Optimizing Airline Overbooking Using a Hybrid Gradient Approach and Statistical Modeling

Sheela Siddappa[*][†], Jay M. Rosenberger[§][¶], Victoria C. P. Chen[§]

February 16, 2007

COSMOS Technical Report 06-07

Sheela Siddappa *is a Researcher at General Motors in Bangalore, India. She holds a PhD in industrial engineering from The University of Texas at Arlington. Her research interests include revenue management, statistical regression, and optimization under uncertainty.*

Jay Rosenberger *is an assistant professor of industrial engineering at The University of Texas at Arlington. He holds a PhD in industrial engineering from the Georgia Institute of Technology. His research interests include mathematical programming and simulation in transportation, defense, and health care. Dr. Rosenberger's graduate research on airlines won the First Place 2003 Pritsker Doctoral Dissertation award. Prior to joining the faculty at UTA, Dr. Rosenberger worked in the Operations Research and Decision Support Department at American Airlines.*

Victoria Chen *is an associate professor of industrial engineering at The University of Texas at Arlington. She holds a PhD in Operations Research and Industrial Engineering from Cornell University. Dr. Chen's primary research interests utilize statistical methodologies to create new*

[*] General Motors Technical Centre -India 3rd Floor, Creator Building, ITPL, Whitefield Road Bangalore, Karnataka, 560066 India
[†] This research was done as part of Siddappa's Ph.D. theses (Siddappa 2006).
[§] Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington Campus Box 19017 Arlington, TX 76019 USA
[¶] Corresponding author: Tel: +1 817 272 5787; Fax: +1 817 272 3406; E-mail: jrosenbe@uta.edu
[§] Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington Campus Box 19017 Arlington, TX 76019 USA

methods for operations research problems appearing in engineering and science. She has expertise in the design of experiments and statistical modeling, particularly for computer experiments and stochastic optimization. She has studied applications in inventory forecasting, airline optimization, water reservoir networks, wastewater treatment, and air quality. Through her statistics-based approach, she has developed computationally-tractable methods for continuous state stochastic dynamic programming, yield management, and environmental decision-making.

Optimizing Airline Overbooking Using a Hybrid Gradient Approach and Statistical Modeling

Key Words: revenue management, overbooking, multivariate adaptive regression splines, computer experiments, Newton's method, steepest ascent

Abstract

We develop an overbooking approach for airline revenue management. We estimate a revenue function by employing a statistical modeling approach, specifically a multivariate adaptive regression splines approximation of a stochastic network model. We develop an overbooking cost function using a binomial distribution to model the number of customers that show up for the flight. We implement a hybrid gradient algorithm that combines Newton's and a steepest ascent method to optimize profit. Finally, we compare our method to one that overbooks based on the probability that the number of customers that show up for a flight exceeds the flight capacity.

Revenue management is defined as, "Selling the right seat at the right time to the right passenger for the right price" (Vinod 1995). After deregulation in 1979, revenue management research has evolved to efficiently utilize resources and generate higher revenue. One of the revenue-capturing concepts identified was *overbooking*, where airlines sell more seats than the capacity of the flight. Overbooking is essential because not all customers who buy tickets show up on the day of departure. Some of them are *cancellations* (customers who cancel a purchased ticket), while others are *no-shows* (customers who do not show up for their ticketed flight). The *overbooking level* (or authorization level) specifies the number of seats that may be overbooked on a flight. In this paper, we present an approach to determine optimal overbooking levels, so as to generate higher profit (revenue − cost). The revenue is estimated using a design and analysis of computer experiments (DACE, Sacks et al. 1989) approach, and only overbooking costs are estimated. (We ignored other airline costs, such as fuel burn, labor, etc. that are generally considered independent of the booking process.) To optimize profit, a hybrid algorithm that combines Newton's method and a steepest ascent method is developed (Bertsekas 1999).

Perishable goods need to be utilized well before the end of their life cycle. The overbooking concept can be applied to various transportation sectors, such as auto rentals, ferries, rail, tour operators, cargo, and cruises. Other areas, like hotels and resorts, extended stay hotel, health care, and companies that produce perishable goods can also use overbooking. In the airline industry, once a flight takes off the seats are said to be "perished." Hence, we should efficiently utilize those seats. Given the schedule, flight capacities, and demand distributions, how many seats should be overbooked on each flight, so that the seats are efficiently utilized and higher profits are achieved? The remainder of this section provides background on airline overbooking, and describes the contribution of this paper. Section 1 explains the methodology adapted in this paper to determine the optimal overbooking levels, and Section 2 presents computational results. Finally, concluding remarks are given in Section 3.

Overbooking Concept

Let n_c be the number of cancellations and no-shows, and let y be the overbooking level.

Figure 1about here.

Figure 1 shows the three possible outcomes of overbooking.

- The number overbooked equals the number of cancellations and no-shows ($y = n_c$), which is ideal.
- The number overbooked is less than the number of cancellations and no-shows ($y < n_c$). In this case, although overbooking is considered, it is still insufficient to fill all of the empty seats. Revenue is lost because a few seats fly empty; this is called spoilage.
- The number overbooked is greater than the number of cancellations and no-shows ($y > n_c$). In this case, some of the passengers are unable to board the flight, called bumped or oversold passengers. A penalty has to be paid to those bumped passengers by providing them with a voucher for another round trip or an accommodation in a hotel to wait for the next flight.

Hence, it is important to identify the optimal number of seats to be overbooked.

