Stochastic Programming for Multiple-Leg Network Revenue Management

Sanne V. de Boer[*]
Richard Freling[+]
Nanda Piersma[#]

Report EI-9935/A

Abstract
Airline seat inventory control is a very profitable tool in the airline industry. Mathematical programming models provide booking limits or bid-prices for all itineraries and fare classes based on demand forecasts. But the actual revenue generated in the booking process fails to meet expectations. Simple deterministic models based on expected demand generate better revenue than more advanced probabilistic models. Recently suggested models put even more effort into demand forecasting. We will show that the dynamic booking process rather than the demand forecasts needs to be addressed. In particular the nesting strategies applied in booking control will counter-effect the profitability of advanced estimation of booking limits and bid-prices.

Keywords: Revenue management, Mathematical programming, Simulation

[*] ORTEC Consultants, Gouda, The Netherlands
[+] ORTEC Consultants, Gouda, The Netherlands and Econometric Institute, Erasmus University Rotterdam, The Netherlands
[#] Econometric Institute, Erasmus University Rotterdam, The Netherlands

1. Introduction

Airline seat inventory control is about “selling the right seats to the right people at the right time”[1]. If an airline would sell tickets on a first-come first-serve basis, its capacity is likely to fill up early with leisure travelers, eager to fix their holiday trip. Later bookers, typically business travelers willing to pay a higher fare, will then find that there aren’t any seats left, and these sales will be lost. By imposing booking limits on the lower fare classes this can be avoided. Delta Airlines has estimated that selling only one seat per flight at full rather than at discount rate can add over $50 million to its annual revenues[2]. This illustrates the economic significance of *Revenue Management*.

Since the enormous potential profitability of Revenue Management was recognized in the early 1970’s, mathematical models have been developed to determine a booking strategy in a sophisticated way. Typically booking limits are determined at the beginning of the booking process based on demand estimates. They are updated several times during the booking period, although it is practically undesirable to recalculate them every time a booking request is made. The models used to determine the booking limits all share the need of accurately forecasting future demand for each class, for which purpose the demand distributions are estimated from historical booking data. But each model differs in the approach and in the level of detail in representing the booking process. For a comprehensive and up-to-date overview of the area we refer to McGill and van Ryzin (1999) containing a bibliography of over 190 references.

Current methodology can be divided into leg-based and network-based models. Leg-based methods are aimed at optimizing the passenger mix on a single-leg flight. Network-based models consider booking requests for multiple legs at the same time. Developments in the field of leg-based models started with Littlewood’s (1972) study of a single-leg example with two fare classes under the assumption that low-fare class passenger's book first. Other references on this subject include Richter (1982), Belobaba (1987), Wollmer (1992) and Brumelle and McGill (1993). Most of this work involves dynamic programming models.

A major flaw of such leg-based models is that they only locally optimize booking control, whereas an airline should strive to maximize revenue from its network as a whole. These objectives might even conflict. To see this, think of a booking request for a high-fare class flight from Brussels to Amsterdam. Alternatively, this seat can be sold to a passenger of a lower fare class traveling from Brussels to Singapore via Amsterdam. When the direct flight from Amsterdam to Singapore has ample capacity, granting the latter request is more profitable because of its higher ticket price. But if booking control is optimized locally, the former request should be granted, because that is the one for the highest fare class. Airlines typically offer hundreds of such combinations of origin, destination and fare class (ODF). Hence determining an overall booking control strategy for the entire network is by far a trivial matter.

Williamson (1992) has performed a comprehensive research into network revenue management. She studied two network-based mathematical programming models. The first model incorporates probabilistic demand and should be solved by probabilistic mathematical programming (PMP). The second model simplifies the problem by substituting uncertain demand by its expectation so that deterministic mathematical programming (DMP) can be used. After an extensive simulation study she concludes that optimizing booking control over the complete network instead of legs yields significant profit gains.

Booking control can be implemented in various ways. The mathematical programming models we consider are explicitly aimed at determining booking limits. Booking requests are

[1] “The Art of Managing Yield,” *American Airlines Annual Report* (1987)
[2] “Yield Managers Now Control Tactical Marketing,” *Lloyd’s Aviation Economist* (May 1985)

rejected if the respective booking limits would otherwise be exceeded. An alternative form of booking control that can be derived from the dual information of these models is based on *bid-prices*. Bid-prices are linked to each itinerary over the network and no tickets are sold for such an itinerary unless its fare exceeds its bid price. Williamson found no significant difference between these two booking control methodologies.

Surprisingly, she did find that implementing the DMP booking limits consistently produces higher revenues than if the PMP booking limits were used. One should expect that the PMP technique would produce better revenues since it puts more effort into dealing with the unknown demand. Talluri and van Ryzin (1998) show that bid-prices based on the DMP model have better asymptotical properties when the seat capacity on each leg of the network approaches infinity. This might explain Williamson's results for bid-price booking control.

Talluri and van Ryzin (1999) propose an elegant way to incorporate the probabilistic nature of demand into the DMP framework for networks, which they call Randomized Linear Programming (RLP). Instead of using demand expectations, they simulate a sequence of realizations of demand for each itinerary and each fare class in the network. For every realization of demand, the optimal booking control policy is determined by DMP. The resulting sequence of bid-prices is then averaged in order to obtain an approximation of the optimal booking control policy under the assumptions of the probabilistic model. But again, the more advanced probabilistic method fails to uniformly dominate the basic DMP model in their concluding simulation experiments.

Ciancimino et al. (1999) in their study of railway revenue management find that their probabilistic model does yield significantly higher revenues than its deterministic counterpart. Unfortunately, solving this model requires non-linear programming techniques, which for large-scale problems tends to push the optimization time beyond what is acceptable in practice. But, as McGill and van Ryzin (1999) point out, recently developed approximation methods for stochastic programming may overcome this problem and make this methodology a promising and practically feasible option in revenue management.

