_s through a regular demand function as described above. We assume this demand function is known. A demand for product j is realized at time s if $dN_s^j = 1$, in which case the firm takes a_{ij} units of each resource i , $i = 1, \dots, m$, out of its inventory, produces the product, and provides it to the customer in return for a revenue of p_s^j .

Prices are chosen using a nonanticipating *pricing policy* $p_s = p(\lambda_s, s)$. The vector of prices p_s must be chosen from a set $\mathcal{P}(s)$ of allowable prices. Equivalently prices may be chosen via nonanticipating intensities $\lambda_s = \lambda(p_s, s)$. The set of allowable intensities is denoted by $\Lambda(s) = \{\lambda(p, s): p \in \mathcal{P}(s)\}$, and we assume initially that $\Lambda(s)$ is convex for all s , although the convexity restriction can be relaxed in certain cases as explained below.

We denote by \mathcal{U} the class of all nonanticipating pricing policies which satisfy

$$\int_0^t A dN_s \leq x \quad (\text{a.s.}), \text{ and} \quad (2)$$

$$p_s \in \mathcal{P}(s) \Leftrightarrow \lambda_s \in \Lambda(s), \quad 0 \leq s \leq t. \quad (3)$$

Constraint (2) acts to "turn off" the demand process for a product j when the firm lacks sufficient resources to provide it. Thus as soon as a resource, say i , is exhausted, at time s , the prices of all products j consuming resource i ($a_{ij} > 0$) are increased to $p_\infty^j(s)$. The existence of a null price $p_\infty^j(s)$ for all j at all times s guarantees that (2) can always be satisfied (with probability 1). Finally, we assume that the salvage value of unused resources is zero. We show below that this assumption is made without loss of generality.

Note that pricing policies are nonanticipating in the sense that the price at time s can depend only on s and on the realization of sales up to but not including time s . Also, since we allow policies where the price depends on the actual realization of demand, the price p_s and the corresponding demand intensity λ_s are themselves random vectors. Finally, note that the price vector does not need to be a continuous function of time; it merely needs to be contained in the set $\mathcal{P}(s)$.

Given an initial vector of resource inventories x , a deadline $t > 0$, and a regular demand process as described above, we want to find a pricing policy in \mathcal{U} that maximizes total expected revenue. More formally, we denote the expected revenue of policy u by

$$J^u(x, t) \doteq E_u \left[\int_0^t p_s' dN_s \right], \quad (4)$$

and the firm's problem is to find a pricing policy u^* , if such a policy exists, that maximizes the total expected revenue generated over $[0, t]$, denoted $J^*(x, t)$. Equivalently,

$$J^*(x, t) \doteq \sup_{u \in \mathcal{U}} J^u(x, t). \quad (5)$$

4.3. Special Cases

We briefly mention some special cases of this formulation to illustrate its range of application. Further examples are given in Section 4.

Nonstationary Demand. Demand may change over time due to seasonalities, day-of-the-week effects, or the increased or decreased desirability of products as the deadline approaches. Such nonstationarity is modeled using a time varying demand function.

Discounted Revenue. One can have discounted revenue functions of the form $r(\lambda, s) = e^{-\beta s} \lambda p(\lambda, s)$, provided $r(\lambda, s)$ satisfies the regularity assumptions.

Resource Salvage Values. It is clear that with salvage values $q^i \geq 0$ for resources $i = 1, \dots, m$, we would never set $p^j < \pi^j = \sum_{i=1}^m a_{ij} q^i$ for any j ; therefore, we can define a new price vector $\hat{p}^j = p^j - \pi^j$ and demand function $\hat{\lambda}(\hat{p}^j) = \lambda(\hat{p}^j + \pi^j)$ that reduce the problem to the zero-salvage-value case. (In this case, “revenue” in the problem is interpreted as revenue in excess of the total salvage value of the initial stock of resources, $q'x$.)

Group Pricing. It is often possible to market multiples of a single product as a new product. For example, this occurs frequently in the airlines when blocks of space are sold to tour groups. This sort of group pricing can be modeled by introducing columns of the form $A_k^j = kA^j$, $k \in \mathcal{L}_+$, and then specifying a separate price and separate demand function for each such column.

Compound Poisson Demand. Suppose demand for product j is compound Poisson. That is, each customer request k items with probability q_k , $k = 1, \dots, K < +\infty$, $\sum_k q_k = 1$. Let A^j be the column corresponding to product j and let $p^j(\lambda, s)$ be the inverse demand function of product j . In this case, one can replace this single product by K products with columns $A_k^j = kA^j$, $k = 1, \dots, K$ and inverse demand functions $p_k^j(\lambda, s) = kp^j(\lambda, s)$, and add the set of convex constraints $\lambda_k^j = q_k \lambda^j$, $k = 1, \dots, K$. This yields an equivalent problem in the standard form of our formulation.

Stochastic Customer Selection of Products. In many cases, firms cannot control the price of each product separately. For example, airlines typically offer a fare for O-D pairs only and cannot force passengers to use a particular itinerary (path) connecting the O-D pairs, since typically all “legal” connections of legs are allowed. If demand functions are linear, this situation can be modeled as follows: Let p denote the price for a given O-D pair connected by K distinct paths. Let q_k , $k = 1, \dots, K$ be the probabilities representing preferences of customers for these paths; that is, a given customer requests path k with probability q_k . By pricing the O-D pair at p , intensities $\lambda_k = q_k \lambda(p)$, are generated for each path $k = 1, \dots, K$. If $\lambda(p)$ is linear and $\mathcal{P}(s)$ is convex, then the set of achievable intensities $\Lambda(s)$ is convex, and the problem reduces to our standard form.

Discrete Fare Classes. In many cases, firms do not have complete flexibility in setting prices. For example, airlines are often forced to match the fares of competitors or are committed to offering a certain number of advertised promotional fares. As a result, a firm may be restricted to a small number of price points for each product. This means that, on a tactical level, the firm may be able to control only the *availability* of its various fares, and may not have the freedom to change the fares themselves.

To model this situation, one can consider the set of allowable prices $P(s)$ to be a finite set, reflecting the possible fares that can be offered for each product together with the null prices p_∞^j that block demand for product $j = 1, \dots, n$. The set $\Lambda(s)$ is then a finite set as well. In this case, the decision is which of the discrete set of prices to make available at any point in time for each product. If the null price is offered, then the product is closed, and no sales are made. This type of decision is more in the spirit of traditional yield management problems, in which one controls the availability of various fare classes for a given product. The discrete price formulation creates a technical problem since $\Lambda(s)$ is not convex in this case; however, we discuss in Section 4.5 how this restriction can be relaxed if the revenue rate is constant or piecewise constant in time. If the revenue rate is continuously varying in time, it is possible to use randomized rules to overcome the convexity problem.

4.4. Optimality Conditions

The formulation (5) is an example of an intensity control problem. An optimal solution can be found in principle through the Hamilton-Jacobi sufficient conditions (Brémaud 1980, Theorem VII.1):

$$\begin{aligned} & \frac{\partial}{\partial t} J^*(x, t) \\ &= \sup_{\lambda \in \Lambda(x, s)} \left\{ r(\lambda, s) - \sum_{j=1}^n \lambda^j [J^*(x, t) - J^*(x - A^j, t)] \right\}, \end{aligned} \quad (6)$$

where

$$\Lambda(x, s) = \Lambda(s) \cap \{ \lambda: \lambda^j = 0 \text{ if } a_{ij} < x^i \text{ for some } i \},$$

denotes the set of allowable intensities in state x at time s and A^j denotes the j th column of A . J^* satisfies the boundary conditions

$$\begin{aligned} J^*(x, t) &= 0 & \forall t & & x: x^i < a_{ij} \text{ for some } i \\ J^*(x, 0) &= 0 & \forall x. & & \text{and all } j = 1, \dots, n \end{aligned}$$

This implicitly defines a system of first-order, typically nonlinear, differential equations.

In general it is very difficult to find closed form solutions to systems of differential equations such as the one presented above. (See Gallego and van Ryzin 1994 for a closed-form solution for a single-product problem with an exponential demand function.) Although some of these systems can be solved numerically, the solutions tend to call for continuous price decreases between sales and for price jumps immediately after each sale. Therefore, we

next examine an approach for finding provably good sub-optimal pricing policies based on analyzing a deterministic version of this problem.

4.5. Formulation of Deterministic Problem

The deterministic problem is defined as follows. At time zero, the firm has a vector of material supplies x , which are *continuous quantities*, and a line of products $j = 1, \dots, n$ which are sold in *continuous amounts*. To produce a unit quantity of product j requires quantities a_{ij} of each material $i = 1, \dots, m$. As before $A = [a_{ij}]$, and the firm has a finite time $t > 0$ to sell its products. Demand at time s is modeled as a vector of *deterministic rates* $\lambda(s) = (\lambda^1(s), \dots, \lambda^n(s))$ that are functions $\lambda(p, s)$ of the current product price vector $p(s) = (p^1(s), \dots, p^n(s))$. We assume the demand function and the revenue function $r(\lambda(s), s) = \lambda'p(\lambda, s)$ satisfy our regularity assumptions as before and that salvage values are zero. The price vector $p(s)$ must again be chosen from a set $\mathcal{P}(s)$ of allowable prices. As before, we can equivalently view the firm as setting the vector of sales rates $\lambda(s) \in \Lambda(s)$, which implies charging a vector of prices $p(s) = p(\lambda(s)) \in \mathcal{P}(s)$.

The firm's problem is to maximize the total revenue generated over $[0, t]$ given x , denoted $J^d(x, t)$.

$$J^d(x, t) = \max \int_0^t r(\lambda(s), s) ds \tag{7}$$

subject to

$$\int_0^t A\lambda(s) ds \leq x, \\ \lambda(s) \in \Lambda(s), \quad 0 \leq s \leq t.$$

This problem is a deterministic optimization problem whose solution, if one exists, is a function from $[0, t]$ to R^n , denoted $\lambda_d(s)$. Such problems can be analyzed either through the calculus of variations or the more general theory of optimization in vector spaces (Luenberger 1969).