Overbooking Literature

Chatwin (1993) stated, "experience shows that nearly 15% of the seats fly empty if overbooking is not considered." Kosten (1960) developed a continuous time stochastic model, but the model was proved impractical by McGill and van Ryzin (1999) because it required finding a solution to many differential equations. Bodily and Pfeifer (1973) considered the probability of a customer cancellation, and determined that it depends on (1) when the reservation was made, and (2) unknown events that might occur before departure. The drawback of this approach was that it did not consider the dynamic nature inherent in the reservation process.

Rothstein (1971) was the first to formulate airline overbooking as a dynamic programming problem. This approach was computationally intractable due to the curse of dimensionality. Chatwin (1993) mentioned two ways to overcome the above problem of computation: (1) approximate the states by aggregating them or (2) develop a theory of the structure of an optimal solution, so as to facilitate more efficient computation. The first approach was followed by Alstrup et al. (1989). They developed a dynamic programming approach to solve the overbooking problem for two fare classes. They assumed that customers requested and cancelled reservations in groups of five, thereby reducing the size of the state space by a factor of 25. Their dynamic yield management model is being implemented worldwide. Chatwin (1993) implemented the second idea in his dissertation. He considered overbooking models with (1) discrete time and discrete state spaces and (2) a continuous time birth and death process. Subramanian et al. (1999) formulated the overbooking revenue management problem as a finite-horizon, discrete-time Markov decision process (MDP). It was an extension to the model developed by Lee and Hersh (1993). Chatwin (1998) solved the multiperiod overbooking problem that relates to a single flight leg and service class. The conditions on the fares, refunds, and distributions of passenger demand for reservations and cancellations in each period, and on the bumping penalty function are given to ensure that a booking-limit policy is optimal. In other words, Chatwin stated "in each period the airline accepts reservation requests up to a booking limit if the number of initial reservations is less than that booking limit; it declines reservation requests otherwise." The model is applied to the discount allocation problem in which lower fare classes book prior to higher fare classes. Karaesmen and van Ryzin (2004) modeled the overbooking problem as a two-period optimization problem. In the first period, given the probabilistic knowledge of cancellations, reservations are accepted. In the second period, cancellations are realized and surviving customers are assigned to the various inventory classes to minimize penalties.

The drawback of the approaches in the literature is that they generally consider smaller simpler airline networks than the one in our computational experiments in Section 2. Consequently, implementing them on a network such as the on in Section 2 would require determining methods to simplify it, which is another topic of research.

Contribution

Our overbooking optimization seeks to determine the overbooking levels for the different flights that maximize the total profit of the flight network. Revenue is estimated using DACE, which generates data via a *computer experiment* then constructs a statistical model that will be employed to improve the performance of the system. Recent reviews of DACE methods are given by Chen, Tsui, Barton and Allen (2003) and Chen et al. (2006). The typical computer experiment is a simulation (Kleijnen 2005; Sacks et al. 1989); however in our approach, the computer experiment is an optimization model that solves for the seat allocation that maximizes expected revenue (Chen, Günther and Johnson 2003). For the "design" aspect, we use a strength two orthogonal array experimental design (Bose and Bush 1952; Chen 2001) to organize a set of overbooking levels that are input into the revenue optimization. For the "analysis" aspect, a multivariate adaptive regression splines (MARS, Friedman 1991; Tsai and Chen 2005) model is fit over the revenue data, as a function of the overbooking level.

An oversold seat is a customer that cannot be accommodated on a flight because more customers show up than there is capacity ($y > n_c$). Overbooking costs due to oversold seats are penalties paid by the airline to maintain good will with customers. In this paper, we conservatively estimate the penalty for each oversold seat as three times the bid price for that flight. A binomial distribution is used to model the number of customers that show up for the flight, and the cost function is the expected total penalty, as a function of the overbooking level.

To conduct the profit optimization, an appropriate algorithm was needed for an objective function that involves a MARS function. Our approach develops a hybrid algorithm that combines Newton's method and a steepest ascent method. Finally, we present a comparison to a simple approach that determines the overbooking levels based on the probability that the number of customers that show up for a flight exceeds the flight capacity.

1 Overbooking Optimization Approach

In this section, we describe the estimation of expected revenue and cost functions, followed by our hybrid Newton's and steepest ascent optimization algorithm for determining the set of overbooking levels that maximize expected profit. For overbooking level y, let $R(y)$ be the revenue function, and let $O(y)$ be the overbooking cost function, where cost is assumed to increase, as y increases. Since the number of seats booked is finite, revenue always increases as y increases. Thus, the profit function is

$$Z(y) = R(y) - O(y).$$

Figure 2 presents a conceptual diagram of the expected revenue, expected cost, and expected profit functions, with respect to overbooking level y.

Figure 2 about here.

Hence, the objective of our overbooking model is: $\max_{y \geq 0} Z(y)$.

1.1 DACE Concept

A design and analysis of computer experiments (DACE) approach is useful when a computer experiment is the only means for representing a complex system (see recent reviews Chen, Tsui, Barton and Allen 2003; Chen et al. 2006). The purpose of DACE is to provide a computationally efficient method for conducting computer experiments with the objective of optimizing the performance of a complex system. In a typical setup, a simulation model is used to represent the complex system, but is too computationally intensive to directly imbed within an optimization. Instead, a metamodel, i.e., a statistical model of the simulation model, is constructed to predict the performance of the system, and the metamodel, which has a closed form and is quick to compute, is used within the optimization. Specifically, our DACE approach conducts the following:

- An optimization model (computer experiment) of system performance is constructed based on knowledge of how the system operates.
- An experimental design is used to select the set of sample points as input to the optimization model, which then provides the corresponding system performance outputs.
- A statistical metamodel is to fit these data.