We study a model (SLP) from the field of stochastic programming that can be seen as a promising alternative for the PMP model by Williamson. In fact, it contains the model by Williamson as a special case. We will show that the observed domination of the DMP model over probabilistic techniques is a fortunate by-product of ignoring the uncertainty related to demand. This phenomenon is based on nesting. The solution of the mathematical programming problem divides the available seats among the various ODF. This suggests that booking requests for an ODF should be accepted as long as they fit into its allocated part of the plane. But this might lead to declining high-fare booking requests in a situation that there are still empty low-fare seats, which cannot be a sensible strategy. Hence passengers from highly profitable classes are given restricted access to seats allocated to less desirable classes, using a nesting strategy. None of the mathematical programming models take nesting strategies into account, but probabilistic models suffer more from this deficiency. In a simulation study, the revenue generated by using bid-prices and booking limits based on the stochastic model SLP is compared to the optimal revenue for a single-leg network and to the DMP revenue for multiple-leg networks.

The outline of this paper is as follows. In Section 2 we discuss the DMP and the PMP models by Williamson and introduce the SLP model. In Section 3 we discuss how their solution can be translated into an effective (nested) booking control policy in multiple-leg networks. In Section 4 we present a dynamic model of the arrival of booking requests. This model is used to simulate the booking process, which enables us to compare the effectiveness of different forms of booking control. In Section 5 we show that booking limits based on DMP, PMP or SLP are not optimal and investigate the cause. In addition, we investigate the factors that influence (the size of) the optimality gap. In Section 6 we compare the revenues generated by DMP and SLP booking limits in a small-scale simulation study for a multiple-leg network. Again we will see that the DMP solution dominates the SLP revenue in most cases and discuss why. Finally, Section 7 summarizes our conclusions.

2. Network models for seat inventory control

The goal of airline seat inventory control is to maximize the expected total revenue from its supply of origin, destination and fare class combinations (ODF), using estimated demand distributions. Each ODF in the network consists of one or more flight legs. The limited capacity on each flight leg has to be used in the most profitable way. This can be achieved by limiting the number of seats available to the less profitable classes. Therefore, let x_{ODF} denote the number of seats reserved for each separate ODF. This definition implies that each seat on each flight leg is available for only one particular ODF. The partitioned approach ensures that the passengers are divided into homogeneous groups whose contribution to the total revenue of the network is clear. Let N be the total number of flight legs in the ODF network. The set of ODF combinations available on flight leg l is denoted by S_l. The probabilistic aggregated demand for each ODF is called D_{ODF}. Although demand is in fact a discrete variable, continuous approximations of the demand distributions are generally used. Furthermore, let f_{ODF} be the fare required for an ODF and let C_l denote the seat capacity on leg l. The x_{ODF} are integer decision variables that should be chosen to maximize the airline's expected profit. At take-off the number of passengers per ODF equals $min\{x_{ODF}, D_{ODF}\}$. Hence the general problem can be formulated as

$$\text{Maximize } E\left(\sum_{ODF} f_{ODF} \min\{x_{ODF}, D_{ODF}\}\right) \qquad \text{(PMP)}$$

subject to

$$\sum_{ODF \in S_l} x_{ODF} \le C_l \quad \text{for all flight legs } l = 1,\dots,N$$

$$x_{ODF} \ge 0 \quad \text{integer for all ODF}$$

2.1 Deterministic approximation methods

A simple approximation method to PMP is to substitute each D_{ODF} by its expectation ED_{ODF}. This results in the deterministic mathematical programming problem DMP.

$$\text{Maximize } \sum_{ODF} f_{ODF}\, x_{ODF} \qquad \text{(DMP)}$$

subject to

$$\sum_{ODF \in S_l} x_{ODF} \le C_l \quad \text{for all flight legs } l = 1,\dots,N$$

$$x_{ODF} \le ED_{ODF} \quad \text{for all ODF}$$

$$x_{ODF} \ge 0 \quad \text{integer for all ODF}$$

Madansky (1960) has shown that the objective value of DMP overestimates the expected outcome of implementing the solution of the deterministic approximation in a booking process. Thus the objective value of DMP provides an upperbound on the actual revenue that will be generated. The LP relaxation DLP is commonly solved instead of the integer-programming problem, since DMP is hard to solve if the number of decision variables and constraints is large. This will often be the case in a real-world situation involving global networks and many different fare classes. The LP approximation is tight if the constraint matrix is totally unimodular. This is only true if the “network” consists of a single multi-leg flight without allowing for connecting traffic (Ciancimino et al. (1999)). We restricted our examples to be of this kind, so we could use the LP approximation to obtain optimal integer solutions.

2.2 Probabilistic approximation methods

One should expect that ignoring the probabilistic nature of demand has a negative impact on total revenue. Suppose that the upper fare classes are highly more profitable. Then it would certainly be rewarding to reserve more high-fare seats than the airline actually expects to sell on average, in order to earn the upward potential of high-fare demand. Therefore, it could be profitable to solve the model PMP exactly. Probabilistic mathematical programming problems explicitly take into account uncertainty of demand.

The most common way to incorporate probabilistic demand is given by a model based on expected marginal revenue (EMR) that was developed by Wollmer (1986):

$$\textit{Maximize } \sum_{ODF} \sum_{i} f_{ODF} P(D_{ODF} \geq i) x_{ODF}(i) \qquad \text{(EMR)}$$

subject to

$$\sum_{ODF \in S_l} \sum_{i} x_{ODF}(i) \leq C_l \quad \textit{for all flight legs } l = 1,...,N$$

$$x_{ODF}(i) \in \{0,1\} \quad \textit{for all ODF and } i = 1,\ldots, \max_l\{C_l | ODF \in S_l\}$$

In this model, the decision variable $x_{ODF}(i)$ is a binary variable representing whether *i* seats or more are allocated to the ODF. Again, the LP relaxation of this model is tight if the network consists of single multiple-leg flights. The objective coefficient represents the expected marginal revenue of allocating an additional i^{th} seat to the ODF. The interpretation of model EMR is intuitively clear, but its large number of decision variables limits its practical applicability.