Notice that in the time invariant case where $r(\lambda, s) = r(\lambda)$ and $\Lambda(s) = \Lambda$, the above problem reduces to a convex programming problem in R^n , and solutions are always constant intensities (prices) over $[0, t]$. It is not hard to show that one can allow Λ to be nonconvex in this case. This is accomplished by replacing Λ by its convex hull and interpreting the resulting constant intensity solution of the new problem as a mixture of intensities from the original problem, the mixture being achieved by varying the amount of time each fare is offered over the sales horizon. Indeed, in this case where Δ is discrete the deterministic problem reduces to a linear program, with decision variables being the amount of time each discrete price (intensity) is offered over $[0, t]$. The resulting solution is a set of piecewise constant intensities. An especially important application of this case is when prices, and hence intensities, are restricted to a finite set of fare classes as was discussed in Section 4.3. (See Gallego and van Ryzin for a detailed explanation of this case for a single-item problem.) This

same approach can be easily extended to the case where the revenue function and constraint sets are piecewise constant over time.

We use the fact that the optimal revenue in the deterministic problem can be expressed as

$$J^d(x, t) = \sum_{j=1}^n \bar{p}^j \alpha^j, \tag{8}$$

where

$$\alpha^j = \int_0^t \lambda_d^j(s) ds, \tag{9}$$

is the total quantity of product j sold under the optimal policy and

$$\bar{p}^j = \frac{\int_0^t p_d^j(s) \lambda_d^j(s) ds}{\int_0^t \lambda_d^j(s) ds}, \tag{10}$$

is the weighted average price obtained for product j .

4.6. Relationship Between Stochastic and Deterministic Problems

Our first theorem proves the rather intuitive result that the deterministic problem provides an upper bound on the stochastic problem. The proof requires a preliminary lemma. Let $\mu = (\mu_1, \dots, \mu_m)$ be a real m -vector. For $\mu \geq 0$ and any $u \in \mathcal{U}$, define the augmented value function

$$J^u(x, t, \mu) = E_u \left[\int_0^t (r(\lambda_s, s) - \mu' A \lambda_s) ds \right] + \mu' x \geq J^u(x, t), \tag{11}$$

and its deterministic equivalent

$$J^d(x, t, \mu) = \max_{\{\lambda(s) \in \Lambda(s)\}} \int_0^t (r(\lambda(s), s) - \mu' A \lambda(s)) ds + \mu' x. \tag{12}$$

We claim the following.

Lemma 1.

$$J^u(x, t, \mu) \leq J^d(x, t, \mu) \quad \forall u \in \mathcal{U}, \mu \geq 0.$$

Proof. The claim follows by viewing the integrand inside the expectation in (11) as purely a function of λ and maximizing pointwise:

$$J^u(x, t, \mu) \leq \int_0^t \max_{\lambda \in \Lambda(s)} \{r(\lambda, s) - \mu' A \lambda\} ds + \mu' x \\ = \max_{\{\lambda(s) \in \Lambda(s)\}} \int_0^t (r(\lambda(s), s) - \mu' A \lambda(s)) ds + \mu' x \\ \doteq J^d(x, t, \mu). \quad \square$$

We are now ready to prove our first theorem.

Theorem 1. *If $\lambda(p, s)$ is a regular demand function, then for all nonnegative integer x and all $0 \leq t < +\infty$*

$$J^*(x, t) \leq J^d(x, t).$$

Proof. By our regularity assumption, we have $\lambda_s^j \leq \bar{\lambda} < \infty$, which implies $\int_0^t A\lambda_s ds$ is finite almost surely for all $t \geq 0$. Because $u \in \mathcal{U}$ satisfies $\int_0^t AdN_s \leq x$ (a.s.), we have by Brémaud, Theorem II.8,

$$E_u \left[\int_0^t AdN_s \right] = E_u \left[\int_0^t A\lambda_s ds \right] \leq x. \quad (13)$$

Since the demand intensity in the control problem (5) is Markovian, it is sufficient to consider only Markovian policies u (Brémaud, Corollary VII.2). That is, policies for which the price at time s is a function $p_s = p_u(x - N_s, s)$ only. (Equivalently, the intensity at time s is a function $\lambda_s = \lambda_u(x - N_s, s)$.) By Brémaud, Theorem II.8, we have,

$$\begin{aligned} J^u(x, t) &= E_u \left[\int_0^t p_u(x - N_s, s) dN_s \right] \\ &= E_u \left[\int_0^t r(\lambda_u(x - N_s, s), s) ds \right], \end{aligned} \quad (14)$$

and by definition

$$J^*(x, t) = \sup_{u \in \mathcal{U}} J^u(x, t).$$

Thus, by (11) and Lemma 1 we have

$$J^*(x, t) \leq \inf_{\mu \geq 0} J^d(x, t, \mu). \quad (15)$$

Theorem 1 then follows by noting that the quantity on the right, above, is the optimal value of the deterministic optimization problem

$$J^d(x, t) = \max \int_0^t r(\lambda(s), s) ds$$

subject to

$$\int_0^t A\lambda(s) ds \leq x$$

$$\lambda(s) \in \Lambda(s), \quad 0 \leq s \leq t,$$

where the constraints we are relaxing are linear and thus convex, $\Lambda(s)$ is convex by assumption, and the objective functional is concave. If $x > 0$, the solution $\lambda(s) = 0$, $0 \leq s \leq t$ gives $\int_0^t A\lambda(s) ds < x$ and thus is an interior point of the relaxed constraint set. Under these conditions, there exists a multiplier $\mu^* \geq 0$ for which

$$\inf_{\mu \geq 0} J^d(x, t, \mu) = J^d(x, t, \mu^*) = J^d(x, t),$$

so the duality gap is zero (Luenberger, Section 8.6, Theorem 1). This together with (15) establishes the result for $x > 0$. If $x^i = 0$ for some resource i , it is not hard to see that any product j that uses resource i can be eliminated and the problem can be reduced to one in which $x > 0$. \square

5. TWO ASYMPTOTICALLY OPTIMAL HEURISTICS

We next examine heuristics suggested by the solution to the deterministic problem. Two implementations of the deterministic solution are considered, a make-to-stock and

a make-to-order heuristic. We show both are asymptotically optimal as the expected volume of sales tends to infinity.

5.1. The Make-to-Stock Policy

In the deterministic problem, we know with certainty the quantity of each product that will be sold under any given policy. Therefore it is possible to produce products *in advance* and hold inventories of finished products rather than hold inventories of resources. This suggests the following heuristic:

Make-to-Stock (MTS) Policy. Let $p_d(s)$ be the optimal deterministic price path and $\lambda_d(s)$ be the corresponding optimal deterministic intensity. Define

$$z^j = \left\lfloor \int_0^t \lambda_d^j(s) ds \right\rfloor = \lfloor \alpha^j \rfloor. \quad (16)$$

Preassemble z^j units of product $j = 1, \dots, n$ and place the products in separate inventories. Price products at $p_d(s)$, $0 \leq s \leq t$ and sell them until the product inventories are exhausted or the deadline t is reached, whichever comes first.

Here $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .

Of course, there is no need to *physically* preassemble products in this heuristic; we need only *logically* reserve resources for specific products. Effectively, however, the MTS heuristic takes away both price flexibility and product mix flexibility. Despite this lack of flexibility, the policy performs quite well when the expected number of sales of each product is large. (Computational examples are given in Section 7.)

Before proving this claim, we need a preliminary result. Consider just a single product being sold over $[0, t]$ using a deterministic (scalar) price path $p_d(s)$ and deterministic (scalar) intensity $\lambda_d(s)$. Suppose there is an infinite supply of the product and let R , a random variable, denote the total revenue received by following this optimal deterministic price path. The distribution of R is characterized as follows.

Lemma 2. Let N be a Poisson random variable with mean $\int_0^t \lambda(s) ds$ and $\{\hat{T}_k: k \geq 1\}$ be an i.i.d. sequence of random variables with density function

$$f(s) = \begin{cases} \frac{\lambda(s)}{\int_0^t \lambda(s) ds}, & 0 \leq s \leq t \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$R \stackrel{D}{=} \sum_{k=1}^N p_d(\hat{T}_k),$$

where $\stackrel{D}{=}$ denotes equality in distribution.

Proof. Since arrivals form a Poisson process, conditional on $N = m$, the arrival times of the m sales, T_1, \dots, T_m , can be represented as the order statistics of the i.i.d.

random variables $\hat{T}_1, \dots, \hat{T}_m$. Therefore, the prices $p_d(\hat{T}_1), \dots, p_d(\hat{T}_m)$ are i.i.d. and $R = \sum_{k=1}^m p_d(T_k) = \sum_{k=1}^m p_d(\hat{T}_k)$. Since $p_d(\hat{T}_1)$ is independent of N , the result follows after unconditioning on N . \square

The following theorem establishes the asymptotic optimality of the MTS heuristic.

Theorem 2. Let $u^j = \sup\{p_d^j(s) : \lambda_d(s) > 0, 0 \leq s \leq t\}$. Then

$$\frac{J^{\text{MTS}}(x, t)}{J^*(x, t)} \geq 1 - \frac{\sum_{j=1}^n u^j \left(\frac{1}{2} \sqrt{\alpha^j} + 1\right)}{\sum_j \bar{p}^j \alpha^j}.$$

Some explanation of this theorem is in order before we present a proof. Here, u^j is an upper bound on the price obtained for item j and α^j is the quantity of sales under the deterministic policy defined by (9). In the context of the stochastic problem, one can think of this last quantity as the mean number of requests for product j over the interval $[0, t]$ when using the deterministic prices $p_d(s)$. Note that the maximum prices u^j must be finite, otherwise this bound is trivial. Theoretically, this is a somewhat a restriction, but it poses no real practical limitation. Provided u^j is finite for all products j , the error term tends to zero as the quantities α^j increase. Thus, if the expected sales volumes are large, the error term above is small. We give a more precise interpretation of this bound after presenting the proof.