The optimization model used in this paper is described in the next subsection. Several options are available for experimental designs and statistical metamodeling. In general, when the underlying form of the true relationship between system performance and system inputs is unknown, a "space-filling" design is chosen to select sample points that cover the input space. These include orthogonal arrays (Bose and Bush 1952; Hedayat et al. 1999, Chen 2001), Latin hypercubes

(McKay et al. 1979, Tang 1993, Ye 1998), and number-theoretic methods (Niederreiter 1992, Fang and Wang 1994). Orthogonal arrays, for example, are based on lattices in the input space.

Choices for statistical metamodeling include spatial correlation models (Sacks et al. 1989, Koehler and Owen 1996), a.k.a., "kriging;" multivariate adaptive regression splines (MARS, Friedman 1991, Tsai and Chen 2005); and neural networks (Haykin 1999). Kriging assumes some form of spatial correlation between points in the multi-dimensional input space, and uses this correlation to predict values between observed points. However, kriging can be very computational intensive and is often impractical for higher than ten dimensions. MARS and neural networks have been successfully applied for high-dimensional DACE-based stochastic dynamic programming (20 dimensions: Tsai et al. 2004, 30 dimensions: Cervellera et al. 2006, 25 dimensions: Yang et al. 2007). A neural network requires significant effort to identify the appropriate model architecture while the MARS architecture is adaptively determined by the data. Hence, for this paper, we employ MARS.

MARS is a linear statistical model with a forward stepwise algorithm to select model terms followed by a backward procedure to prune the model. The approximation bends to model curvature at "knot" locations, and one of the objectives of the forward stepwise algorithm is to select appropriate knots. If appropriate, the model terms can be "smoothed" to achieve a desired degree of continuity, e.g., continuous first derivatives. A critical parameter of the forward stepwise algorithm is the maximum number of basis functions, which determines when to stop adding model terms. Friedman's original MARS algorithm requires the user to supply this parameter, but the version by Tsai and Chen implements an automatic stopping rule, based on model fit. For stochastic dynamic programming applications, the automatic stopping rule enables higher flexibility in the use of MARS.

1.2 DACE-Based Revenue Function

A purchased ticket specifies an *itinerary* of connecting *flight legs*. Due to an airline's complex fare structure, equivalent seats on the same flight are sold at different prices to different types of customers. *Itinerary-fare classes* define different customers by their chosen itinerary and fare class. The airline revenue management problem allocates seat capacity for all the flight legs to the itinerary-fare classes to maximize revenue.

Chen, Günther and Johnson (2003) used a DACE-based statistical modeling approach to model the value functions of a revenue management Markov decision process. Specifically, they used a deterministic linear programming formulation as an upper bound on the value function and a stochastic network model as a lower bound. A statistical modeling module constructs MARS approximations for these upper and lower bounds, and then these are used on-line within the availability processor module that conducts the ticket bookings. Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) revised the statistical modeling approach of Chen, Günther and Johnson (2003) to (i) derive more realistic remaining capacity state spaces and enable more accurate MARS approximations, and (ii) develop a more accurate algorithm for a modified version of the stochastic network model. Each value function approximation requires solving for the upper and lower bound values at each point of an experimental design and then fitting two MARS approximations, one for the upper bound and one for the lower bound. For a 20-city, 31-leg network, computational effort per value function was less than 10 minutes on a Dual 2.8-GHz Intel Xeon Workstation. A similar, but far less computationally efficient approach, was presented by Bertsimas and de Boer (2005), where they reported a CPU effort of 50–60 hours to estimate their value function for a 15-city hub-and-spoke network.

In this paper, we use the stochastic network model (STOCH) of Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) as a computer experiment that maximizes the expected revenue over the entire booking period for a given set of overbooking levels. Let y be the vector of overbooking levels for all the flight legs, and let q be the corresponding vector of flight capacities. Let $A = [a_{jf}]$ denote the leg-itinerary incidence matrix with $a_{jf} = 1$ if leg j is utilized in itinerary-fare class f, and 0 otherwise; For itinerary-fare class f, let d_f be a random variable for its demand, r_f be its fare, and u_f be its seat allocation decision variable. Then for m itinerary-fare classes, the (STOCH) formulation in this paper is:

$$\begin{aligned} \text{(STOCH)} \ \max \sum_{f=1}^{m} r_f E[\min(d_f,\ u_f)] \\ \text{s.t. } Au \le q + y \\ u_f \ge 0. \end{aligned}$$

The dual solution to (STOCH) provides an initial bid price for each flight leg (D'Sylvia 1982). In a bid price revenue management approach, if a customer's booking request is greater than or equal to the sum of the bid prices along the desired itinerary, then the request is accepted; otherwise it is rejected. These bid prices are used in Section 1.2 to estimate the penalty for overbooking. Bid prices can be updated over time by updating the remaining capacities and re-solving the stochastic network model.

Our DACE-based approach uses a strength two orthogonal array experimental design, where each point in the design specifies a set of overbooking levels y. The (STOCH) model is solved for each design point, and a MARS model is fit over the expected revenue data to provide our metamodel revenue function $R(y)$. Since our optimization algorithm uses a combination of Newton's and steepest ascent, the profit function $Z(y)$ in equation (1) must be continuous and have continuous first and second derivatives. Consequently, we utilize a smoothed version of MARS for *R(y)* (see Siddappa, Chen and Rosenberger 2006).