Another formulation of PMP can be a stochastic programming problem with *simple recourse* (e.g. Birge and Louveaux (1997)). This class of problems is characterized by a separable objective function that includes a deterministic part and a part consisting of functions of a single random variable. Suppose D_{ODF} can only take the values $\{d_{ODF,1} < d_{ODF,2} < \ldots < d_{ODF,Kodf}\}$. Following a result from Wets (1983) the LP relaxation of PMP is equivalent to

$$\textit{Maximize } \sum_{ODF} f_{ODF} x_{ODF} - \sum_{ODF} f_{ODF} \sum_{j=0}^{K_{ODF}} P(D_{ODF} \leq d_{ODF,j}) x_{ODF,j} \qquad \text{(SLP)}$$

subject to

$$\sum_{ODF \in S_l} x_{ODF} \leq C_l \quad \textit{for all flight legs } l = 1,...,N$$

$$x_{ODF} = \sum_{j=0}^{K_{ODF}} x_{ODF,j}$$

$$x_{ODF,0} \leq d_{ODF,1}$$

$$x_{ODF,j} \leq d_{ODF,j+1} - d_{ODF,j} \qquad (j=1,...,K_{ODF} - 1)$$

$$x_{ODF,j} \geq 0 \qquad (j=0,...,K_{ODF})$$

The left-hand term of the objective function is the revenue that would be generated if all planes would depart fully loaded, the term on the right is a correction for the uncertainty in demand. Each x_{ODF} is split up in several smaller allocations $x_{ODF,j}$, which are meant to accommodate the part of the demand D_{ODF} that is between $d_{ODF,j}$ and $d_{ODF,j+1}$. However, if D_{ODF} does not exceed $d_{ODF,j}$, these seats will all remain unsold. The expected "costs" of this event are exactly $P(D_{ODF}\text{£}\, d_{ODF,j})f_{ODF}x_{ODF,j}$. Birge and Louveaux remark (page 194) that practical experience suggests that SLP can be solved in a time of about the same order of magnitude as DLP. Although we have modeled demand as a discrete random variable defined on all positive integers, limiting the demand distribution to the set of integers bounded by its 1% and 99% percentiles is a good approximation. In addition, it reduces the computational load of stochastic optimization significantly. Therefore, we propose to solve SLP to obtain a close approximation of the solution of PMP. For single multiple-leg flights the LP relaxation of PMP is tight and the solution of SLP consists of integer values.

In fact, the LP relaxation of EMR is just a special case of model SLP, as can be seen by letting $d_{ODF,j+1}-d_{ODF,j}=1$ and $d_{ODF,1}=0$. But model SLP is a more efficient representation because it involves fewer decision variables. In addition, model SLP can be simplified to any extends by considering only a limited number of possible scenarios, e.g. low, average and high demand. This reduces the problem size significantly, while still taking into account part of the uncertainty about future demand. If only one scenario is considered (the expected demand) the SLP model reduces to the LP relaxation of DMP. Hence both models can be seen as special cases of the SLP model where DMP contains only one and EMR covers all possible scenarios.

3. Booking control policies

The mathematical models from the previous section determine booking limits for each ODF based on a partitioning of the seats in the planes. In this section we describe how booking limits can be determined by nesting seat allocations and discuss the bid-price approach.

3.1 Nested booking limits

A nesting order of the booking limits for the ODF's should be established based on the contribution of each ODF to the network revenue. Williamson (1992) suggests three ways of measuring this: ordering by fare class, by fare or by shadow prices. Ordering by fare classes, where a full-fare economy class passenger is always rated higher than one from a discount class, does not take into account the magnitude of the total fare. The passenger with a long journey will contribute more than one with only a short flight leg. Ordering by fare will rank flights with the highest fares on top while also the load factors of the flights should be taken into account. Shadow prices measure the incremental network revenue that would be generated if an extra seat would be made available to an ODF. Customer contribution is best measured by shadow prices.

Williamson suggests using the shadow prices of the demand constraints in the deterministic programming formulation. A more general approach is to sum up the shadow prices of the capacity constraints of the legs that the ODF traverses. This is an approximation of the opportunity costs of this itinerary. Subtracting these costs from the ODF's fare results into an estimate of an ODF's contribution to network revenue. For the deterministic model this is equivalent to Williamson's suggestion, but the generalization can be used for the stochastic programming problem as well.

After a ranking order has been established, a booking control strategy has to be determined. Every time a booking request arrives for any ODF in the network, a quick decision should be made whether or not to accept the request. Let *(j,k)* denote the ODF combination of fare class *k* with OD pair *j*. We assume that model PMP or an approximation has been solved and let *x(j,k)* be the number of seats that are allocated to ODF *(j,k)*. We assume that the ODF have been ranked according to their shadow price and we let *(j,k)>(p,q)* say that ODF *(j,k)* is ranked higher than *(p,q)*. Let *n(j,k)* be the number of booking requests for *(j,k)* that have already been *accepted* and let RC_l be the remaining capacity on leg *l*. If leg *l* is part of itinerary *j*, this is denoted by l Î T_j. We propose to use the following algorithm for nested ODF-based booking control.

0. Initialize the variables RC_l by the leg capacities of the flight network. Let $n(j,k)=0$ for all ODF (j,k).
1. A booking request for an ODF (p,q) arrives and should be considered for acceptation.
2. Define $b(j,k)=\max\{x(j,k)-n(j,k),0\}$ for all ODF (j,k).
3. Define $b_l = \sum_{(j,k)\in S_l:(j,k)>(p,q)} b(j,k)$ for all legs $l \in T_p$.
4. Define $c_{\min} = \min\{RC_l - b_l \mid l \in T_p\}$.
5. If $c_{\min} > 0$ accept the booking request, let $RC_l = RC_l - 1$ for all legs $l \in T_p$ and let $n(p,q)=n(p,q)+1$. Decline the request otherwise.
6. If another booking request arrives within the booking period, go to step 1. End the algorithm otherwise.

