OVERBOOKING IN AIRLINE REVENUE MANAGEMENT

TANG YANPING

NATIONAL UNIVERSITY OF SINGAPORE

2003

OVERBOOKING IN AIRLINE REVENUE MANAGEMENT

TANG YANPING

(B.Sc.(Hons), ECNU)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2003

Acknowledgements

This report would not have been possible without the help, invaluable suggestion and patient guidance from my supervisor, Associate Professor Zhao Gong Yun. If not for him, I would not have learned so much. Thank you very much, sir, for everything you have done for me!

During the 10 months that Prof. Zhao was on study leave, my seniors and my friends, Wan Mei and Chee Khian gave me generous guidance and encouragement during that period. My heartfelt appreciation also goes to Dr Tan Geok Choo, Prof. Sun Defeng, Prof. Toh Kim Chuan, Prof. Koh Khee Meng, etc. for your help and encouragement, I really admire you for your dedication to teaching.

I would like to say a very special thank you to my parents. You always encourage me and support my decision, your words give me great motivation in my study and in my life. I am also extremely grateful to my boyfriend, who has never failed to offer me love and encouragement. Furthermore, I would like to express a million thanks to all my friends in Singapore and those in other countries. I may not list all your names here because there are really too many to be listed, your friendship and encouragement keep me on the hard-working track to finish this report.

A million thanks are also extended to others including the friendly and helpful academic and non-academic staffs in Department of Mathematics.

Last but not least, I would also wish to thank the National University of Singapore for awarding me the Research Scholarship which financially supports me throughout my M.Sc. candidature.

This report is dedicated to you all for your help, encouragement, advice, love and understanding.

Tang Yanping, Helen

2003

Contents

List of Notations

- *Time:*

T	Length of time horizon (number of periods), in reverse order.
t	Time periods left until departure (count-down).

- *Fares, refunds and penalties:*

f	Single fare class, $f > 0$.
f_i	The fare category of demand class i, time-independent.
$\widehat{f}_{it}$	The revenue that airline earns if the booking agent accepts a request for a seat in fare class i at t (Charging different prices at different points in time).
R^c	The refund to the customer who cancels.
R^{ns}	The refund to the customer who is a no-show.
R^o	The overbooking penalty/Denied boarding cost.
R^{sp}	Spoilage cost per passenger, which is the revenue lost by not being able to fill the capacity due to show up falling short of capacity.

- *Capacity and Booking limits:*

C	Capacity (Physical seats).
Q	Overbooking Pad, i.e. how much to overbook.
B	Booking Limit for all fare classes/Overbooking level, i.e. the maximum number of bookings will be accepted by the airline.
B_i	The booking limit for fare class i.
B_{it}	The booking limit for fare class i at time period t.

- *Expected Revenue:*

x	The current number of reserved seats.
$\mathbf{x} = (x_1, \cdots, x_m)$	The reservation vector, where x_i denotes the number of seats currently reserved in fare class i.
$U_t(x)$	The maximum total expected net revenue of operating the system from period t to 0.

- *Demand and Cancellation Process:*

p_{it}	Prob. of a booking request for a seat in fare class i at time t.
q_{it}	Prob. of a class i cancellation occurring at time t.
p_0^t	Prob. of no request (reservation or cancellation) at time t.
D^t	Demand (to come) process (m-dimensional).
$\overline{D}^t$	Aggregate demand (to come) distribution (m-dimensional).
$\mathbf{D}^t = E[\overline{D}^t]$	Expected aggregate demand to come (m-dimensional).

- *Others:*

S	Survivals, i.e. those who bought the ticket and show up.
β	The probability for each customer holding a seat reservation to be a no-show at the time of departure (Same for all fare classes).
β_i	Prob. of each customer in fare class i being no-show.
$\alpha = 1 - \beta$	The probability of surviving that does not depend on when the reservation was booked and independent of other customers. i.e. show-up rate.
α_i	The show-up rate for fare class i.

Summary

In the airline industry, it is of crucial importance to optimize passenger bookings as this is a main source of income for the airline. Even when a flight is booked solid, there is a possibility of a passenger not showing up at the departure time resulting in an empty seat which otherwise could earn a revenue for the airline. It is common knowledge that once an aircraft departs, the revenue from the empty seats on that flight will never be recouped. In an attempt to reduce vacant seats, airline resorts to "Overbooking" — that is, accepting more reservations than the capacity of the aircraft which is effective at increasing load factors and revenues.