Proof. Define $\alpha^j = \int_0^t \lambda^j(s) ds$ as in the deterministic problem, and let $N(j)$ denote a Poisson random variable with mean α^j . Let $\{T_k : k \geq 1\}$ denote the sequence of arrival times for a Poisson process with intensity $\lambda_d^j(s)$, $0 \leq s \leq t$, and let R^j denote the revenue from product j under the MTS heuristic. Then we have

$$\begin{aligned} E[R^j] &= E \left[\sum_{k=1}^{N(j)} p_d^j(T_k^j) - \sum_{k=z^j}^{N(j)} p_d^j(T_k^j) \right] \\ &\geq E \left[\sum_{k=1}^{N(j)} p_d^j(T_k^j) \right] - u^j E[(N(j) - z^j)^+], \end{aligned} \quad (17)$$

where $x^+ \doteq \max(x, 0)$ and u^j is the upper bound on $p_d^j(s)$ as defined above. From Lemma 2 and Wald's equation we have that the first term above is

$$E \left[\sum_{k=1}^{N(j)} p_d^j(T_k^j) \right] = E[N(j)] E[p_d^j(\hat{T}_k^j)] = \alpha^j \bar{p}^j.$$

For a random variable D with mean μ and standard deviation σ , and for any real number d , we have the following inequality due to Gallego (1992):

$$E[(D - d)^+] \leq \frac{\sqrt{\sigma^2 + (d - \mu)^2} - (d - \mu)}{2}. \quad (18)$$

Applying this inequality to the line of (17) and noting that $\mu = \sigma^2 = \alpha^j$ for the Poisson distribution we obtain

$$\begin{aligned} E[(N(j) - z^j)^+] &\leq \frac{\sqrt{\alpha^j + (z^j - \alpha^j)^2} - (z^j - \alpha^j)}{2} \\ &\leq \frac{\sqrt{\alpha^j}}{2} + \frac{|z^j - \alpha^j| - (z^j - \alpha^j)}{2} \\ &\leq \frac{\sqrt{\alpha^j}}{2} + 1, \end{aligned}$$

where the last inequality follows from the fact that $z^j = \lfloor \alpha^j \rfloor$. Thus

$$E[R^j] \geq \alpha^j \bar{p}^j - u^j \left(\frac{\sqrt{\alpha^j}}{2} + 1 \right).$$

Summing these revenues over all products j and recalling that Theorem 1 gives us $J^*(x, t) \leq J^d(x, t) = \sum_j \bar{p}^j \alpha^j$ establishes the desired result.

Remark. One can strengthen the theorem somewhat by noting that it remains true for u^j defined as the smallest u satisfying

$$u \geq \int_0^y p_d^j(s) \frac{f^j(s)}{\int_0^y f^j(\xi) d\xi} ds, \quad \forall 0 \leq y \leq t,$$

where $f^j(\cdot)$ is the density defined in Lemma 2 for product j . Note that an upper bound on $p_d^j(s)$ as defined in Theorem 2 certainly satisfies this inequality, though in general smaller values may work. To see this definition of u^j is valid, condition on the time Y at which the allocation z^j is exhausted. Then by an argument identical to that in Lemma 2 we have

$$\begin{aligned} E \left[\sum_{k=z^j}^{N(j)} p_d^j(T_k^j) | Y = y \right] \\ = \left[\int_0^y p_d^j(s) \frac{f^j(s)}{\int_0^y f^j(\xi) d\xi} ds \right] E[N(j) - z^j | Y = y], \end{aligned}$$

where the first term is defined to be zero when $Y \geq t$ (equivalently, when $N(j) \leq z^j$). Applying u^j as an upper bound on the first term for $Y < t$ and unconditioning we obtain a bound of $u^j E[(N(j) - z^j)^+]$ as before.

To understand the significance of Theorem 2 more easily, consider a sequence of problems indexed by integers $k = 1, 2, \dots$ defined by an initial vector of resources $x_k = kx$ and a demand function $\lambda_k(p, s) = k\lambda(p, s)$ for some fixed $x, \lambda(p, s)$ and t . This generates a sequence of problems with proportionately larger sales volumes and initial stocks. It is not hard to see that this sequence of problems has the same optimal deterministic price path $p_d(s)$ for all k ; hence

$$\begin{aligned} \frac{J_k^{\text{MTS}}(x_k, t)}{J_k^*(x_k, t)} &\geq 1 - \frac{\sum_{j=1}^n u^j \left(\frac{\sqrt{k}}{2} \sqrt{\alpha^j} + 1 \right)}{k \sum_{j=1}^n \bar{p}^j \alpha^j} \\ &= 1 - O(k^{-1/2}), \end{aligned}$$

where α^j is defined as before for the case $k = 1$. This clearly shows the MTS heuristic is asymptotically optimal as the scale of the problem increases ($k \rightarrow \infty$).

5.2. The Make-to-Order Policy

An alternative to the MTS policy is to price products according to the deterministic prices $p_d(s)$, $0 \leq s \leq t$ and then simply satisfy requests in a first-come-first-serve order. This idea leads to our next policy.

Make-to-Order (MTO) Policy. *Follow the deterministic optimal price path $p_d(s)$ over $[0, t]$ and assemble and sell products in the order in which they are requested. Reject requests for a product j when the inventory of one or more of its resources i drops below a_{ij} .*

Before stating our main result on this policy, several definitions are needed. Let $S_i = \{j: a_{ij} > 0\}$ denote the set of products j that use components of type i . As before, let $\alpha^j = \int_0^t \lambda_d^j(s) ds$ denote the mean number of requests for product j over $[0, t]$, and define $\alpha(S_i) = \sum_{j \in S_i} \alpha_j$. Note that $\alpha(S_i)$ is the total expected number of request for products that use components of type i . Let $v^i = \max\{u^j: j \in S_i\}$, where $u^j, j = 1, \dots, n$ are the uniform upper bounds on the price paths defined in Theorem 2, and define $\bar{v}^i = v^i \max\{a_{ij}: j \in S_i\}$. Finally, let $\hat{p}^i = \sum_{j \in S_i} \bar{p}^j \alpha^j / \alpha(S_i)$ and note that with this definition, the deterministic revenue can be written $J^d(x, t) = \sum_i \hat{p}^i \alpha(S_i)$.

Theorem 3.

$$\frac{J^{\text{MTO}}(x, t)}{J^*(x, t)} \geq 1 - \frac{\sum_{i=1}^m \bar{v}^i \sqrt{\alpha(S_i)}}{2 \sum_{i=1}^m \hat{p}^i \alpha(S_i)}.$$

Proof. In the MTO heuristic, requests for product j are filled up until the first time the inventory for a resource i drops below a_{ij} . Consider a modified system in which we allow inventories of resources to be negative (i.e., backlogging is allowed) but we charge a *penalty* v^i for every unit of resource i that is backlogged at the end of the horizon. We claim the net revenue (revenue minus backlog penalties) collected in this modified system is, pathwise, a lower bound on the revenue of the MTO heuristic.

To see this, consider a request for product j at time s that is rejected under the MTO heuristic because the inventory of some component, call it i , is strictly less than a_{ij} . In the MTO heuristic, no revenue is collected and the inventories are left intact. In the modified system, the request is satisfied and revenue $p^j(s)$ is collected; however, the backlog of resource i is increased by at least one unit. Thus, the net revenue received from the arrival in the modified system is no more than $p^j(s) - v^i$, which by definition of v^i , satisfies $p^j(s) - v^i \leq p^j(s) - u^j \leq 0$. Hence, revenue net of penalties does not increase. Further, there are fewer resources left for future requests. Thus, the net revenue of the modified system is no more than the revenue of the MTO heuristic for every sample path as claimed.

As before, let $\{T_k^j: k \geq 1\}$ denote the sequence of arrival times of these requests. Then by comparison to the modified system we have

$$J^{\text{MTO}}(x, t) \geq \sum_{j=1}^n E \left[\sum_{k=1}^{N(j)} p_d^j(T_k^j) \right] - \sum_{i=1}^m v^i E \left[\left(\sum_{j=1}^n a_{ij} N(j) - x^i \right)^+ \right]. \quad (19)$$

The first term above is simply $J^d(x, t)$, as was shown in the proof of Theorem 2. To bound the second sum, notice that if $d \leq \mu$ in (18), then $E(D - d)^+ \leq \sigma/2$. Consequently,

$$E \left[\left(\sum_{j=1}^n a_{ij} N(j) - x^i \right)^+ \right] \leq \frac{\sigma_i}{2},$$

where

$$\sigma_i^2 \doteq \text{Var} \left[\sum_{j=1}^n a_{ij} N(j) \right] = \sum_{j=1}^n a_{ij}^2 \alpha^j,$$

and, by construction,

$$\mu_i \doteq E \left[\sum_{j=1}^n a_{ij} N(j) \right] = \sum_{j=1}^n a_{ij} \alpha^j \leq x^i.$$

Noting that $\sum_{j=1}^n a_{ij}^2 \alpha^j \leq \max\{a_{ij}^2: j \in S_i\} \alpha(S_i)$, using the definition of \bar{v}^i , and using Theorem 1, we establish the desired result. \square

Again, by considering a sequence of problems defined by an initial vector of resources $x_k = kx$ and a demand function $\lambda_k(p, s) = k\lambda(p, s)$ for some fixed $x, \lambda(p, s)$ and t , one can see that as $k \rightarrow \infty$ the relative error of the MTO heuristic is $1 - O(k^{-1/2})$. Notice, however, that in this case the error bounds depend on $\alpha(S_i)$, the total expected requests for products that use component i , and not directly on α^j , the number of requests for product j . Thus, this bound is useful even in cases where there are arbitrarily small numbers of requests for each product, as long as the total volume of components sold is high.