Step 0 is the initialization phase that occurs only at the beginning of the booking period. In step 1 we consider a booking request for a specific ODF that is made at some time in the booking period. The variables *b(j,k)* that are defined in step 2 represent the number of seats we wish to protect for ODF *(j,k)* against lower ranked ODF's in the remainder of the booking period. The initial value for each ODF depends on the solution of the PMP model and decreases each time that a booking request for this ODF is accepted. Note that by the nesting principle it is possible that the number of accepted booking requests for an ODF exceeds the number of seats that was assigned to it by the solution of the model. If this is the case, we do not wish to hold any additional seats for this ODF. The variables b_l represent the number of seats on each leg that should be protected for ODF that are ranked higher than *(p,q)*. They are calculated in step 3 for each leg that is part of ODF *(p,q)*. In step 4 we calculate how many seats are available for ODF *(p,q)* after the remaining capacity on each leg in this itinerary has been corrected for the desired level of protection. This number, c_{min}, is the minimum of the available number of seats per leg. If there are still seats available, the booking request is accepted and the remaining capacity on each leg and the number of accepted booking requests for this ODF are adjusted accordingly in step 5. If this is not the case, the booking request is declined and no further adjustments have to be made. Finally, if within the booking period another booking request is made, the whole procedure is repeated. Otherwise, the algorithm is ended.

Other booking control strategies could be considered to incorporate the nesting principle. Further research should be done whether other nesting heuristics have a positive impact on total revenue. For our study on comparing PMP and DMP models, however, we restrict ourselves to the above booking control method.

3.2 Bid-prices

The booking limit control process described above requires much administration. This can be avoided by using the dual information of the optimization model. For each OD its contribution to network revenue can be approximated by summing the shadow prices of the capacity constraints of the legs it crosses. If a passenger is willing to pay more for an itinerary than its actual value, it is profitable to grant the request. The value of an OD to the network is called its *bid-price*. Only if an OD's fare is higher than its bid-price, this class is open for bookings. Otherwise, the class is closed. The advantage of this approach is that one only has to reckon with remaining capacity and ODF status (open/closed). Moreover, the nesting principle is automatically satisfied. But when a class is open to bookings, there is no limit on the number of accepted requests. It would therefore be possible that passengers from an ODF that only marginally contributes to network value start filling up capacity. This is an undesirable situation. The problem can be solved by frequently recalculating bid-prices while taking into account bookings at hand. This may lead to closure of some previously open classes and to reopening certain others. Williamson (1992) found through simulation that in this case, bid-prices and nested booking limits lead to comparable results.

A better way to calculate bid-prices (although only for single-hub networks) is given by Günther (1998), but since his approach does not involve mathematical programming, his results are outside the scope of this research.

4. Simulating the booking process

In order to compare the effectiveness of different forms of booking control we have simulated the airline booking process. In modeling the arrival of booking requests, two characteristics should definitely be included. Naturally, their total number is not known beforehand. But in addition, the arrival intensity of each fare class can change over time. The non-homogeneous Poisson process (NHPP) allows modeling both aspects simultaneously. According to the overview given by McGill and van Ryzin (1999), the Poisson process is by far the most popular dynamic model of airline demand. The arrival intensity $l_{ODF}(t)$ consists of two factors. The expected total number of bookings A_{ODF} is assumed to have a Gamma distribution, which has been shown to fit empirical reservation data (e.g. Beckmann and Bobkoski (1958)). The Gamma distribution also is the *conjugate prior* of the Poisson distribution (e.g. DeGroot (1970)), which strongly simplifies analytical study of the model. Including such a factor in the Poisson process solves the problem that the mean and variance of Poisson distributed demand are equal, which need not hold in practice. The arrival pattern is modeled by the standardized beta distribution $b_{ODF}(t)$, which is chosen for its flexible shape. Hence the model we have used for our simulation experiments is a NHPP with arrival intensity

$$l_{ODF}(t) = b_{ODF}(t) * A_{ODF} \qquad A_{ODF} \sim Gamma(p, g) \tag{4.1}$$

It is well known (e.g. Stuart and Ord (1994), page 182) that the total number of booking requests follows a negative binomial distribution. Model (4.1) also introduces a positive correlation between the number of ODF bookings in separate parts of the booking period, hence bookings at hand provide information about bookings to come. If during the booking period the booking control policy is re-optimized, the demand forecast should take this into account. It is easy to derive that the conditional distribution of the number of bookings in the remaining part of the booking period follows a negative binomial distribution as well.

Our model is considered a simplification of the model proposed by Weatherford et al. (1993), who let demand for different fare classes on the same flight be positively correlated. But demand for different fare classes is often motivated in different ways and factors that influence business travel need not affect tourists as well. Moreover, pricing strategies are often aimed at isolating such influences. This suggests that for revenue management purposes, demand for separate fare classes can be assumed independent. According to Swan (1993), most researchers in the field share this view.

In order for our simulation results to be of any practical value, the model presented above should be a realistic representation of the airline booking process. We are not aware of any recent empirical study of airline demand that might confirm this. A popular reference on this subject still is Beckmann and Bobkoski (1958). For this reason, we have tested the model against real airline booking data, provided by a major European carrier (De Boer (1999)). The main result was that both the negative binomial aggregate demand distribution as well as the beta arrival pattern fitted the data very accurately. We should point out, however, that the data were aggregated on flight leg level, hence we were not able to distinguish between local and through traffic. Moreover, the data consisted of accepted booking requests only, hence they are censored by the presence of booking limits and capacity constraints. We have dealt with this problem by only considering demand for high-fare classes, which are unlikely to be closed for bookings. Better booking data should be available for a more complete test.

5. DMP versus PMP: single-leg networks

In this section we consider a well-known basic example (Littlewood (1972)) of a single-leg flight with two fare classes under the assumption that low-fare passengers book before high-fare passengers. We will show that the solution of the probabilistic model follows the effects of the optimal solution where the deterministic solution does not. However, both methods tend to reserve too many seats for the high-fare customers, probably due to the absence of nesting in both models. The single leg case can be used to get a better insight into the effectiveness of DMP and PMP booking limits for multi-leg networks.