Overbooking problem may seem simple. However, beneath that surface impression, a good deal of complexity lurks. The crux of the problem lies in how much to overbook. Due to the unpredictable nature of passengers' behavior, there is a great degree of variance in the number of people who cancel the reservations or do not show up for a particular flight. Consequently, numerous flights end up taking off with empty seats while other flights end up denying some passengers' boarding.

The number of articles that have been published in the area of Airline Overbooking Problem is relatively not big, in spite of the huge financial impact of a yield management system. This is partly due to the fact that overbooking is part of yield management, which is a strategic tool to increase corporate profitability and most airlines generally do not publish their yield management approaches, models

and implementation aspects due to their proprietary nature. We tried our best to find all possible important papers published up to this day.

In **Chapter One**, efforts are made to survey the important results in this field. We give a rough overview of the airline overbooking problem with regards to the overbooking models in use today, and analyze 3 different techniques for the airline overbooking problem: Static Models on Single Leg, Dynamic Models on Single Leg, and the Models in Network Environment(Corresponding to Chapters 2, 3 and Chapter 4). Furthermore, we will explain the difference between the defined Static and Dynamic models in details.

Chapter Two focuses mainly on Static Overbooking Problem on the single-leg, which is separated into two sections 2.2 and 2.3. We describe 3 rules for the Single-class, Single-leg problem in Section 2.2. The model for rule one is similar to the one by Beckmann (1958) [3]. The models of Rule Two and Three are same as Bodily and Pfeifer's (1992) in [10]. We complete their proofs in our report. In section 2.3, we revised Littlewood's rule (1972) [39] to include overbooking for a single-leg, two fare classes case based on the nested reservation system. Similar to what Belobaba (1987) [5] has done, we extend the results to the overbooking problem for multiple fare classes, which we call revised EMSR (rEMSR) to find the protection levels for higher fare classes from lower ones. We present an example to show that, given all these protection levels in the nested reservation system, how the optimal booking limits are determined to maximize the total expected revenue.

The Dynamic Model on the single-leg is the subject of **Chapter Three**. The main difficulty for overbooking models with cancellations in the dynamic environment lies in the fact that there are two concurrent stochastic processes: booking and cancellation. We discuss two models from Janakiram, Stidham, and Shaykevich (1999) [54]. The booking control policy is proposed and the optimality of the policy is proved in Section 3.3 which is not provided in [54]. Model 2 considers

the refund to cancellations and no-shows. The model is in multi-dimension as cancellation and no-show probabilities are fare-dependent. In this report, the way how a multi-dimensional problem is converted into a one-dimensional problem was represented. We tried to make the steps more clear and we completed some proofs which were not provided in [54]. A numerical example is quoted to show that we do not always need the full multi-dimensional model, and to imply several other important results.

In **Chapter Four**, the Airline Overbooking problem is put into the network environment. Bertsimas and Popescu (2001) [7] proposes two approximate dynamic programming algorithms: Bid-price Control and Certainty Equivalent Control (CEC belongs in the class of approximate dynamic programming mechanisms in which the cost-to-go function is approximated by the value of a linear programming relaxation). We discuss these two algorithms handling cancellations and no-shows by incorporating overbooking control in the underlying mathematical programming formulation in depth. We extend the results from [7] by providing and proving structural properties of the two algorithms allowing overbooking which [7] hasn't considered. These results offer insights into the behavior of both algorithms. One computational example is quoted to show that the CEC policy improves upon the performance of the bid price control policy.

Finally, in **Chapter Five**, we conclude this report with an overview and suggest a direction for future research in the integration of revenue management.

For clear interpretation, a glossary of sometimes-confusing terminologies used in Overbooking problems in Airline Revenue Management is provided in **Appendix 1** as they can be very useful for the future researchers in this area. We also collect and outline some important results throughout the literature of Overbooking Problem in **Appendix 2**.

Chapter 1

Airline Overbooking Problem

1.1 Introduction

Airline industry is one of the capacity constrained services, such as transportation, tourism, entertainment, media and internet providers. They all constantly face with the problem of intelligently allocating the fixed capacity of perishable products to demand from different market segments, with the objective of maximizing total expected revenue. For the airline, a seat on any particular flight departure is an extremely perishable commodity. Once the doors are closed on a plane, the value of any unsold seats is lost forever.