1.3 Derivation of Overbooking Cost Function

A simplified cost estimate is used and assumes: (1) the numbers of customers that show up are independent and identically distributed, (2) the probability that any customer shows up is a constant, and (3) the penalty per oversold seat is three times the bid price of the itinerary the customer is willing to travel. The penalty in assumption (3) is obtained by solving (STOCH) for the bid prices. Given the first two assumptions, we derive here the cost function for a single flight leg, and we denote its penalty by c. Let q be the capacity of the flight; let x be the sum of the flight capacity and the overbooking level, $x = (q + y)$; let Y be a random variable for actual number of seats overbooked, realized on the day of departure; and let W be a random variable for the total number of customers that show up on the day of departure. The number of customers that show up is assumed to be binomially distributed, hence from the definition of the binomial distribution, we have

$$P[W = w] = \binom{x}{w} \alpha^{w} (1-\alpha)^{x-w}.$$

To simplify notation, let $[x]^{+} = \max(x,0)$. From the definition of Y, we have

$$\begin{aligned} Y &= [W-q]^{+} \\ E[Y] &= E[W-q]^{+} \\ &= E[(W-q)^{+} \mid W > q]P[W > q] + E[(W-q)^{+} \mid W \le q]P[W \le q] \\ &= E[(W-q)^{+} \mid W > q]P[W > q]. \end{aligned}$$

Thus, the discrete overbooking cost function is

$$O_d(y) = cE[Y]$$
$$= c\sum_{w=q+1}^{q+y}(w-q)P[W=w].$$

As for the revenue function *R(y)*, we must use an overbooking cost function *O(y)* that is continuous with continuous first and second derivatives. Hence, we approximate the inherently discrete random variable *W* as a continuous random variable using the normal approximation to the binomial distribution:

$$O(y) \;=\; c\int_{w=q+1}^{q+y} h_W(w)(w-q)dw$$
$$= c\int_{w=q+1}^{q+y} wh_W(w)dw - q\int_{w=q+1}^{q+y} h_W(w)dw.$$

From the definition of the normal distribution, $h_W(w) = (1/\sigma\sqrt{2\Pi})e^{(w-\mu)^2/2\sigma^2}$, we have

$$O(y) \;=\; (c/\sigma\sqrt{2\Pi})\int_{w=q+1}^{q+y} e^{(w-\mu)^2/2\sigma^2}dw - q\int_{w=q+1}^{q+y} h_W(w)dw.$$

To compute the integral, function int *gsl integration qags*(const *gsl function* * *O(y)*, double $q+1$, double $q+y$, double *epsabs*, double *epsrel*, size *sz limit, gsl integration workspace* * *workspace*, double * *res*, double * *abserr*) from the GNU Scientific Library (GSL) was used. From GSL Team (2006): This function applies the Gauss-Kronrod 21-point integration rule adaptively until an estimate of the integral of *O(y)* over ($q+1$, $q+y$) is achieved within the desired absolute and relative error limits, epsabs and epsrel. The results are extrapolated using the epsilon-algorithm, which accelerates the convergence of the integral in the presence of discontinuities and integrable singularities. The final approximation is obtained from the extrapolation, *res*, and an estimate of the absolute error *abserr*. The subintervals and their results are stored in the memory provided by workspace. The maximum number of subintervals is given by the limit, which may not exceed the allocated size of the workspace. Differentiating *O(y)* with respect to *y*, we have

$$\frac{dO(y)}{dy} = (c/\sigma\sqrt{2\Pi})e^{-(q+y-\mu)^2/2\sigma^2}$$

$$\frac{d^2O(y)}{dy^2} = (c/\sigma\sqrt{2\Pi})e^{-(q+y-\mu)^2/2\sigma^2}[y(\mu-q-y)/\sigma^2+1].$$

Finally, the overbooking cost function for the entire network *O(y)* is simply the sum of the overbooking costs over the flight legs.

1.3 Hybrid Newton's and Steepest Ascent Method

In this section, we seek to find the maximum on the profit curve *Z(y)*, which has been constructed to be continuous and have continuous first and second derivatives. Gradient methods are used for either minimization or maximization of a function. These methods use an iterative formula that contains the gradient of the function to find the minimum or maximum, hence, the name gradient methods. Newton's method is said to be the fastest of all the gradient methods (Bertsekas 1999). Upon implementing Newton's method, it was observed that the smoothed MARS revenue function *R(y)* is still mostly composed of linear or nearly linear components; thus, its second derivative and the determinant of the profit function's Hessian matrix were frequently (near) zero. As a consequence, we developed a hybrid Newton's and steepest ascent algorithm, shown in Figure 3 and specified in Algorithm 1.

Figure 3 about here.

Our steepest ascent implementation estimates the stepsize *k* using the Armijo Rule (Bertsekas 1999), so the next point is given by

$$x_{k+1} = x_k - \gamma_k \nabla_x Z(x_k).$$

Algorithm 1 Hybrid Newton's and steepest ascent algorithm.

Initialize $x_k = q, k = 0, \gamma_k = 1$.

Step 1: Find $Z(x_k)$, $\nabla_x Z(x_k)$ and $\nabla^2_{xx} Z(x_k)$.

Step 2

if norm of $\nabla_x Z(x_k) = 0.0$ **then**

Stop. Maximum profit is obtained at x_k.

else

Step 3: Find the determinant (det) of $\nabla^2_{xx} Z(x_k)$.

if det=0 **then**

Use the steepest ascent method to find the next point.

$x_{k+1} = x_k - \gamma_k \nabla_x Z(x_k).$

Update step size using the Armijo Rule.

Go to Step 1.

else

Use Newton's method.

$x_{k+1} = x_k - \nabla_x Z(x_k)\{\nabla^2_{xx} Z(x_k)\}^{-1}$.