This behavior provides a distinct advantage over the MTS heuristic in cases where the number of products is much larger than the number of component types. For example, in airline networks there are typically a very large number of O-D combinations (products) offered with potentially few sales of any given combination occurring on a specific leg (component). A pure O-D allocation, such as in the MTS heuristic, can result in reserving only a handful of seats for each O-D market. Indeed, precisely for this reason, the airline industry has developed *virtual nesting* (see Smith et al. 1992 for a discussion) and other aggregation schemes to compensate for this “small numbers” problem.

While the bounds in Theorems 2 and 3 are theoretically satisfying, they are nevertheless quite crude from a practical standpoint. However, in Section 7 we examine some

results of numerical experiments that indicate the performance of the heuristics is good, even for instances where sales volumes are moderate.

6. A MODEL WITH OVERBOOKING AND NO-SHOWS

Because of the possibility of no-shows (customers who reserve a seat but do not show-up at the departure time), most airlines accept reservations in excess of capacity. This, however, can result in flights being overbooked and the possibility of refusing seats to ticketed passengers. When realized demand exceeds capacity, customers typically are serviced by an alternative source (e.g., another flight or a nearby hotel). In addition, they may be compensated for the inconvenience (lunch, cocktails, a free taxi ride) and to restore goodwill (a free flight coupon).

We next extend our model to allow for overbooking and no-shows. We assume that revenues are collected as reservations are made, that customers show up independently of each other just prior to the delivery of service, and that additional capacity can be secured at a given unit cost. Customers who do not show are refunded the price paid minus a fixed plus variable penalty. The problem is to determine the pricing, and overbooking policy that maximizes expected revenue.

We begin by deriving an expression for the expected revenue. Let γ^j denote the probability that a customer with a reservation for product j shows up; let $c^j \geq 0$ denote the fixed fee; and $\beta^j \in (0, 1)$ denote the variable percentage fee charged to a product j customer who does not show up. Thus, if a customer pays p dollars for product j , he/she obtains a refund of $p(1 - \beta^j) - c^j$. We assume that the quantity refunded is nonnegative. If necessary, additional components of type i can be secured at a unit cost ρ_i . Given a nonanticipating intensity control policy $\lambda_s = (\lambda_s^1, \dots, \lambda_s^n)$ based on the initial inventories $x = (x^1, \dots, x^n)$ and the current history of reservations, the number of reservations for product j , N_s^j is a stochastic Poisson process with random intensity $\int_0^s \lambda_u^j du$. Let $\{T_k^j: k \geq 1\}$ denote the jump points of the counting process N_s^j , $0 \leq s \leq t$. Then the revenue collected from sales is given by

$$\int_0^t \sum_j p^j(\lambda_s^j, s) dN_s^j = \sum_j \sum_{k \geq 1} p^j(\lambda_{T_k^j}^j, T_k^j) \mathbf{1}(T_k^j \leq t),$$

where $\mathbf{1}(A)$ is the indicator function of the event A .

To account for no-shows, let $\{Z_k^j: k \geq 1\}$ be a sequence of independent Bernoulli random variables taking value 1 with probability γ^j and taking value 0 with probability $1 - \gamma^j$. We assume that these random variables are also independent of the counting processes N_s^j .

In our model, revenues are collected as reservations are made and later refunded to customers who do not show up. If we disregard the time value of money, the sales revenue net of refunds can be written as

$$\sum_j \sum_{k \geq 1} p^j(\lambda_{T_k^j}^j, T_k^j) \mathbf{1}(T_k^j \leq t) \mathbf{1}(Z_k^j = 1).$$

Thus, we can write the expected revenue net of refunds as

$$\begin{aligned} E \sum_j \gamma^j \sum_{k \geq 1} p^j(\lambda_{T_k^j}^j, T_k^j) \mathbf{1}(T_k^j \leq t) \\ = E \int_0^t \sum_j \gamma^j p^j(\lambda_s^j, s) dN_s^j. \end{aligned}$$

If variable fees are imposed, the expected revenue net of refunds can be obtained from the last expression by replacing γ^j by $\gamma^j + \beta^j(1 - \gamma^j)$. If fixed fees are also imposed, revenues are increased by

$$\sum_j c^j \sum_{k \geq 1} (T_k^j \leq t) \mathbf{1}(Z_k^j = 0).$$

Taking expectations, this results in

$$E \int_0^t \sum_j c^j (1 - \gamma^j) dN_s^j.$$

The number of components of type i necessary to satisfy the demands for reservations that show up is

$$\sum_j a_{ij} \sum_{k \geq 1} (T_k^j \leq t) \mathbf{1}(Z_k^j = 1) = \sum_j a_{ij} \int_0^t d\bar{N}_s^j,$$

where \bar{N}_s^j is a stochastic Poisson process with random intensity $\gamma^j \int_0^s \lambda_u^j du$. If $\sum_j a_{ij} \bar{N}_t^j > x_i$ we must purchase additional type i components at unit cost ρ_i . Under this model, the expected net revenue under a nonanticipating Markovian policy u is

$$\begin{aligned} J^u(x, t) = E_u \int_0^t \sum_j (\gamma^j + \beta^j(1 - \gamma^j)) p^j(\lambda_s^j, s) \lambda_s^j ds \\ + E_u \int_0^t \sum_j c^j (1 - \gamma^j) \lambda_s^j ds \\ - E_u \sum_i \rho_i \left(\sum_j a_{ij} \bar{N}_t^j - x_i \right)^+. \end{aligned} \quad (20)$$

As before, let

$$J^*(x, t) = \sup_{u \in \mathcal{U}} J^u(x, t).$$

The deterministic problem can be written as

$$\begin{aligned} J^d(x, t) \\ = \max_{\lambda(s)} \left\{ \int_0^t \sum_j (\gamma^j + \beta^j(1 - \gamma^j)) p^j(\lambda^j(s), s) \lambda^j(s) ds \right. \\ + \int_0^t \sum_j c^j (1 - \gamma^j) \lambda^j(s) ds \\ \left. - \sum_i \rho_i \left(\sum_j a_{ij} \gamma^j \int_0^t \lambda^j(s) ds - x_i \right)^+ \right\}. \end{aligned}$$

Theorem 4. For the overbooking and no-show problems defined above,

$$J^*(x, t) \leq J^d(x, t).$$

Proof. Proceed by applying Jensen's inequality to the third term of $J^d(x, t)$ in (20) and by viewing the integrand inside the expectation as purely a function of λ and maximizing pointwise. The result then follows as before. \square

Again, the solution to the deterministic problem suggests two heuristics for the stochastic problem. Let $\lambda_d(s)$ be the vector of optimal deterministic intensities, and set $\alpha^j = \int_0^t \lambda_d^j(s) ds$, and $z^j = \lfloor \alpha^j \rfloor$. The make-to-stock (MTS) heuristic accepts up to z^j requests for product j . The make-to-order (MTO) heuristic accepts all requests. Both heuristics can be shown to be asymptotically optimal, though in the interest of space we shall provide only the theorem and proof for the MTO heuristic. The next theorem shows that the revenue loss of the MTO heuristic is $O(\sum_i \sqrt{\alpha(S_i)})$. Since $J^d(x, t)$ is $O(\sum_i \alpha(S_i))$, this result together with Theorem 4 implies asymptotic optimality of the MTO heuristic.

Theorem 5.

$$J^{\text{MTO}}(x, t) \geq J^d(x, t) - \frac{1}{2} \sum_i \bar{w}^i \sqrt{\alpha(S_i)},$$

where $\bar{w}^i = \rho_i \max\{a_{ij} \sqrt{\gamma^j} : j \in S_i\}$.

Proof. The expected value of the make-to-order heuristic is given by

$$\begin{aligned} J^{\text{MTO}}(x, t) &= \int_0^t \sum_j (\gamma^j + \beta^j(1 - \gamma^j)) p^j(\lambda_d^j(s), s) \lambda_d^j(s) ds \\ &\quad + \int_0^t \sum_j c^j(1 - \gamma^j) \lambda_d^j(s) ds \\ &\quad - E \sum_i \rho_i \left(\sum_j a_{ij} \hat{N}_t^j - x_i \right)^+, \end{aligned} \quad (21)$$

where \hat{N}_t^j is a Poisson process with deterministic intensity $\int_0^s \gamma^j \lambda_d^j(u) du$. Note that the first two terms of $J^{\text{MTO}}(x, t)$ are equal to those of $J^d(x, t)$.

To establish the asymptotic optimality of the MTO heuristic we need a slight variant of (18). Let D be a random variable with mean μ and variance σ^2 , writing $(D - d)^+ = \frac{1}{2}(|D - d| + (D - d))$, taking expectations and using the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} E(D - d)^+ &\leq \frac{1}{2} \sqrt{\sigma^2 + (\mu - d)^2} + \frac{1}{2} (\mu - d) \leq \frac{1}{2} \sigma \\ &\quad + \frac{1}{2} (|\mu - d| + (\mu - d)) = \frac{1}{2} \sigma + (\mu - d)^+. \end{aligned}$$

Applying this to the last term of $J^{\text{MTO}}(x, t)$ in (21) we obtain

$$\begin{aligned} E \sum_i \rho_i \left(\sum_j a_{ij} \hat{N}_t^j - x_i \right)^+ &\leq \frac{1}{2} \sum_i \rho_i \sqrt{\sum_j a_{ij}^2 \gamma^j \alpha^j} \\ &\quad + \sum_i \rho_i \left(\sum_j a_{ij} \gamma^j \alpha^j - x_i \right)^+, \end{aligned}$$

which when substituted for the third term of (21) and using the definition of \bar{w}^i proves Theorem 5. \square

7. NUMERICAL EXAMPLES

We next examine several example applications of our model and the results of some numerical experiments. Because the model presented in the previous sections is quite general, there are many problems that one can investigate using it. We have selected only a representative sample, each one chosen to illustrate a specific problem characteristic. In general, one could combine the characteristics of these examples to obtain increasingly realistic formulations. In addition to illustrating model formulations, the examples also provide insights into some fundamental characteristics of yield management problems.