Revenue management originates from the airline industry, where deregulation of the fares in the 1970's led to heavy competition and the opportunities for revenue management schemes were acknowledged in an early stage. Revenue management can be defined as the art of maximizing profit generated from a limited capacity of a product over a finite horizon by: *selling the right product to the right type of customer, at the right time and for the right price.* But, this process involves consumer behavior and past data analysis, it can be very challenging.

The airline revenue management problem has received a lot of attention throughout the past years and will continue to be of interest in the future. Smith et al. (1992) [52] describe the airline revenue management problem as a non-linear, stochastic, mixed-integer mathematical program that requires data such as passenger demand pattern, cancellations, group reservations, cargo load, and other estimates. Solving this problem would require approximately **250** million decision variables! For the sake of feasibility and time, it has been reduced to three distinct smaller problems: Overbooking, Discount allocation, and Traffic management in [52].

The air transport industry operates its passengers service almost entirely on the basis of pre-reservation. Passengers who make reservations may, with minor exceptions, cancel them or even not show up at the departure time without economic penalty. Airlines, in turn, compensated for this flexibility by taking reservations in excess of the capacity, i.e. overbooking. By this, the planes would not so often depart with empty seats for which there was a demand.

> As long as forty years ago, the major U.S. Airlines had a significant "no-show" problem. In 1961, the CAB (Civil Aeronautics Board) reported that the 12 leading carriers were experiencing a very significant no-show rate: only 1 out of 10 passengers actually boarded. This statistic resulted from an investigation undertaken because of reports, ultimately confirmed, that several major carriers were deliberately overbooking. In the sixties, the so-called "no-shows" were becoming a major problem for airlines who found they had many flights that were fully booked departing with empty seats.— Rothstein (1985) [46]

So, in fact, the airline overbooking problem arises from the propensity of airline customers, who have made a reservation for a flight, to subsequently cancel that reservation or make a no-show. In airline revenue management, cancellations refer to return or changes of booked seats prior to flight departures, which can be rebooked in the future, while no-shows refer to passengers that do not check in without notifying the airline in advance which lead to ultimate vacancies. In anticipation that cancellations and no-shows will occur, the airline may overbook the flight, thereby reselling a seat vacated by a customer who cancels or will be a no-show. The potential extra revenue from overbooking a flight must be balanced against its costs. This arises because in overbooking, the airline runs the risk of not having sufficient capacity which is relatively fixed[1] to accommodate all its customers, in which case it must deny reservation requests or deny boarding to some of them (i.e. bumping), thereby incurring a cost measured both financially and in loss of goodwill.

One might think that a good strategy would be to avoid overbooking completely in the attempt to keep all customers satisfied. However, because of passengers' uncertainties, airlines have to adjust the policies to offset the effects of passenger cancellations and no-shows, which is necessary and not so easy. Without overbooking, it is estimated that 15 percent of seats would be spoiled on sold-out flights [52]. Figure (1.1) from [52] shows that when there are cancellations, the capacity of a plane can only be filled through overbooking.

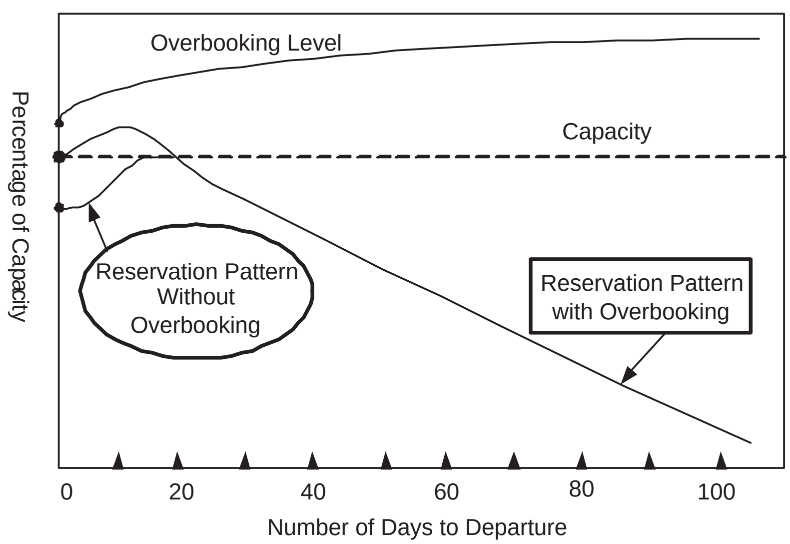


Figure 1.1: Overbooking allows more reservations to be accepted. For flights close to departure, there are more reservations accepted with overbooking to compensate for cancellations and no-shows.