Go to Step 1.

end if

end if

2 Computational Results

Our methodology was tested on the 31-leg airline hub application used by Chen, Günther and Johnson (2003), Siddappa (2006), and Siddappa, Günther, Rosenberger and Chen (2007). Their work applied statistical modeling approaches for revenue management, which use the value functions from a Markov decision process to estimate the fair market value of a seat over time; thus, fares will change over time depending on the remaining capacities. Their application used fifteen reading dates, included 123 itineraries, and had a maximum demand of 50. Data on flight capacities and demand distribution parameters were provided by a domestic airline carrier. To determine the revenue function *R(y)*, we conducted DACE using a strength two orthogonal array experimental design with 31^2 points, and we fit a version of MARS that uses an automatic stopping rule (Tsai and Chen 2005). The experimental design covered overbooking levels from 0 to 20% capacity ($0.2*q$).

We compared our optimized overbooking approach to two variants of a simple probabilistic approach. The probabilistic approach considers the probability that the number of customers that show up is greater than capacity ($P[W>q]$). Anecdotal evidence suggests that this approach is prevalent in the airline industry. The two variants set the overbooking levels, so that this probability is less than 0.05 and 0.01, respectively. Table 1 shows the overbooking levels (OL) from the three approaches. Since the revenue function *R(y)* depends on the load factor (LF), defined to be the quotient of total requested capacity over available capacity, our optimized approach is presented for two load factors. Airlines use nominal load factors of up to 150% to test revenue management methods.

The overbooking levels from the three approaches were all tested using the statistical modeling approach of Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) and compared to the case of no overbooking. Table 2 presents the average profit (revenue − overbooking costs) from 2000 simulation runs for different load factors (LF) and coefficients of variation (CV). The standard errors and percent increases from no overbooking are given below each average profit. It can be seen that overbooking generally provides higher profit than no overbooking, and that our optimized approach provides higher profit than all other cases. It should be noted that the run times for fitting the MARS approximation and conducting the hybrid optimization were about 1.2 minutes and one minute, respectively, on a Dual 2.8-GHz Xeon Workstation.

3 Concluding Remarks

The presented DACE-based optimized overbooking approach provides a computationally practical method for achieving higher profit through overbooking than a simple probabilistic approach. Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) demonstrated that their statistical modeling approach for revenue management, based on a Markov decision process, is computationally tractable and provides higher revenue than state-of-the art bid pricing approaches. The overbooking optimization presented in this paper was implemented as an overbooking module within the statistical modeling approach. For future work in the derivation of the cost function, a more realistic customer show-up pattern and different estimates of overbooking penalties can be considered.

4 Acknowledgements

We would like to thank the referees for there comments that improved the quality of this paper.

References

Alstrup, J., Andersson, S. E., Boas, S. and Madsen, O. B. G.: 1989, Booking control increases profit at Scandinavian airlines, *Interfaces* **19**, 10–19.

Bertsekas, P. D.: 1999, *Nonlinear Programming*, Athena Scientific, Belmont, MA.

Bertsimas, D. and de Boer, S.: 2005, Simulation based booking limits for airline revenue management, *Operations Research* **53**, 90–106.

Bodily, S. E. and Pfeifer, P. E.: 1973, Overbooking decision rules, *OMEGA* **20**, 129–133.

Bose, R. C. and Bush, K. A.: 1952, Orthogonal arrays of strength two and three, *Annals of Mathematical Statistics* **23**, 508–524.

Cervellera, C., Chen, V. C. P., and Wen, A.: 2006, Optimization of a large-scale water reservoir network by stochastic dynamic programming with efficient state space discretization, *European Journal of Operational Research* **171**, 1139–1151.

Fang, K. T. and Wang, Y.: 1994, *Number-Theoretic Methods in Statistics*, Chapman & Hall, New York, NY.

Haykin, S.: 1999, *Neural Networks: A Comprehensive Foundation* (2nd Edition), Prentice Hall, NJ.

Hedayat, A. S., Sloane, N. J. A., and Stufken, J.: 1999, *Orthogonal Arrays*, Springer-Verlag, New York, NY.

Koehler, J. R. and Owen, A. B.: 1996, Computer experiments, In *Handbook of Statistics* (S. Ghosh and C. R. Rao, eds.), Elsevier Science, New York, NY, 261–308.

McKay, M. D., Beckman, R. J., and Conover, W. J.: 1979, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* **21**, 239–245.

Niederreiter, H.: 1992, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, PA.

Chatwin, R. E.: 1993, *Optimal Airline Overbooking*, PhD thesis, Stanford University.

Chatwin, R. E.: 1998, Multiperiod airline overbooking with a single fare class, *Operations Research* **46**, 805–819.

Chen, V. C. P.: 2001, Measuring the goodness of orthogonal array discretizations for stochastic programming and stochastic dynamic programming, *SIAM Journal of Optimization* **12**(2), 322–344.

Chen, V. C. P., Günther, D. and Johnson, E. L.: 2003, Solving for an optimal airline yield management policy via statistical learning, *Journal of the Royal Statistical Society, Series C* **52**(1), 1–12.

Chen, V. C. P., Tsui, K.-L., Barton, R. R. and Allen, J. K.: 2003, A review of design and modeling in computer experiments, in R. Khattree and C. R. Rao (eds), *Handbook of Statistics: Statistics in Industry*, Vol. 22, Elsevier Science, Amsterdam, pp. 231–261.

Chen, V. C. P., Tsui, K.-L., Barton, R. R. and Meckesheimer, M.: 2006, Design, modeling, and applications of computer experiments, *IIE Transactions* **38**, 273–291.

D'Sylvia, E.: 1982, O and d seat assignment to maximize expected revenue, *Technical Report unpublished*, Boeing Commercial Airplane Company.