Example 1 examines a network problem that is time homogeneous, has no product differentiation and has no constraints on the O-D prices. Example 2 looks at this same network with two well-differentiated fare class products for each O-D pair. In this example, we assume the differentiation is not based on time, so fare class products are sold simultaneously. Finally, Example 3 looks at this same network when overbooking and no-shows are allowed. A common finding in all these cases is that using correctly chosen fixed prices along with a simple allocations scheme like FCFS (the MTO heuristic) is quite close to optimal.

These findings would seem to suggest that current yield management, with its multiple fares and booking limit allocation schemes, is unlikely to be more effective than simple FCFS allocation. However, we next look at several scenarios in which these practices are indeed effective. The first (Example 4) is when, as a matter of practical necessity, fares for each problem instance cannot be set freely, as was assumed in Examples 1–3. The second case (Example 5) is when the demand function changes over time, either due to changes in the perceived value of the products or through time-of-purchase product differentiation. The final case (Example 6) is when products are mispriced. In all these cases, we show that a multiple-price, dynamic-allocation scheme can indeed be quite effective. The examples help explain more precisely much of the phenomenon that yield management systems are exploiting.

7.1. Airline Network Pricing Examples

Consider the following network model. A carrier network is described by a directed graph $G = (V, E)$, where V is the set of n cities serviced by the carrier and an edge $(i, j) \in E$ represents a scheduled flight on a leg from city i to city j . We let c_{ij} denote the capacity of edge (i, j) , i.e., the number of seats available on this flight leg. All capacities are defined with respect to a given day t . At each time s , $0 \leq s \leq t$, the carrier can set the $n(n - 1)$ prices, p_s^{ij} , for all possible origin-destination pairs i and j , $i \neq j$. These

prices induce demand intensities for the various origin-destination (O-D) pairs through demand functions $\lambda^{ij}(p^{ij})$. There is a fixed path (itinerary) π_{ij} , indexing a subset of E that forms a directed path from i to j , associated with each O-D pair i, j . We shall index O-D pairs by π . The objective is to adjust origin-destination prices and allocate leg capacity among the O-D pairs over time in order to maximize the expected revenue received prior to the departure time t .

We note that several generalizations of this formulation are possible. In particular, there is no need to have a unique path connecting each pair of nodes. An arbitrary (or the complete) set of paths between nodes could be offered and priced either individually or jointly as discussed in Section 4.3. Indeed, the paths for all our examples were in fact selected by solving a problem that initially makes available all possible paths between nodes. For simplicity, however, we assume a unique path connects each pair of nodes.

In our network examples, we use a revenue function that is separable and time homogeneous,

$$r(\lambda) = \sum_{\pi} \lambda_{\pi} p_{\pi}(\lambda_{\pi}).$$

A log-linear demand function of the form

$$\lambda(p) = \lambda_0 e^{-\epsilon_0(p/p_0 - 1)},$$

was used for each itinerary π where p_0 is interpreted as a reference price, λ_0 is the demand rate at the reference price, and ϵ_0 is the magnitude of the elasticity of demand at the reference price. This log-linear demand function has a unique inverse

$$p(\lambda) = (\epsilon_0^{-1} \ln(\lambda_0/\lambda) + 1)p_0.$$

There is a column in the bill-of-materials matrix A for each itinerary π and rows of A correspond to flight legs (edges). An element $a_{\pi e}$ of A is one if the itinerary π uses edge e and is zero otherwise. With this identification, the problem reduces to an instance of the multiproduct problem. The deterministic version of this problem is a separable nonlinear program. We emphasize that the separability and time homogeneity assumptions are imposed to simplify the numerical calculations only; they are not required for the theoretical results of the previous sections to hold.

7.1.1. Example 1: Undifferentiated Network. Our first example illustrates the basic pricing problem with no price constraints and no product differentiation. It is based on the six-node example of a hypothetical airline network shown in Figure 1. Nodes 2 and 3 are “hub” nodes. Leg seat capacities are shown and were chosen to approximate the number of seats on a single aircraft.

Parameter values for the demand functions of all O-D pairs are shown in Table I along with the path (itinerary) used by each pair. The values shown in Table I are essentially arbitrary and were chosen merely to illustrate the performance of the heuristics; however, it is not hard to see that the revenue of the heuristics is affected only by the

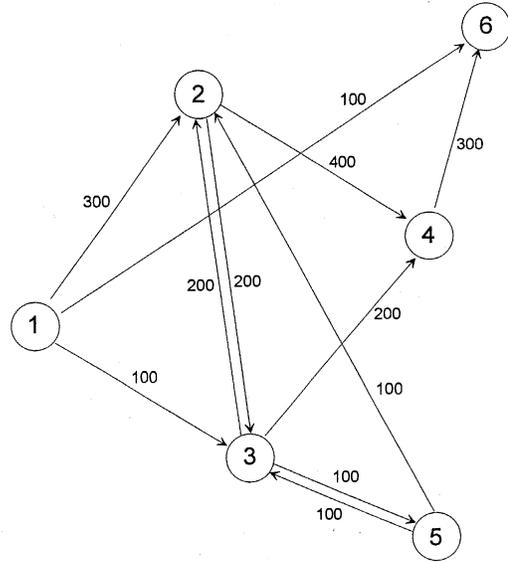


Figure 1. A six-node, two-hub airline network.

deterministic O-D prices and seat allocations, which are shown in the last two columns in Table I. These prices and seat allocations are reasonable approximations of those found in actual airline applications.

The deterministic problem was formulated and solved using a nonlinear programming routine. The resulting prices and seat allocations are given in Table I. The solution resulted in a total revenue of \$661,200 across all O-D pairs.

We next simulated both the MTS and MTO heuristic derived from this deterministic solution. To illustrate their

Table I
Demand Function Parameters, Itineraries, and Optimal Deterministic Solution for Example 1

Market		Demand Function			Path	Opt. Det. Sol.	
O	D	λ_0	ϵ_0	p_0		# Seats	Price
1	2	300	1.0	220	1-2	135	\$396.62
1	3	300	1.2	220	1-3	67	\$495.86
1	4	300	2.0	400	1-2-4	165	\$520.11
1	5	300	1.0	250	1-3-5	33	\$752.04
1	6	300	0.8	200	1-6	100	\$525.58
2	3	300	1.0	230	2-3	168	\$364.28
2	4	300	0.9	200	2-4	143	\$365.74
2	5	300	2.0	200	2-3-5	32	\$423.79
2	6	300	1.0	200	2-4-6	92	\$436.80
3	2	300	1.0	200	3-2	200	\$281.76
3	4	300	2.0	230	3-4	131	\$325.30
3	5	300	2.0	120	3-5	35	\$249.51
3	6	300	2.0	150	3-4-6	14	\$378.60
4	6	300	1.0	150	4-6	162	\$243.30
5	2	300	1.0	200	5-2	100	\$420.39
5	3	300	2.0	150	5-3	47	\$289.90
5	4	300	1.0	160	5-3-4	21	\$585.20
5	6	300	1.0	230	5-3-4-6	32	\$748.50

Table II
Simulation Results for Example 1

Problem Instance	MTO			MTS	
	Upper Bound (\$)	Revenue (\$)	% UB	Revenue (\$)	% UB
0.1	66,120	58,001	87.72%	54,510	82.44%
0.5	330,600	311,767	94.30%	306,690	92.77%
1.0	661,200	636,858	96.32%	629,875	95.26%
2.0	1,322,400	1,280,404	96.82%	1,272,904	96.26%
10.0	6,612,000	6,542,195	98.94%	6,517,793	98.58%

performance over a range of demand volumes, the common value of $\lambda^0 = 300$ in Table I as well as the leg capacities were scaled as shown in Table II. For example, a scale factor of 2 corresponds to leg capacities that are twice those shown in Figure 1 and a value of $\lambda_0 = 600$ rather than 300 for all demand functions in Table I. This scaling does not change the optimal prices shown in Table I. In practice, such an increase in volume may occur when aggregating the capacity of several flights.

Both heuristics were simulated for each instance, and the resulting expected revenue was estimated. All point estimates for expected revenues have a relative error of $\pm 1\%$ with 95% confidence. Table II shows both the dollar revenue and the revenue as a percentage of the deterministic upper bound.

The simulation results reveal several interesting insights. First, observe that the revenue from the MTO heuristic dominates that of the MTS heuristic in all cases. Thus, it appears that if prices are set correctly, “protecting” space for specific O-D pairs, as is done in the MTS heuristic, is not effective; the simple FCFS allocation of the MTO heuristic is better. Second, even relative to an optimal policy the performance of MTO is surprisingly good. Indeed, in the original problem (scale factor 1.0), the results show that at most 3.68% (in expectation) additional revenue could possibly be captured by using a scheme more sophisticated than a fixed-price policy based on the deterministic model together with a FCFS allocation (MTO)—and the potential increase could be even less, since this figure is only an upper bound on the optimality gap. Nevertheless, for most commercial airlines a 3–4% potential increase in revenues is a significant amount of money and would likely justify more sophisticated pricing and/or allocation schemes.

One way to capture some of the additional potential expected revenue is to find the *fixed* prices that maximize the expected revenue subject to a prespecified allocation scheme such as FCFS; see Gallego and van Ryzin for a discussion on optimal fixed prices for the stochastic problem. An alternative way to capture additional expected revenue is to find more sophisticated allocation policies given a fixed pricing policy.