While unpopular with passengers, overbooking is effective at increasing load factors and revenues. This raised the issue of determining the right booking limit

[1]The reason for this characteristic is very simple. If capacity were flexible, there would be no need for a tradeoff. If airlines could add or remove seats on aircraft at will, there would be no reason to try to manage capacity. Unfortunately, the plane cannot be enlarged, the only flexibility allowed is to schedule the passenger on a later flight.

(Overbooking Level), which is the maximum number of seats that can be sold to passengers. The level of overbooking for each class of passenger has been the topic of research for many years. If the booking limit is set too low, there will be lots of empty seats. On the other hand, if the booking limit is set too high, the benefits of filling the aircraft would be overwhelmed by the denied boarding costs. Determining the optimal booking limit is the focus in the airline overbooking problem. And, the airline has the opportunity to change the limit for the latest demand forecast and changing human behaviors as departure approaches.

The following section will provide some main results in literature of airline revenue management and trace the development of “overbooking” concept.

1.2 Models in Use

Current models can be grouped as leg-based and network-based. Leg-based methods are aimed at optimizing the expected revenue on a single-leg flight. Network-based models consider booking requests for multiple legs at the same time. In either case, the booking control policy can be static, in which decisions are based on pre-calculated booking limits, or dynamic, where the decision rules will be changed during the booking period.

In details, the entire network of the flight can be separated into smaller flight legs. The leg-based airline overbooking policy allows the airline to maximize the total expected revenue from each leg, setting booking limits on all the fare classes available in that leg. Therefore, reservations on that flight leg are accepted based on the availability of a particular fare class on that leg. A passenger’s ultimate destination, overall itinerary[2] which includes multiple legs, or total revenue contribution to the airline is not taken into account.

For the single leg airline overbooking problem, as mentioned above, the booking control policy can be *Static* or *Dynamic*: those assuming that the demands fare

[2] Airline, typically, offers tickets for many origin-destination itineraries in various fare classes.

classes[3] arrive separately in a predetermined order and we get one-time setting of booking limits for each class (Static), and those allowing customers of different fare classes to book concomitantly, and we may change the booking limits during the booking period (Dynamic).

If the route structure of an airline served each distinct origin and destination (OD) market with isolated, non-stop, point-to-point flights, as shown in Figure [1.2], a Leg-based approach would be all that was necessary. However, in real world situations, the typical airline route structure is a more complex network built around one or more connecting hubs, as shown in Figure [1.3].

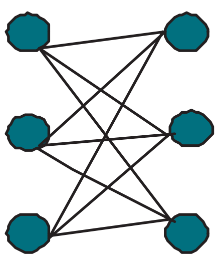


Figure 1.2: Distinct origin and destination (OD) market with isolated, non-stop, point -to-point flights.

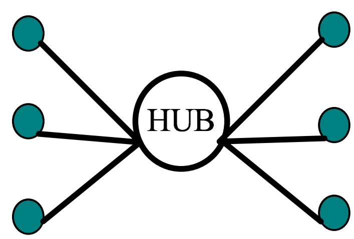


Figure 1.3: Network is built around one connecting hub.

A major flaw of 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. Overbooking control focusing on individual flight legs does not guarantee that revenues will be maximized across an entire network of flight, as Williamson (1992) [62] stated.

[**Example:**] Consider a passenger travelling from A to C through B. That is, travelling from A to C using flight legs AB and BC. If the single leg approach is used, this passenger can be rejected on one of the flight legs because another passenger is willing to pay a higher fare on this flight leg. But by rejecting this

[3]In this report, the terminology 'fare classes' actually refers to the buckets. I do not concern myself with how airlines define their fare classes because the model presented in this report is independent of the method of fare classification.

demand, the airline loses an opportunity to create revenue for the combination of the two flight legs. If the other flight leg does not get full, it could have been more profitable to accept the passenger to create revenue for both flight legs.

Hence, determining an overall booking control strategy for the entire network is far from trivial. Network overbooking control allows the airline to differentiate between the many types of fares[4] and the variety of itinerary values determining seat allocations. The purpose of such control manages overall network traffic, limiting sales by origin-destination itinerary, as well as fare class by methods which incorporate mathematical programming and network flow techniques.