5.1 The optimal, DMP and PMP booking limits

Let Φ_i be the cumulative distribution function of D_i and let x_i be the number of seats assigned to class i=1,2 ($x_2 = C - x_1;\ f_1 > f_2$). Littlewood (1972) suggested that class 2 should be closed for further bookings when the revenue of accepting another class-2 booking request is exceeded by the marginal expected revenue of holding the seat for class 1. Richter (1982) proved that this rule indeed is optimal. Hence the optimal protection level x_1^{opt} for class 1 satisfies

$$\frac{f_2}{f_1} = P(D_1 > x_1^{opt}) \qquad (5.1)$$

The (PMP) for this example becomes

Maximize $E\left(f_2 \min\{x_2, D_2\} + f_1 \min\{x_1, D_1\} \right)$

Subject to

$x_1 + x_2 \leq C$

$x_i \geq 0$ *integer for i=1,2.*

The unconstrained solution of (PMP) satisfies

$$\frac{f_2}{f_1} P(D_2 > C - x_1^{pmp}) = P(D_1 > x_1^{pmp}) \qquad (5.2)$$

which is the well-known optimality condition that the marginal value of adding a single extra seat is equal for both classes. Note that this condition only holds when high-fare demand is unlikely to exceed capacity, but otherwise the discount rate wouldn't have been introduced in the first place. Comparing (5.1) and (5.2) learns that the PMP solution always protects more seats for the high-fare class than the optimal protection level. The solution to the deterministic problem is to assign as many seats to class 1 as one expects to sell, hence $x_1^{dmp} = E[D_1]$. The deterministic protection level can be either too high or too low, depending on the circumstances.

Both x_1^{opt} and x_1^{pmp} are increasing in f_1 whereas x_1^{dmp} is a constant. Indeed, for higher values of f_1 it is profitable to protect additional seats for this class. Furthermore, from (5.1) and (5.2) it also follows that both x_1^{opt} and x_1^{pmp} are increasing in $s_1^2=Var(D_1)$, provided that x_1^{opt} (or x_1^{pmp}) $> E(D_1)$. This effect can be explained by the fact that a higher variance implies a larger upward potential in D_1. On the other hand, if the ratio of f_1 and f_2 decreases, the impact of $Var(D_1)$ reverses. In this case, the uncertainty about high–fare demand grows, whereas its incremental profit is relatively low. The negative effect of a high protection level on the load factor is not sufficiently compensated by the positive effect of carrying more high-fare passengers. Neither factor affects the DMP solution, which is an obvious disadvantage of ignoring the probabilistic nature of demand.

5.2 Numerical example

We consider a numerical example to get a better insight into the nature of the optimality gap of the DMP and PMP booking limits. Let $c = 130$, $f_2 = 100$, $D_2 \sim N(100,1000)$ and $D_1 \sim N(50, s_1^2)$. According to McGill and van Ryzin (1999), empirical studies have shown that the Gaussian distribution is a good continuous approximation to the distribution of airline demand. Moreover, (5.1) and (5.2) now can easily be solved in a spreadsheet to obtain the optimal and the PMP booking limits. The results are given in Table 5.1, where $D_{dmp}=x_1^{dmp}-x_1^{opt}$ and $D_{pmp}=x_1^{pmp}-x_1^{opt}$ denote the size of the optimality gap. Note that we did not bother to get integer solutions, since we are not interested in actually implementing the booking limits but only in the differences between them.

Table 5.1
Comparison of the optimal, DMP and PMP booking limits

	f_1=130			f_1=180			f_1=230		
s_1	**10**	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
x_1^{dmp}	50	50	50	50	50	50	50	50	50
x_1^{pmp}	48.61	48.07	47.63	52.01	52.84	53.58	54.18	55.97	57.62
x_1^{opt}	42.64	38.98	35.33	48.60	47.90	47.21	51.65	52.43	53.28
Δ_{dmp}	7.36	11.02	14.67	1.4	2.1	2.79	-1.65	-2.43	-3.28
Δ_{pmp}	5.97	9.10	12.30	3.41	4.93	6.38	2.53	3.54	4.33

It is clear from Table 5.1 that both mathematical programming models tend to overprotect seats for the upper fare class, unless (DMP-case) its fare is exceptionally high in relation to the fare of the other class. Moreover, it depends solely on the difference between the fares what method suffers from this most.

There is only one possible explanation for the optimality gap. The models assumes distinct seat inventories for different fare classes, whereas in practice seat allocations are nested. Hence in order to carry larger numbers of high-fare passengers, it is not necessary to reserve an accordingly large number of seats. Nesting can often accommodate the remaining part of high-fare demand. Lower protection levels lead to higher load factors and, by the principle described above, initially only at the cost of a relatively small fraction of high-fare passengers.

5.3 Simulation results

To examine the impact of non-optimal booking limits on leg revenue, we have used the NHPP demand model from Section 4 to simulate the airline booking process. The PMP booking limits were approximated by the SLP solution. Demand expectations and variances as well as the seat capacity are as in Section 5.2, with $s_1=20$. The simulation statistics include expected total revenue (ER), expected load factor (LF), expected yield per passenger (YP) and the expected load factor of the highest class (LFHC). The LFHC is calculated by dividing the number of accepted high-fare bookings by its non-nested seat allocation. Hence if classes are nested, it can take values greater than 1. The parameters α and β of the beta distributions that determine the arrival patterns are such that high-fare bookings almost certainly arrive after all low-fare bookings[3]. The simulation results are given in Table 5.2. To make the results more insightful, we have presented the optimality gap $(ER-ER_{opt})/ER_{opt}$ rather than the expected dollar amounts.

Table 5.2
Simulation results with optimal, DMP and SLP booking limit (10000 iterations)

	$f_1=130$			$f_1=180$			$f_1=230$		
method	**Opt.**	**DMP**	**SLP**	**Opt.**	**DLP**	**SLP**	**Opt.**	**DLP**	**SLP**
x_1	**34**	50	47	**45**	50	52	**51**	50	56
Gap ER_{opt}	**0%**	-1.7%	-1.0%	**0%**	-0.1%	-0.6%	**0%**	0%	-0.5%
LF	**0.94**	0.91	0.92	**0.93**	0.92	0.91	**0.91**	0.91	0.89
LFHC	**1.09**	0.86	0.90	**0.92**	0.87	0.84	**0.86**	0.86	0.80
YP	**109**	110	110	**127**	129	129	**147**	147	150

With expected total revenue (ER), expected load factor (LF), expected yield per passenger (YP), expected load factor of the highest class (LFHC)

The simulation results support our thesis that booking limits based on mathematical programming models tend to overprotect seats for high-fare classes. The optimal booking limit leads to higher load factors and only a slightly lower average fare per passenger. Note that the opportunity costs of using a non-optimal booking limit are significant and can be as high as 1.7%.