Friedman, J. H.: 1991, Multivariate adaptive regression splines, *Annals of Statistics* **19**(1), 1–141.

GSL Team: 2006, *GNU Scientific Library- Reference Manual.*

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

Kleijnen, J. P. C.: 2005, An overview of the design and analysis of simulation experiments for sensitivity analysis, *European Journal of Operational Research* **164**(2), 287–300.

Kosten, L.: 1960, Een mathematisch model voor een reservings probleem, *Statistical Neerlandica* **14**, 31–37.

Lee, T. and Hersh, M.: 1993, A model for dynamic airline seat inventory control with multiple seat bookings, *Transportation Science* **27**, 252–265.

McGill, J. and van Ryzin, G. J.: 1999, Revenue management: Research overview and prospects, *Transportation Science* **33**, 233–256.

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

Sacks, J.,Welch,W. J., Mitchell, T. J. andWynn, H. P.: 1989, Design and analysis of computer experiments, *Statistical Science* **4**, 409–423.

Siddappa, S.: 2006, *Statistical Modeling Approach to Airline Revenue Management with Overbooking*, PhD thesis, University of Texas at Arlington.

Siddappa, S., Chen, V. C. P. and Rosenberger, J. M.: 2006, Multivariate adaptive regression splines and its derivative, Technical Report *IMSE 06-08*, The University of Texas at Arlington, Arlington, TX.

Siddappa, S., Günther, D., Rosenberger, J. M. and Chen, V. C. P.: 2007, Refined experimental design and regression splines method for network revenue management, *Journal of Pricing and Revenue Management* **6**(3), 188–199.

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

Tang, B.: 1993, Orthogonal array-based Latin hypercubes, *Journal of the American Statistical Association* **88**, 1392–1397.

Tsai, J. C. C. and Chen, V. C. P.: 2005, Flexible and robust implementations of multivariate adaptive regression splines within a wastewater treatment stochastic dynamic program, *Quality and Reliability Engineering International* **21**, 689–699.

Tsai, J. C. C., Chen, V. C. P., Chen, J., and Beck, M. B.: 20045, Stochastic dynamic programming formulation for a wastewater treatment decision-making framework, *Annals of Operations Research* **132**, 207–221.

Vinod, B.: 1995, Origin and destination yield management, in D. Jenkins (ed.), *Handbook of Airline Economics*, The Aviation Weekly Group of the McGraw-Hill Companies, New York, 459–468.

Yang, Z., Chen, V. C. P., Chang, M. E., Murphy, T. E., and Tsai, J. C. C.: 2007, Mining and modeling for a metropolitan Atlanta ozone pollution decision-making framework." *IIE Transactions* **39**, 607–615.

Ye, K. Q.: 1998, Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association* **93**, 1430–1439.

Table Captions

Table: 1 Overbooking levels (OL) obtained from our optimized approach at load factors (LF) 120% and 150%, and the two variants of the simple probabilistic approach.

Table: 2 Average revenues from 2000 simulations of the 31-leg hub using the statistical modeling approach of Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) for the case of no overbooking (OL = 0) and cases using the overbooking levels obtained in Table 1. Various coefficients of variation (CV) are tested for two different load factors (LF). Standard errors are given in parentheses and percent increases compared to no overbooking are shown.

Figure Captions

Figure: 1 Illustration of the three possible outcomes under overbooking, where y is the allowable number of overbooked seats, n_c is the number of cancellations and no-shows, and OD is the origin-destination of the flight.

Figure: 2 Theoretical representation of the profit curve.

Figure: 3 Flow Chart illustrating the generation of the optimal overbooking level y*.

Table 1: Overbooking levels (OL) obtained from our optimized approach at load factors (LF) 120% and 150%, and the two variants of the simple probabilistic approach.

Flight Leg	Capacity	Optimized OL for LF 120%	Optimized OL for LF 150%	$P[W > q]<$ 0.05	$P[W > q]<$ 0.05
1	113	8	9	20	17
2	152	7	7	28	24
3	164	3	5	30	26
4	208	11	12	40	35
5	151	4	5	28	24
6	107	8	9	18	15
7	152	4	5	28	24
8	154	2	4	28	24
9	153	6	7	28	24
10	179	2	4	34	29
11	155	8	8	28	25
12	166	1	2	31	27
13	168	1	2	31	27
14	167	8	8	31	27
15	150	2	2	27	24
16	156	9	10	29	25
17	221	13	13	43	38
18	159	3	5	29	25
19	168	13	14	31	27
20	166	16	22	31	27
21	160	12	12	26	34
22	201	1	5	38	34
23	155	1	3	28	25
24	188	14	14	35	31
25	157	2	5	29	25
26	156	18	20	29	25
27	165	8	12	30	27
28	150	8	10	27	24
29	155	2	5	28	25
30	149	19	24	27	24
31	193	19	20	37	32

Table 2: Average revenues from 2000 simulations of the 31-leg hub using the statistical modeling approach of Siddappa (2006) and Siddappa, Günther, Rosenberger and Chen (2007) for the case of no overbooking (OL = 0) and cases using the overbooking levels obtained in Table 1. Various coefficients of variation (CV) are tested for two different load factors (LF). Standard errors are given in parentheses and percent increases compared to no overbooking are shown.