7.1.2. Example 2: Differentiated Network. Suppose a well-differentiated super-saver and full-coach product exists for

each O-D pair of the network from Example 1. For simplicity, we assume that these products are differentiated by travel restrictions, cancellation policies or other mechanisms not related to the time of purchase. (Time-dependent restrictions are discussed in Section 7.1.4.) In this way, sales of the two products occur concurrently throughout the time horizon. Sequential sales, as is usually assumed in airline yield management models, or partial overlapping of sales could be modeled by making each product’s demand function time dependent.

We model product differentiation using virtual nodes at each city i to represent the demand from each market segment. These virtual nodes are then connected to the physical node i via infinite capacity links. Thus, the virtual nodes “compete” for the same physical capacities on the legs of the network. Figure 2 shows the network of Figure 1 modified using virtual nodes to account for two classes of demand originating at each node. The revenue function is separable and demand functions are the same log-linear form as in Example 1. Data function parameters, itineraries and the optimal deterministic prices and sales are shown in Table III. The magnitude of the super-saver demand elasticities is either 3.0 or 3.5, and this demand has a higher intensity (λ_0) for each O-D pair; full-coach demand has an elasticity of magnitude 0.5 and a lower intensity for each O-D pair. Again, these values were chosen merely to illustrate the model; further, they are not related to the values from Example 1. In addition, the paths for each O-D pair are also not necessarily related to those in Example 1.

The simulation results for a series of scaled versions of this problem are shown in Table IV. As in Example 1, the

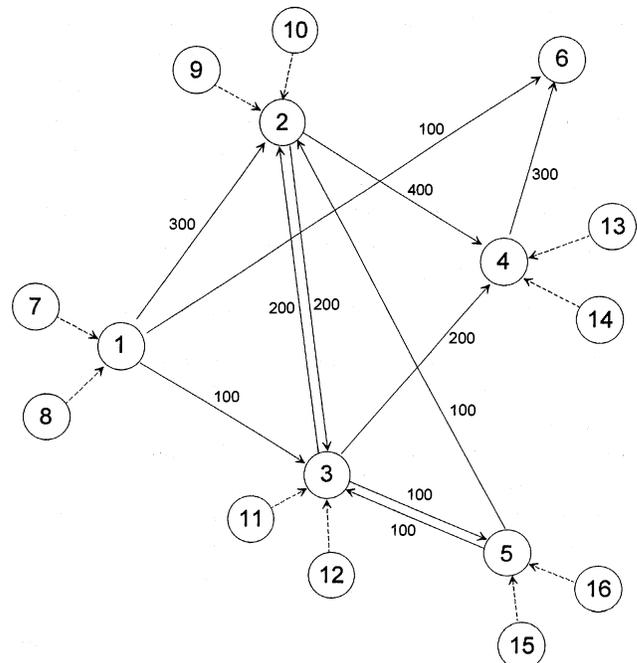


Figure 2. Example with two fare classes at each origin (dashed lines are infinite capacity edges).

Table III
Demand Function Parameters, Itineraries, and Optimal Deterministic Solution for Example 2

Market		Demand Function			Opt. Det. Sol.		
O	D	λ_0	ϵ_0	p_0	Path	# Seats	Price
7	2	30	0.5	320	7-1-2	15	\$ 749.17
7	3	30	0.5	320	7-1-3	14	\$ 804.34
7	4	30	0.5	400	7-1-2-4	14	\$1001.79
7	5	30	0.5	300	7-1-6	14	\$ 866.25
7	6	30	0.5	350	7-1-3-5	11	\$ 879.32
8	2	200	3.0	160	8-1-2	197	\$ 160.89
8	3	200	3.0	160	8-1-3	69	\$ 216.85
8	4	200	3.0	200	8-1-2-4	74	\$ 266.44
8	5	200	3.0	150	8-1-6	85	\$ 224.58
8	6	200	3.0	175	8-1-3-5	6	\$ 328.55
9	3	30	0.5	330	9-2-3	15	\$ 775.23
9	4	30	0.5	300	9-2-4	16	\$ 692.18
9	5	30	0.5	300	9-2-3-5	12	\$ 830.13
9	6	30	0.5	300	9-2-4-6	14	\$ 740.35
10	3	150	3.5	165	10-2-3	164	\$ 160.92
10	4	150	3.5	150	10-2-4	213	\$ 135.04
10	5	150	3.5	150	10-2-3-5	9	\$ 271.67
10	6	150	3.5	150	10-2-4-6	69	\$ 183.21
11	2	30	0.5	300	11-3-2	16	\$ 674.87
11	4	30	0.5	300	11-3-4	16	\$ 681.95
11	5	30	0.5	250	11-3-5	14	\$ 615.04
11	6	30	0.5	150	11-3-4-6	12	\$ 429.95
12	2	100	3.0	150	12-3-2	165	\$ 124.87
12	4	100	3.0	150	12-3-4	144	\$ 131.64
12	5	100	3.0	125	12-3-5	47	\$ 156.71
12	6	100	3.0	75	12-3-4-6	4	\$ 154.80
13	6	40	0.5	150	13-4-6	21	\$ 348.17
14	6	90	3.5	90	14-4-6	168	\$ 73.88
15	2	30	0.5	300	15-5-3-2	12	\$ 838.75
15	3	30	0.5	250	15-5-3	13	\$ 663.88
15	4	30	0.5	260	15-5-3-4	11	\$ 765.79
15	6	30	0.5	230	15-5-3-4	10	\$ 753.93
16	2	100	3.0	150	16-5-3-2	6	\$ 288.75
16	3	100	3.0	175	16-5-3	45	\$ 222.21
16	4	100	3.0	130	16-5-3-4	3	\$ 288.85
16	6	100	3.0	115	16-5-3-4-6	0	\$ 298.45

results show that the both the MTO and MTS heuristics are effective with large sales volumes (scale factor 10.0), but again MTO dominates MTS in all instances. Note the relative performance of the MTS heuristic is somewhat worse than in Example 1. This suggests that on problems with a higher level of detail (e.g., many itineraries and fare

Table IV
Simulation Results for Example 2

Problem Instance		MTO		MTS	
Scale	Upper Bound (\$)	Revenue (\$)	% UB	Revenue (\$)	% UB
0.1	41,070	32,048	78.03%	27,899	67.93%
0.5	205,350	191,595	93.30%	180,722	88.01%
1.0	410,700	392,385	95.54%	375,628	91.46%
2.0	821,400	793,262	96.57%	771,564	93.93%
10.0	4,107,000	4,047,899	98.56%	4,006,430	97.55%

Table V
Overbooking Charges for Example 3

Leg (Edge)	Charge
(1, 2)	\$576.6
(1, 3)	\$812.5
(1, 6)	\$775.6
(2, 3)	\$634.3
(2, 4)	\$643.5
(3, 2)	\$581.2
(3, 4)	\$710.3
(3, 5)	\$689.5
(4, 6)	\$593.3
(5, 2)	\$717.5
(5, 3)	\$714.9

classes) the performance of the MTS heuristic may suffer. This observation is consistent with our theoretical bounds.

7.1.3. Example 3: Overbooking and No-Shows. We next consider the effect of overbooking and no-shows on the network in Figure 1. We assume a uniform probability of a customer showing up of γ for all O-D pairs. To make comparisons to Example 1 easier, we assume the demand data are those given in Table I except that the common value of $\lambda_0 = 300$ is multiplied by the factor $1/\gamma$. In this way the expected demand net of refunds has the same value as in Example 1 for any given set of prices. We further assume that the airline collects no revenue from those customers who do not show up and that the overbooking charges imposed for each leg are the values shown in Table V. These overbooking charges are the penalties/costs paid when the airline is unable to accommodate a customer. These values are sufficiently large that there is no incentive to have net expected demand exceed leg capacities.

Together, these assumptions imply the following implementation of the heuristics. First, the optimal prices in Table I remain the same. For the MTS heuristic, the allocations in Table I are scaled up by a factor $1/\gamma$ and reservations are accepted for each itinerary up to this limit. In the MTO heuristic, all reservations are accepted. In either case, each reservation shows up with probability γ , and any excess demand for leg capacities must be satisfied at the costs shown in Table V.

This example was run for a variety of values for γ and the results are shown in Table VI. Again, all figures have a

Table VI
Simulation Results for Example 3

Problem Instance		MTO		MTS	
γ	Upper Bound (\$)	Revenue (\$)	% UB	Revenue (\$)	% UB
1.0	661,200	636,858	96.32%	629,875	95.26%
0.95	661,200	622,347	94.12%	627,898	94.96%
0.90	661,200	623,717	94.33%	631,570	95.52%
0.80	661,200	627,274	94.87%	629,106	95.15%
0.60	661,200	616,295	93.21%	625,321	94.57%

relative error of $\pm 1\%$ at a 95% confidence level. Note that the revenue is only about 2% lower than that in Example 1, except for $\gamma = 0.6$ where it is about 3% lower. What is rather surprising is that the revenue does not go down appreciably as the probability of showing up decreases; not until $\gamma = 0.6$ is there a statistically significant drop in revenue.¹

Also note that, unlike Examples 1 and 2, the MTS heuristic in this case appears to be slightly better than the MTO heuristic.

We note that scaling up allocations by $1/\gamma$ is a common strategy used by many airlines. The use of scaling factors (not necessarily $1/\gamma$) is described in Belobaba (1987a), and its accuracy as an approximation to optimal overbooking levels is discussed at some length in McGill (1989).

Though this is only one example, it again suggests that a simple deterministic correction, in this case scaling up the allocation limits by $1/\gamma$, is surprisingly effective. The stochastic effects of no-shows appear to produce only a minor drop in revenue. Indeed, for this example one can show that imposing even a modest fee of \$25 for not showing up more than compensates for the lost revenue due to overbooking charges. This is due to the fact that, roughly speaking, no-show fees grow linearly with the number of reservations while the overbooking charges grow proportional to the square root of the number of reservations. Thus, for large scale problems, modest no-show fees can effectively offset quite large overbooking charges.