Tables 5.1 and 5.2 show that our thesis holds for different fare levels and different degrees of demand uncertainty. However, our results may have been affected by the stringent assumption about the arrival order. Therefore, we have performed an additional simulation in which we have relaxed the assumption that low-fare passengers always book first[4]. Titze and Griesshaber (1983) have found that even if low-fare requests only *tend* to arrive before high-fare requests, this does not have a significant impact on the optimal seat allocation, provided that D_1 and D_2 are assumed to be independent. Consequently, condition (5.1) can still be used

[3] To be precise: $\alpha_1=2;\beta_1=13;\alpha_2=13;\beta_2=2$.
[4] Now: $\alpha_1=2;\beta_1=5;\alpha_2=5;\beta_2=6$.

to approximate the optimal booking limit. Additional simulation results indicate that when there is no strict arrival order, the effect of the overprotection even worsens. This is because if part of low-fare demand comes after high-fare demand, the chances of being able to accept additional high-fare bookings by nesting have improved, which favors setting a low protection level. This illustrates another shortcoming of the models under consideration. They are static, whereas booking control is typically a dynamic process.

6. DLP versus SLP: multiple-leg networks

In this section we simulate the booking process for a single multiple-leg flight to compare the performance of the DLP and SLP models in a network environment. Notice that in this case the totally unimodularity assumption holds and that the solutions are integers. We consider a flight over three legs, each with a capacity of 200 seats. The airline offers all six possible OD pairs. Passengers can choose from three different fare classes for each OD. Booking requests are accepted starting from 150 days before departure. The initial demand and fare settings are given in the appendix. Long haul flights are relatively cheap compared to single-leg flights. The highest fare class is relatively expensive. The parameters of the customer arrival process are chosen to let low-fare passengers in general book before high-fare passengers.

6.1 Bid-pricing versus booking limits

We are interested in the differences between the bid-price and the booking limit approach. In our first example, the booking control strategy is optimized only once, in the beginning of the booking period. The optimization results for both models are in Table 6.1.

Table 6.1
Solutions to DLP and SLP for the initial example

	Class 3		Class 2		Class 1	
Leg	DLP	SLP	DLP	SLP	DLP	SLP
1	41	42	40	40	30	40
2	0	0	25	18	20	22
3	0	0	24	21	20	17
4	30	23	20	19	20	27
5	1	15	20	16	20	22
6	45	38	40	36	30	35

Objective value DLP: 84915, SLP: 71767.35

The SLP solution assigns more seats to the upper fare classes than the DLP solution (as was observed in the single-leg network). In addition, the simulation results from Table 6.2 show that the DLP method outperforms the SLP approach.

Table 6.2
Simulation results with the initial settings (5000 iterations)

Booking Control	**booking limit control**		**bid price control**	
Policy	**DLP**	**SLP**	**DLP**	**SLP**
ER	75983	74726	73501	73416
LF	0.897	0.886	0.96	0.96
YP	195	197	177	177
LFHC[5]	0.96	0.875	0.60	0.54

With expected total revenue (ER), expected load factor (LF), expected yield per passenger (YP) and the expected load factor of the highest valued class (LFHC)

[5] Keep in mind that classes are ranked by shadow prices, hence the highest class might differ between models.

We see that applying DLP instead of SLP booking limits leads to significantly higher revenue (1.7 %). The positive effect of the higher LF apparently more than makes up for the lower YP. It seems that the SLP solution keeps too many seats away from low-fare classes, which is especially clear from the much lower LFHC resulting from this method. This is consistent with the tendency to overprotect high-fare seats on single-legs that we found in Section 5. Another interesting observation from Table 6.2 is that the DLP solution overestimates expected revenue whereas SLP underestimates it.

Table 6.2 suggests that bid-price-booking control is less profitable than the booking limit approach. As we pointed out in Section 3.2, bid-price-booking control does not have a limit on the number of requests that are accepted once a class is open for bookings. When low-fare classes requests arrive before the other classes, the plane will fill up quickly with relatively low yielding passengers, resulting in the lower YP and LFHC for the bid-price approach. The higher LF is exclusively caused by accepting more low-yield passengers, and apparently does not form sufficient compensation. For the bid-price approach to work, the bid-prices have to be adjusted regularly during the booking period.

With the next example we aim to illustrate the positive effects of frequently updating the booking control strategy. Booking limits and bid-prices are recalculated at 1/3 and 2/3 of the booking period using bookings at hand. The simulation results are in Table 6.3.

Table 6.3
Simulation results implementing two updates (1000 iterations)

Booking Control Policy	**booking limit control**		**bid price control**	
	DLP	**SLP**	**DLP**	**SLP**
ER	76248	75863	76431	75962
%ER increase	0.35%	1.52%	3.99%	3.47%
LF	0.91	0.90	0.95	0.96
YP	195	195	187	185

Indeed, expected revenue has risen and bid-price-booking control now results into similar (or slightly better) expected revenue. The higher LF of the bid-price approach compensates its lower YP, which are both caused by the absence of booking limits. DLP is again more profitable than SLP, although the difference has decreased. The uncertainty about future bookings decreases throughout the booking period and the solutions of the deterministic and stochastic model become more alike. Additional simulations with more updates confirmed these conclusions, hence our results about the effectiveness of the DLP and the SLP model are essentially independent of which method for booking control is used. For this reason, we have not considered the bid-price approach in our further analysis.

6.2 Sensitivity analysis: demand distribution and fares

We will now investigate the factors that may have affected our comparison of the DLP and the SLP model. We start by examining the influence of demand variance on booking limits and on total expected revenue. Recall that we have shown in Section 5 that more uncertainty concerning high-fare demand leads to a larger degree of overprotection and less revenue. We therefore limit our examination to the consequences of increased low-fare demand variance (exact values in the appendix). Since Belobaba (1987) has reported 0.33 as a common coefficient of variance for demand in the airline industry, our initial variances were rather low. Because the DLP model only involves the demand expectations, its solution does not change. The new solution to the SLP model is given in Table 6.4.