1.2.1 Static Overbooking Problem on Single Leg

We could unearth no scientific work or even discussion of the overbooking problem published earlier than 1958. In that year, an article by Beckmann (1958) [3] employing a static one-period model with reservation requests, booking, and finally cancellations was issued. It contains a mathematical model to determine the booking level that minimized the lost revenue due to empty seats plus the costs of over-sales. The model of Kosten (1960) [31] has the same objective but is more exact, in that it provides the interspersion of reservations and subsequent cancellations (which Beckmann ignored, as he took it that all cancellations occurred at departure time). An easier-to-implement model is published by Thompson (1961) [59], which entirely ignores the probability distribution of passenger demand as well as costs, and which requires data only on the cancellation proportions out of any fixed number of reserved passengers. Thompson's work influenced much subsequent research.

The next important work was done by Taylor (1962) [58]. He adopted Thompson's approach and presented a model similar in spirit, but featuring a much more exact treatment of cancellation, no-shows and group sizes. Deetman (1964) [20] at

[4] These fare classes not only include business and economy class, which are settled in separate parts of the plane, but also include fare classes for which the difference in fares is explained by different conditions for cancellation options or overnight stay arrangements or etc.

KLM studied Taylor's model to test its behavior and implementability. Rothstein and Stone (1967) [47] developed a computer system for booking levels by using a slightly simplified version of the Taylor's model and capitalizing on the copious cancellation statistics available from Sabre. Belobaba (1987) [5] and in part of his Ph.D. dissertation [4], discussed the problem of overbooking in multiple fare classes and suggested a heuristic approach to solve the problem. American Airlines implemented (in 1976, with a major revision in 1987 [1]) such a model with additional constraints to ensure that the level of service was not overly degraded (Smith et al. 1992) [52]. More review in this area is given by Rothstein (1985) [46] and is further discussed in Chatwin (1993, chapter 1) [15]. Chatwin dealt exclusively with the overbooking problem and provided a number of new structural results. More Recent work on the static overbooking problem is discussed by McGill (1989) [40], Bodily and Pfeifer (1992) [10].

In Chapter Two, we critically explore two simple models (single-fare-class) and generate 3 Rules corresponding to different policies to determine the optimal booking limit. The model for rule one is similar to the one by Beckmann (1958) [3]. The model for Rule Two and Three are same as in [10]. The rest of this Chapter is about static multi-fare-class problem on Single-leg. We revise Littlewood's rule (1972) [39] including overbooking for a Single-Leg, two-fare-class case. Similar to what Belobaba (1987) [5] has done, we extend the result to get a revised EMSR method for multi-fare-class case considering overbooking. Finally, we propose a method incorporating overbooking based on the nested reservation system to find out the optimal overbooking level and the optimal booking limits for each fare class. The results generated by these solutions are optimal under the sequential arrival assumption as long as no change in the probability distributions of the demand is foreseen.

1.2.2 Dynamic Overbooking Problem on Single Leg

A Drawback of the aforementioned models is that the dynamic nature inherent in the reservations process and cancellation process is not considered. In the "Dynamic Overbooking Problem", the demand for each fare class is modelled as a time-dependent process, where the inter-arrival time is lengthen or shorten as the scheduled departure time approaches. Dynamic solution methods do not determine the booking limits at the start of the booking period as the static solution methods do. Instead, we should monitor the state of the booking process over time and decide whether to accept a particular booking request when it arrives or reject it, based on the state of the booking process at that point in time.

Rothstein (1968, 1971) [43][44] first formulated the airline overbooking problem as a dynamic programming model, and he later did the same for the similar hotel overbooking problem in [45]. Hersh and Ladany (1977) [35] modelled flights with an intermediate stop using dynamic programming, and Ladany [32] [33] developed models for the hotel/motel industry, and considered the extension to two or more fare classes.

A characterization of the optimal dynamic policy based on a threshold time property was done by Diamond and Stone (1991) [21], and later by Fend and Gallego (1995) [25]. Lee and Hersh (1993) [38], considered a discrete time dynamic programming model, where demand for each fare class is modelled by a non-homogeneous Poisson process. Using a Poisson process gives rise to the use of a Markov decision model. They also provided an extension to their model to incorporate batch arrivals. Janakiram, Stidham, and Shaykevich (1999) [54] extended the model proposed by Lee and Hersh to incorporate cancellations, no-shows and overbooking. They also considered a continuous time arrival process as a limit to the discrete time model by increasing the number of decision periods. In Chapter Three, we will discuss more in details of two dynamic models which permit cancellations, no-shows and overbooking. The difference between "Static Models" and "Dynamic Models" defined in this report will also be described in section 1.3.