7.1.4. Example 4: Time Varying Demand. In this example we examine the effect of a time varying demand function on a simple, one-product problem. Unlike the previous examples where a single fixed price is used for each product, in this case different prices are offered in response to a temporal change in demand.

Consider a single leg problem with an initial capacity of $x = 525$ seats. Let $t = 1$ be the sales horizon, and assume that demand at time $s \in (0, 1]$ is given by the linear function

$$\lambda(p, s) = a(s) - b(s)p,$$

where the time-dependent coefficients $a(s)$, and $b(s)$ are piecewise constant as shown in Table VII.

To solve the deterministic problem, one first computes the intensity that maximizes the revenue rate over each interval. If the resulting intensities are feasible, then they are optimal. Otherwise, the optimal intensities equalize the marginal revenue rates of both time intervals and

exhaust the capacity. For this example, the solution to the deterministic problem is given by

$$p_d(s) = \begin{cases} 200, & 0 \leq s \leq 0.75, \\ 300, & 0.75 < s \leq 1, \end{cases}$$

and

$$\lambda_d(s) = \begin{cases} 600, & 0 \leq s \leq 0.75, \\ 300, & 0.75 < s \leq 1. \end{cases}$$

Notice that the optimal deterministic price increases from \$200 to \$300 at time 0.75. Using the pricing policy that is optimal for the deterministic problem, we simulated the performance of three heuristics. The MTO heuristic accepts all demands until the inventory is exhausted. The MTS heuristic allocates $450 = \int_0^{0.75} \lambda_d(s) ds$ units for sales over $[0, 0.75]$ and $75 = \int_{0.75}^1 \lambda_d(s) ds$ units for sales over $(0.75, 1]$.

Finally, we consider a refinement of the MTS heuristic that allows unsold units allocated for sales during $[0, 0.75]$ to be made available for sale during $(0.75, 1]$. Thus, if under the MTS heuristic 448 (out of the 450 reserved under the MTS heuristic) units are sold during $[0, 0.75]$, then 77 instead of 75 units are made available for sale during $(0.75, 1]$. As an additional refinement, sales at \$300 are allowed to start before time 0.75 if the 450 reserved by the MTS heuristic to be sold for \$200 over $[0, 0.75]$ are sold before time 0.75. Thus, if under the MTS heuristic 450 units are sold by time 0.65 then sales at \$300 start at time 0.65. Of course, demand over the time interval $[0.65, 0.75]$ consists of people who are willing to pay the higher fare. We call this the booking limit (BL) heuristic. The refinement of the BL heuristic over the MTS heuristic improves revenue by opening the high fare early as a consequence of brisk sales at the lower fare, and by allowing sales at the high fare of inventory that was not sold at the low fare. The BL heuristic is similar in spirit to traditional yield management policies, e.g., Brumelle and McGill, that protect inventory for sale at higher fares.

The columns of Table VIII under the headings MTO, MTS, and BL present the performance of these allocation policies in absolute and relative terms. The first column of Table VIII gives the deterministic upper bound. Notice that all heuristics performed well, with the BL heuristic having a slight advantage. Unlike most of the network examples, it appears that protecting space works better than a FCFS allocation. However, the effect of the allocation scheme is relatively minor (0.6% improvement in BL over MTO), again suggesting that if prices are set correctly, FCFS allocation is still quite good.

Table VII
Coefficients of Linear Demand Function
for Example 4

$0 < s \leq 0.75$	$0.75 < s \leq 1$
$a(s) = 1200$	$a(s) = 600$
$b(s) = 3$	$b(s) = 1$

Table VIII
Performance of Heuristics on Example 4

	Upper Bound	MTO	MTS	BL
Rev. (\$)	112,500	109,586.25	109,597.50	110,283.75
% UB	100%	97.4%	97.6%	98.0%

7.2. Fixed Allocation and Booking Limit Examples

The previous examples show that in many cases allocation policies and booking limits have only a marginal impact on revenue, provided prices are set correctly. In particular, a simple FCFS allocation works nearly as well as fixed (MTS) or booking limit (BL) allocation schemes. This seemingly contradicts the fact that, in practice, booking limit policies typically produce significant revenue increases. Indeed Smith et al. report a 5% increase in revenue for American Airlines, worth approximately \$1.4 billion dollars over a three-year period, attributable to effective yield management. In this section, we propose two explanations for this discrepancy: (1) price flexibility may be limited, and (2) prices may be set incorrectly (misspricing occurs). In both cases, we show that using booking limits can significantly improve revenues.

7.2.1. Example 5: Limits on Price Flexibility. A key assumption in the previous examples is that prices could be chosen without constraints. If each instance of a yield management problem corresponds to a particular day or even a particular set of flights within a day, then effectively we are assuming that each day or individual flight can be priced separately. In practice, this degree of price flexibility is rarely achievable; it would be quite confusing for customers, make advertising difficult, and complicate the airlines' own planning and reservation systems. Indeed, most airlines offer prices that typically remain in effect for several weeks at a time. Underlying demand, however, often changes dramatically from day to day or even within a day.

One mechanism for increasing revenues under such volatile conditions is precisely to offer several prices for each product and then vary the allocation of space to these prices in response to changes in demand. In this way, one retains the advantage of a stable set of prices while achieving some ability to respond to short-term fluctuations in demand; in a sense, varying the allocation provides a means of *synthesizing* a continuous range of prices from a relatively small set of fare classes.

A simple, deterministic example illustrates this effect. Consider a single leg flight with a capacity of 100 seats that operates once each day of the week. There is a fixed sales horizon t for each day, and without loss of generality, we assume $t = 1$. Total demand for each day varies and has an exponential form $\lambda(p) = \lambda_0 e^{-\epsilon_0(p/p_0-1)}$. The reference price and elasticity are assumed fixed at $p_0 = \$300$ and $\epsilon_0 = 3$, respectively. For simplicity, we assume that the only parameter that varies from one day to the next is λ_0 , the demand at the reference price p_0 . Demand during the week is divided into heavy, medium, and light days. On heavy days (Monday and Friday), $\lambda_0 = 125$; on medium days (Tuesday to Thursday), $\lambda_0 = 50$; and on light days (Saturday and Sunday), $\lambda_0 = 25$.

If one had complete flexibility in choosing prices, each type of day could be priced separately, i.e., one would treat each day as an independent instance and apply a different pricing policy to it. We shall call this option the FLEX

Table IX
Demand Data and Revenues: Example 5

	Light	Medium	Heavy	Wkly Total
# Days/Wk	2	3	2	
λ_0	25	50	125	
Opt. Price	\$ 161	\$ 231	\$ 322	
Rev. FLEX	\$16,100	\$23,100	\$32,200	\$165,900
Rev. SP	\$ 9,874	\$19,768	\$25,700	\$130,452
Rev. BL	\$16,100	\$20,125	\$32,200	\$156,975

pricing policy. Table IX shows the optimal prices to charge on each type of day along with the daily and weekly revenues obtained using the FLEX prices. (All revenues in Table IX are based on deterministic models.)

Since offering a separate price for each day of the week may not be feasible in practice, two alternative options are considered: (1) offer only a single fixed price for the entire week (SP policy); and (2) offer two fares (discount and full), again fixed for the entire week, but vary the allocation of seats to these fares on a daily basis. We call this latter option the booking limit (BL) policy.

Of course, we would like to use the best possible implementation of these two policies. For the SP policy, one can show that the best price to offer is \$257. For the BL policy, we used the FLEX price for light days (\$161) as the discount price, and the FLEX price for heavy days (\$322) as the full price. Table X shows the demand if each of these fare levels was used exclusively over the horizon and gives the optimal allocation of seats to these fare classes. (See Gallego and van Ryzin for a detailed discussion of the optimal policy for this type of discrete price problem.)

Table IX show the daily and weekly revenues obtained from the optimal implementations of the SP and BL policies. Note that the weekly revenue of the SP policy is substantially less than that obtained from using the FLEX policy. The loss in revenue in this case is about 21%. The BL policy revenue loss relative to the FLEX policy, at about 5%, is significantly less than that for the SP policy. Thus, we see that in this example booking policies truly provide substantial increases in revenue; they allow one to make up approximately half the gap between the revenue of a pure fixed price policy and a completely flexible pricing policy.

This example illustrates how one can use a limited number of price classes together with a dynamic allocation scheme to recapture some of the revenue achievable

Table X
Booking Limit Demands and Allocations: Example 5

	Light	Medium	Heavy
Demand at Disc.	100	201	502
Demand at Full	20	40	100
Disc. Alloc.	100	75	0
Full Alloc.	0	25	100

Table XIPerformance of Heuristics Based on Pricing Policy-*a*

	UB	MTO	MTS	BL
Rev. (\$)	\$90,000	\$78,848	\$73,002	\$88,994
% Det.	100%	87.6%	88.1%	98.8

through flexible pricing without introducing an extremely complicated price structure.

7.2.2. Example 6: Misspricing. Another case in which an allocation policy can have a significant impact is when misspricing occurs. Consider again the time varying demand function of Example 4 with a capacity of 525 seats. Suppose that the decision maker mistakenly underestimates the parameters $a(s)$ over $(0, 0.75)$ to be 900 instead of 1200.

The optimal price for the deterministic problem, based on the estimated expected demand, is obtained by setting the marginal revenue rates equal to zero in each of the two time intervals. This results in

$$p_a(s) = \begin{cases} 150, & 0 \leq s \leq 0.75, \\ 300, & 0.75 < s \leq 1. \end{cases}$$

The estimated expected demand rate at p_a is

$$\lambda_a(s) = \begin{cases} 450, & 0 \leq s \leq 0.75, \\ 300, & 0.75 < s \leq 1. \end{cases}$$

Consequently, the deterministic solution based on the *estimated* expected demand, allocates $337 \approx 450 \times 0.75$ seats for sales at \$150 over the interval $(0, 0.75]$ and $75 = 300 \times 0.25$ seats for sales at \$300 over the interval $(0.75, 1]$. Call this pricing/allocation policy, Policy-*a*.