Table 6.4
SLP solution for increased low-fare demand variance

Objective value: 70679.23	class 3	class 2	class 1
1	41	41	41
2	0	15	23
3	0	21	18
4	22	19	28
5	17	15	22
6	35	36	36

Comparison of Tables 6.1 and 6.4 learns that expected revenue has decreased and that the seats allocations to both lower classes are slightly smaller. The simulation results are given in Table 6.5.

Table 6.5
Simulation results with increased low-fare demand variance (5000 iterations)

Booking limit control	**DLP**	**SLP**
ER	75362	74662
%ER decrease against average low-fare demand	0.08%	0.01%
LF	0.883	0.878
YP	197	201

Once again, DLP outperforms SLP, but the increased demand uncertainty has affected the deterministic model most. The number of accepted booking requests for an ODF is truncated from above by the respective booking limit, hence using demand expectations overestimates the occupation of the plane. The effect of the truncation is worsened by the increased uncertainty about demand, and the deterministic model is not able to take this into account and adjust the seat allocations accordingly. As a result, DLP overestimates accepted high-fare demand by failing to recognize the truncation, while SLP underestimates it by ignoring the possibility of nesting. The optimal expected revenue will be somewhere in the middle, which provides an airline useful bounds on the incremental profit of implementing network-based booking control.

In Section 5.2 we found that a smaller spread between fares for different classes favors SLP compared to DLP. We now examine whether this conclusion holds in a network environment. The adjusted fares are in given in the appendix.

Table 6.6
Solutions to DLP and SLP for smaller spread between fares

	Class 3		Class 2		Class 1	
Leg	DLP	SLP	DLP	SLP	DLP	SLP
1	41	45	40	41	30	36
2	0	4	25	20	20	14
3	0	22	24	18	20	25
4	30	20	20	22	20	21
5	1	21	20	17	20	17
6	45	38	40	37	30	30

Objective value DLP: 70615, SLP: 60549.43

Comparison of Table 6.6 with Table 6.1 learns that the SLP seat allocations to the highest fare class have significantly decreased, as was expected. In some cases the SLP solution even allocates fewer seats to a first class OD than the DLP solution. A possible explanation is that the number of seats allocated to the highest fare class nearly equals expected demand, which aggravates the demand truncation. The number of seats allocated to the lowest fare class will often be smaller than its expectation, which weakens the impact of the truncation. Moreover, this class is not unattractive for the airline anymore. It is therefore a good strategy to shift reserving seats from first to third class, but the DLP model fails to recognize this opportunity. The simulation results are in Table 6.7.

Table 6.7
Simulation results for a smaller spread among fares (5000 iterations)

Booking limit control	**DLP**	**SLP**
ER	63356	63181
LF	0.90	0.92
YP	164	163

The revenue generated by DLP is now only slightly higher (0.3 %). It is interesting to see that the roles have reversed: the YP resulting from the DLP booking limits is higher, and the LF lower. Our earlier conclusion that DLP causes the largest overprotection of high-fare seats if the spread among fares is small appears to hold in a network environment as well.

We can conclude that the SLP model can only produce better revenue than DLP if the differences among the fares become very small and the variance for low-fare demand becomes large. Indeed we were able to construct an example where the SLP model outperformed the DLP. However, the example is not representative for airline demand.

Notice that the SLP model is basically the same as EMR when the full demand distributions are incorporated. One might wonder whether distinguishing fewer demand scenarios will improve the performance of SLP compared to DLP. However, fewer scenarios will only diminish the gap between DLP and SLP revenue by removing some of the uncertainty related to demand. Beside the fact that it is a rather odd strategy to simplify a model in order to get a better solution, the DLP model will still outperform SLP regardless of the number of scenarios.

7. Conclusions

It is well known that taking into account the network aspect of seat inventory control significantly increases an airline's revenue. We have examined two mathematical programming models that are aimed at doing so: one that incorporates probabilistic demand and a deterministic approximation of this model. The solution of both models is a partitioning of network capacity into seat assignments to all possible combinations of flight-origin, destination and fare class. When translated to an actual booking control policy, however, these seat allocations are nested. This roughly implies that profitable passengers also have access to seats assigned to travelers that contribute less to network revenue. As a result, the booking control policy is non-optimal and in general leads to an overprotection of seats for the most attractive classes. This results into lower load factors but only in a marginally higher average fare per passenger. We have shown that this can lead to a loss of revenue of up to 2% even in a simple network environment.

The probabilistic model suffers heavier from ignoring the nesting property than its deterministic counterpart. In order to earn the upward potential of high-fare demand, it assigns even more seats to these classes, which in fact only aggravates the degree of overprotection. Only if these classes are highly profitable, this approach will pay off. The deterministic model is unable to recognize this potential at all, which is a major weakness. For this reason, the probabilistic model will yield higher profits in a non-nested environment, but in a nested environment this weakness of the deterministic model turns to be profitable. This explains the results by Williamson (1992) and shows that stochastic programming techniques will not be successful either. In railway practice (Ciancimino et al. (1999)) the assumption about non-nested seat allocations actually holds. This explains why this application area is an exception to this rule, where the more sophisticated probabilistic model yields better results. Here model SLP offers a practical alternative to the non-linear programming technique that they propose to use. The problem size can easily be limited by considering only a small number of demand scenarios and the optimal integer solution can be obtained by ordinary linear programming.

Factors that turn out to have a negative impact on the effectiveness of both probabilistic and deterministic models are uncertainty about the number of future bookings, small spreads between fares for different classes and booking requests for different fare classes arriving simultaneously. This indicates that an airline can diminish the degree of non-optimality by offering well-differentiated fare classes to market segments of significant size. However, optimal booking control is only possible by including the possibility of nesting into the models and only then will the more advanced probabilistic model pay off. Ideally, this method should also be dynamic of nature to deal with unknown future demand and be capable of incorporating corrections in the booking limits or bid-prices for nesting. Hence alternative modeling approaches are needed.

References

Belobaba, P.P. (1987), “Air Travel Demand and Airline Seat Inventory Management,” PhD dissertation, MIT, Cambridge, Mass.