LF	CV	OL = 0	Optimized	$P[W > q] < 0.05$	$P[W > q] < 0.01$
120	0.32	820075 (467.5)	836917.1 (345.4) 2.05%	825483.3 (414.7) 0.66%	829761.5 (827.4) 1.18%
120	0.45	817417 (428.4)	829462.9 (536.8) 1.47%	823840.7 (722.2) 0.79%	825728.2 (638.4) 1.02%
120	0.6	739939 (632.7)	753143.4 (747.5) 1.78%	726422.5 (683.3) −1.83%	736193.6 (638.4) −0.51%
150	0.48	799575 (425.7)	825714.5 (625.7) 3.27%	818534.1 (535.2) 2.37%	820472.7 (637.4) 2.61%
150	0.56	794365 (700.6)	818648.7 (725.5) 3.06%	805242.1 (833.2) 1.37%	813274.5 (683.5) 2.38%
150	0.70	753298 (737.8)	773779.5 (764.2) 2.72%	756328.4 (628.4) 0.40%	765982.7 (759.2) 1.68%

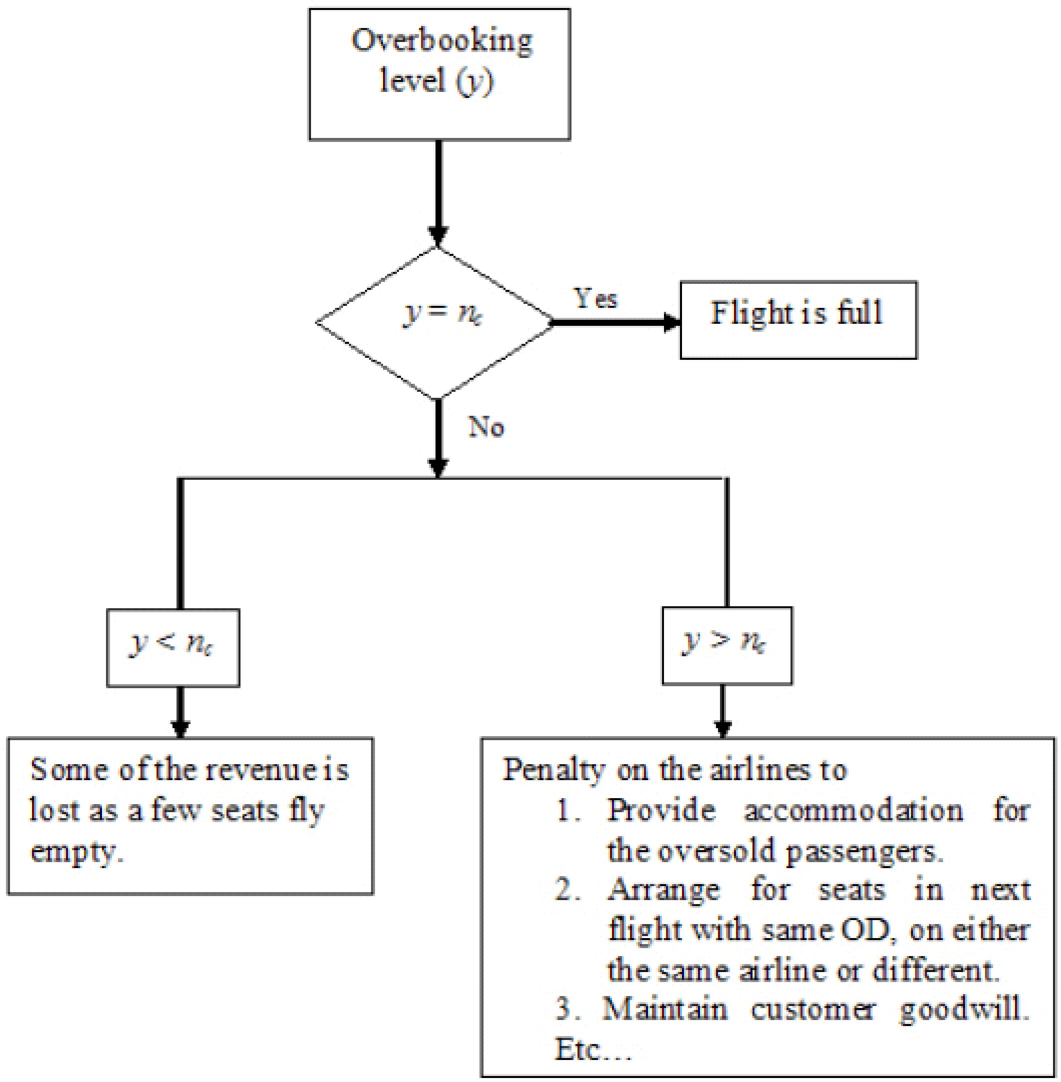

Figure 1: Illustration of the three possible outcomes under overbooking, where y is the allowable number of overbooked seats, n_c is the number of cancellations and no-shows, and OD is the origin-destination of the flight.