The MTO heuristic accepts all requests at \$150 over $(0, 0.75]$, and all requests at \$300 over $(0.75, 1]$ on a first-come-first-serve basis until the capacity is exhausted. The MTS heuristic reserves 337 seats for sales at \$150, and 75 seats for sales at \$300.

The BL heuristic reserves 75 seats for sales at \$300, but makes all unsold units at time 0.75 available for sales at \$300 during $[0.75, 1]$. See Example 4 for a more complete description of the BL heuristic.

Notice the *actual* expected demand rate at p_a is

$$\lambda(s) = \begin{cases} 750, & 0 \leq s \leq 0.75, \\ 300, & 0.75 < s \leq 1. \end{cases}$$

An upper bound on the revenue under Policy-*a* is obtained by finding the best allocation of space assuming that demand is deterministic. This calculation yields $\$90,000 = 450 \times \$150 + 75 \times \$300$. Table XI reports the performance of the MTO, MTS and BL heuristics under pricing Policy-*a*.

For Policy-*a*, the MTO and the MTS heuristics perform poorly. Under the MTO heuristic virtually all the capacity is sold at the promotional fare of \$150 per seat. Under the MTS heuristic only 412 seats are made available for sale. In contrast, the BL heuristic performs near optimally capturing 98.8% of the deterministic upper bound. This is because, unlike the MTO heuristic, the BL heuristic reserves 75

seats for the high fare, and unlike the MTS heuristic, it makes all the capacity available for sale. The average number of unsold seats was zero under the MTO heuristic, 113 under the MTS heuristic, and 3 under the BL heuristic.

This example illustrates the sensitivity of both the MTO and the MTS heuristics to pricing errors and shows that performance can be improved by protecting space from excess low fare demand. One can show that if the parameters of $a(s)$ are overestimated in this example, the MTO and MTS heuristics are less sensitive to the resulting pricing errors. In this case, higher than optimal prices are charged, and hence low fare sales do not consume space needed for the high fare demand.

8. CONCLUSIONS

We have shown how a rich class of revenue management problems can be modeled using a single, unified framework. For this large class of problems, our results show that the pricing policies derived from deterministic models are quite close to optimal.

We also showed that when prices are chosen based on these deterministic models, the effect of allocation schemes appears to be relatively minor. These are, in a way, negative results and suggest that yield management is relatively ineffective when pricing decisions are made correctly. In practice, yield management's main benefit may be to compensate for prices that are not optimally matched to demand, rather than to exploit revenue opportunities due to second-order, stochastic effects. In addition, the practice may—and probably does—serve as a simple mechanism for synthesizing a wide variety of prices using only a limited number of relatively stable fare classes. We believe such insights will prove to be quite beneficial in guiding the development of better yield management systems in practice. Further, our bounds provide a means to evaluate the relative benefit of a given system.

There are certainly many possible directions for future research in this area. We have focused exclusively on airline applications; no doubt there are other interesting applications, such as the multi-day hotel problem, that could be investigated using a similar approach. A more challenging problem is to try to combine demand function estimation with pricing/allocation decisions. This would reflect the more realistic case where demand functions are not directly observable. Also, understanding the role that competitors play by explicitly modeling their networks and pricing decisions is an important future research topic.

ENDNOTE

1. The fact that revenue decreases in γ over the range 0.95–0.80 is not statistically significant. Moreover, the MTO heuristic uses integer allocations, which introduces round-off errors. As a result, changing γ does not result in an exact proportional change in the allocations, which may also account for some of this slightly unusual revenue behavior.

ACKNOWLEDGMENT

This research was sponsored by National Science Foundation Grant SES-93-09394.

REFERENCES

- BELOBABA, P. P. 1987a. Air Travel Demand and Airline Seat Inventory Management. Unpublished Ph.D. Dissertation. MIT, Cambridge, MA.
- BELOBABA, P. P. 1987b. Airline Yield Management: An Overview of Seat Inventory Control. *Trans. Sci.* **21**, 63–73.
- BELOBABA, P. P. 1989. Application of a Probabilistic Decision Model to Airline Seat Inventory Control. *Opns. Res.* **37**, 183–197.
- BITRAN, G. R., AND S. M. GILBERT. 1992. Managing Hotel Reservations with Uncertain Arrivals. Working Paper, MIT Sloan School.
- BRÉMAUD, P. 1980. *Point Processes and Queues, Martingale Dynamics*. Springer-Verlag, New York.
- BRUMELLE, S. L., J. I. MCGILL, T. H. OUM, K. SAWAKI AND M. W. TRETHERWAY. 1990. Allocation of Airline Seats Between Stochastically Dependent Demand. *Trans. Sci.* **24**, 183–192.
- BRUMELLE, S. L., AND J. I. MCGILL. 1993. Airline Seat Allocation with Multiple Nested Fare Classes. *Opns. Res.* **41**, 127–137.
- CHATWIN, R. E. 1992. Optimal Airline Overbooking. Ph.D. Thesis, Stanford University, Palo Alto.
- COHEN, M. A. 1977. Joint Pricing and Ordering Policy for Exponentially Decaying Inventory with Known Demand. *Naval Res. Logist.* **24**, 257–268.
- CURRY, R. E. 1989. Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations. *Trans. Sci.* **24**, 193–204.
- DROR, M., P. TRUDEAU AND S. P. LADANY. 1988. Network Models for Seat Allocation of Flights. *Tran. Res.* **22B**, 239–250.
- FENG, Y., AND G. GALLEGO. 1995. Optimal Stopping Times for Promotional Fares and Optimal Starting Times for End-of-Season Sales. *Mgmt. Sci.* **41**, 1371–1391.
- GALLEGO, G. 1992. A Minmax Distribution Free Procedure for the (Q,R) Inventory Model. *O. R. Letts.* **11**, 55–60.
- GALLEGO, G., AND G. VAN RYZIN. 1994. Optimal Dynamic Pricing of Inventories with Stochastic Demand Over Finite Horizons. *Mgmt. Sci.* **40**, 999–1020.
- GLOVER, F., R. GLOVER, J. LORENZO AND C. MCMILLAN. 1982. The Passenger-Mix Problem in the Scheduled Airlines. *Interfaces*, **12**, 73–79.
- KARLIN, S., AND C. R. CARR. 1962. Prices and Optimal Inventory Policies. In Arrow, K. J., Karlin, S. and Scarf, H. (eds.), *Studies in Applied Probability and Management Science*. Stanford University Press, Stanford, California.
- KIMES, S. E. 1989. The Basics of Yield Management. *Cornell H.R.A. Quarterly*, **4**, 14–19.
- LADANY, S. P., AND A. ARBEL. 1991. Optimal Cruise-liner Passenger Cabin Pricing Policy. *Eur. J. Opnl. Res.* **55**, 136–147.
- LEE, T. C., AND M. HERSH. 1993. A Model for Dynamic Airline Seat Inventory Control with Multiple Seat Bookings. *Trans. Sci.* **27**, 252–265.
- LIBERMAN, V., AND U. YECHIALI. 1978. On the Hotel Overbooking Problem—An Inventory System with Stochastic Cancellations. *Mgmt. Sci.* **11**, 1117–1126.
- LITTLEWOOD, K. 1972. Forecasting and Control of Passengers. *12th AGIFORS Symposium Proceedings*. 95–128.
- LUENBERGER, D. G. 1969. *Optimization by Vector Space Methods*. John Wiley & Sons, New York.
- MCGILL, J. I. 1989. Optimization and Estimation Problems in Airline Yield Management. Unpublished Ph.D. Dissertation, University of British Columbia, Vancouver, B.C.
- ROBINSON, L. W. 1995. Optimal and Approximate Control Policies for Airline Booking with Sequential Fare Classes. Working Paper, Cornell University, *Opns. Res.* **43**, 252–263.
- ROTHSTEIN, M. 1971. An Airline Overbooking Model. *Trans. Sci.* **5**, 180–192.
- ROTHSTEIN, M. 1974. Hotel Overbooking as a Markovian Sequential Decision Process. *Dec. Sci.* **5**, 389–404.
- SMITH, B., J. LEIMKUHNER, R. DARROW AND J. SAMUELS. 1992. Yield Management at American Airlines. *Interfaces* **22**, 8–31.
- SOUMIS, F., AND A. NAGURNEY. 1993. A Stochastic, Multiclass Airline Network Equilibrium Model. *Opns. Res.* **41**, 721–730.
- TALLURI, T. K. 1993. Airline Revenue Management with Passenger Routing Control: A New Model with Solution Approaches. Working Paper, USAir, Arlington, VA.
- WANG, K. 1983. Optimum Seat Allocation for Multi-Leg Flights with Multiple Fare Types. *Proceedings of the AGIFORS 23rd Annual Symposium*, 225–246.
- WEATHERFORD, L. R., AND S. E. BODILY. 1992. A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking and Pricing. *Opns. Res.* **40**, 831–844.
- WEATHERFORD, L., S. BODILY AND P. PFEIFER. 1993. Modeling the Customer Arrival Process and Comparing Decision Rules in Perishable Asset Revenue Management Situations. *Trans. Sci.* **27**, 239–251.
- WILLIAMSON, E. L. 1992. Airline Network Seat Control. Ph.D. Thesis, MIT, Cambridge, MA.
- WOLLMER, R. D. 1986a. A Hub-Spoke Seat Management Model. Unpublished Internal Report, McDonnell Douglas Corporation, Long Beach, CA.
- WOLLMER, R. D. 1986b. An Airline Reservation Model for Opening and Closing Fare Classes. Unpublished Internal Report, McDonnell Douglas Corporation, Long Beach, CA.
- WOLLMER, R. D. 1992. An Airline Seat Management Model for a Single Leg Route when Lower Fare Classes Book First. *Opns. Res.* **40**, 26–37.