Beckman, M.J. and Bobkoski, F. (1958), “Airline Demand: An Analysis of Some Frequency Distributions,” *Naval Research Logistics Quarterly*, **5**, 43-51

Birge, J.R. and Louveaux, F. (1997), “Introduction to Stochastic Programming,” Springer Series in Operations Research

Brumelle, S.L. and McGill, J.I. (1993), “Airline Seat Allocation with Multiple Nested Fare Classes,” *Operations Research,* **41**, 127-137

Ciancimino, A., Inzerillo G., Lucidi S. and Palagi, L. (1999), “A Mathematical Programming Approach for the Solution of the Railway Yield Management Problem,” *Transportation Science*, **33**, 168-181

De Boer, S.V. (1999), “Mathematical Programming in Airline Seat inventory Control,” Master’s Thesis, Vrije Universiteit, Amsterdam, The Netherlands

DeGroot, M.H. (1970), “Optimal Statistical Decisions”, McGraw-Hill

Günther, D. (1998), Airline Yield Management: Optimal Bid Prices, Markov Decision Processes and Routing Considerations,” PhD thesis, Georgia Institute of Technology

Littlewood, K. (1972), “Forecasting and Control of Passengers,” in *Proceedings 12th AGIFORS Symposium*, American Airlines, New York

Madansky, A. (1960), “Inequalities for Stochastic Linear Programming Problems,” *Management Science*, **6**, 197-204

McGill, J.I. and van Ryzin, G.J. (1999), "Revenue Management: Research Overview and Prospects," *Transportation Science*, **33**, 233-256

Richter, H. (1982), "The Differential Revenue Method to Determine Optimal Seat Allotments by Fare Type," *AGIFORS Symposium Proceedings*, **22**, 339-362

Stuart, A. and Ord, K. (1994), "Kendall's advanced theory of statistics vol. 1: distribution theory," Edward Arnold Publishers, London

Swan, W.M. (1993), "Modeling Variance for Yield Management," prepared for AFIFORS Yield Management Group, 2-5 May 1993, Boeing Commercial Aircraft

Talluri, K.T. and van Ryzin, G.J. (1998), "A Randomized Linear Programming Method for Computing Network Bid Prices," *Transportation Science,* **33**, 207-216

Talluri, K.T. and van Ryzin, G.J. (1999), "An Analysis of Bid-Price Controls for Network Revenue Management," *Management Science*, **44**, 1577-1593

Titze, B. and Griesshaber, R. (1983), "Realistic Passenger Booking Behavior and the Simple Low-Fare/High-Fare Seat Allotment Model," *AGIFORS Symposium Proceedings*, **23**, 197-223

Weatherford, L.R., Bodily, S.E. and Pfeifer, P.E. (1993), "Modeling the Customer Arrival Process and Comparing Decision Rules in Perishable Asset Revenue Management Situations," *Transportation Science*, **27**, 239-251

Wets, R. J-B. (1983), "Solving Stochastic Programs with Simple Recourse," *Stochastics*, **10**, 219-242

Williamson, E.L. (1992), "Airline Network Seat Control," PhD thesis, MIT, Cambridge, Mass.

Wollmer, R.D. (1986), "A Hub-Spoke Seat Management Model," Unpublished Internal Report, Mc Donnell Douglas Corporation, Long Beach, CA

Wollmer, R.D. (1992), "An Airline Seat Management Model for a Single Leg when Lower Fare Classes Book First," *Operations Research,* **40**, 26-37

Appendix

In Tables A. 1 – A. 4 we give the parameter settings for the examples considered in Section 6. Recall from Section 4 that we simulate the booking process by a non-homogeneous Poisson process with arrival intensity $l_{ODF}(t) = A_{ODF}\ b_{ODF}(t)$, in which $A_{ODF} \sim Gamma(p, g)$ and $b_{ODF}(t)$ is the standardized beta distribution with parameters α and β. We let *E* and *Sdev* denote the expectation and standard deviation of the aggregate demand for an ODF during the booking period, which follows a negative binomial distribution.

Table A.1: Initial fare settings

OD Number	Origin - Destination	Fare class 3	Fare class 2	Fare class 1
1	A – B	75	125	250
2	A – C	130	170	400
3	A – D	200	320	460
4	B – C	100	150	330
5	B – D	160	200	420
6	C – D	80	110	235

Table A.2: Initial demand settings

	Class 3 a=**5,** b=**6**				**Class 2** a=**2,** b=**5**				**Class 1** a=**2,** b=**13**			
OD	**P**	g	**E**	**Sdev**	**p**	g	**E**	**Sdev**	**p**	g	**E**	**Sdev**
1	80	1.6	50	9.01	80	2	40	7.75	3	0.1	30	18.17
2	80	2	40	7.75	50	2	25	6.12	2	0.1	20	14.83
3	60	2	30	6.71	72	3	24	5.66	2	0.1	20	14.83
4	60	2	30	6.71	40	2	20	5.48	2	0.1	20	14.83
5	60	2	30	6.71	60	3	20	5.16	6	0.3	20	9.31
6	80	1.6	50	9.01	80	2	40	7.75	6	0.2	30	13.42

Table A.3: Increased variance of demand for classes 2 and 3

	class 3				***class 2***				***Class 1***			
OD	**P**	g	**E**	**Sdev**	**p**	g	**E**	**Sdev**	**p**	g	**E**	**Sdev**
1	20	0.4	50	13.23	20	0.5	40	10.95	3	0.1	30	18.17
2	20	0.5	40	10.95	5	0.2	25	12.25	2	0.1	20	14.83
3	15	0.5	30	9.49	18	0.75	24	7.48	2	0.1	20	14.83
4	15	0.5	30	9.49	10	0.5	20	7.75	2	0.1	20	14.83
5	15	0.5	30	9.49	15	0.75	20	6.83	6	0.3	20	9,31
6	20	0.4	50	13.23	20	0.5	40	10.95	6	0.2	30	13.42

Table A 4: Smaller spread between fares

OD Number	Fare class 3	Fare class 2	Fare class 1
1	75	125	175
2	130	170	220
3	200	320	440
4	100	150	210
5	160	200	250
6	80	110	160