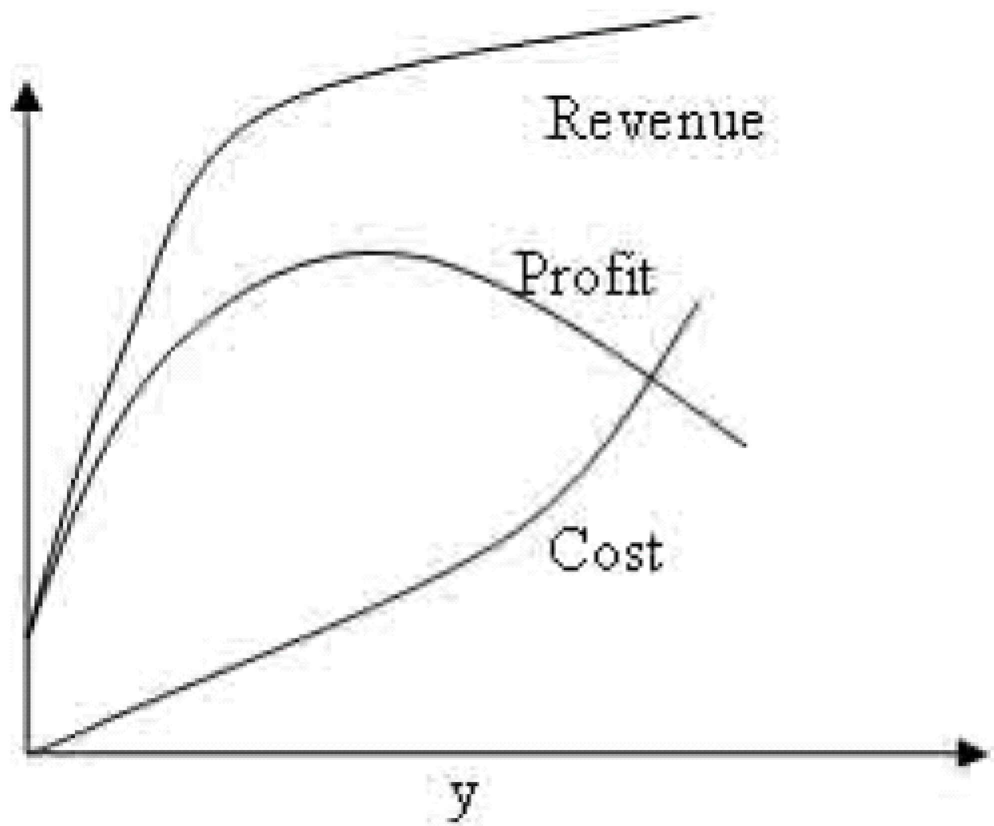

Figure 2: Theoretical representation of the profit curve.

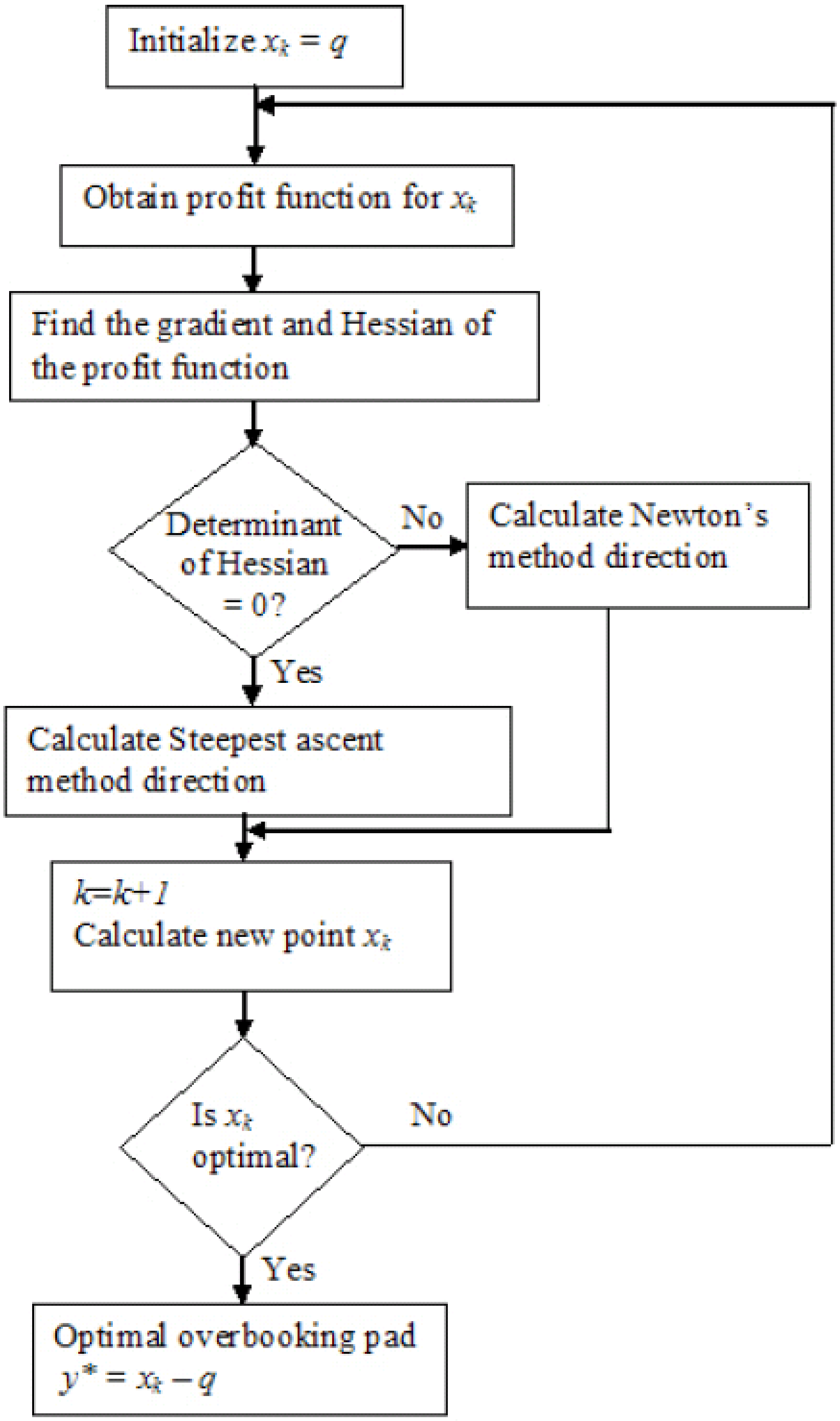

Figure 3: Flow Chart illustrating the generation of the optimal overbooking level y^*.