1.2.3 Network Model

Revenues are maximized for each individual flight leg in the research described above, but the flow of traffic and the interaction between flight legs are not taken into account.

In the *Static network models*, the core problem is determining optimal decision rules for sequentially accepting or denying Origin-Destination-Fare (ODF) itinerary requests at the start of the booking period. We can create such model incorporating probabilistic demand and solve it by probabilistic mathematical programming techniques. Alternatively, we can simplify the problem by substituting uncertain demand by its expectation, which allows the use of deterministic mathematical programming. Booking control can be implemented in various ways. We can aim at determining booking limits. Booking requests are rejected if the respective booking limits would otherwise be exceeded. An alternative form of booking control that can be derived from the dual form of these models is based on *bid-prices*.

Bid-price control is, perhaps, one of the hottest decision rules in the last 10 years. Instead of setting booking limits, this approach assigns a bid-price to each of the flight legs in the network. The simple rule for this control policy is: A booking request for an ODF itinerary is accepted if and only if the associated fare exceeds the sum of the bid-prices of those legs along the itinerary. Simpson (1989) [51] and Williamson (1992) [62] first studied this method and proposed approximations to generate bid prices based on dual prices of various mathematical programming formulations of the problem. The mathematical programming approach will handle realistically large problems and will account for multiple origin-destination itineraries and additional constraints. Even though, in general, the bid-price controls are not optimal, they can still provide asymptotic optimal bid-prices when the leg capacities and the sales volumes are large.

In fact, the most practical and relevant, yet least investigated model for Network Revenue Management (NRM) is the *dynamic network model*. Talluri and Van Ryzin (99A,B) [56] and [57] studied a dynamic network model using bid-price control mechanisms, argued why bid-price policies are not optimal, and provided an asymptotic regime when certain bid-price controls, based on a probabilistic programming formulation of the problem, are asymptotically optimal. Gunther and Johnson (1998) [28] formulated the problem as a Markov Decision Problem, and used linear programming and regression splines to approximate the value function. In the following years, they introduced a new method to compute bid prices for single hub airline network. However, none of these NRM-approaches which base on additive bid prices handles cancellations and no-shows. Actually, most of the network models ignore the cancellations, no-shows and overbooking.

However, we can use *the typical overbooking method* to decide an initial allocation of overbooking pads, which are virtual increased in leg-capacity. By such method, we can handle cancellations and no-shows to some extent.

Ladany and Bedi (1977) [34], and Hersh and Ladany (1978) [36], considered the overbooking problem in the network environment which incorporated the time distribution at which reservations and cancellations were actually made. Dror et al. (1988) [22] also proposed a method by using a network flow representation of the problem incorporating both cancellations and no-shows.

Bertsimas and popescu (2001) [7] proposed a new algorithm — Certainty Equivalent Control, also handling cancellations and no-shows by incorporating oversales decisions in the underlying linear programming formulation. This policy conceptually improves the current NRM-approach which bases on additive bid-pricing, by using more insightful, piecewise linear approximations of opportunity cost. They also reported more encouraging computational performance than Bid-price control. We will look through that algorithm incorporating the cancellations, no-shows and overbooking and obtain more results in Chapter Four.

1.3 Models on Single Leg: Static vs. Dynamic

The simpler approach to the Single-leg airline overbooking problem is often solved by using the static models to find one-time setting of booking limits, while more complicated approaches use historical data, competitors' actions, and current trends to set initial booking limits, and then to make adjustments in these limits as bookings materialize, which are indicated as dynamic models. Obviously, the static models are less data demanding and, hence, have been well accepted by the airline industry, even though they are not so practical. Therefore, it is compelling to compare the static models from the dynamic models.

Static Overbooking Problem

Focus:

What limit to place on booking for each fare class, considering the cancellation and no-show behaviors at the beginning of the booking process based on demand estimates. Once the booking limits are calculated, they will not be changed until the flight departs (Time-independent).

Assumptions:

1. Fare classes are booked sequentially[5], but not exclusively in order of increasing fare level. Once the bookings of one fare class stop, it will not be reopen again.

2. We do not consider the passenger arrival process over time, requiring instead only the total demand for each class. It ignores the airline reservation process, precisely, the stochastic evolution of demand over time.

3. We assume that arrival time uniquely determines the class of each request. In other words, within each time period, all arriving customers request the same fare.

4. Customers can cancel their reservations (cancellations) or simply do not utilize their reservation (no-shows), getting full, partial or no refund, which is depending on the fare category.

5. Statistical independence of demands between booking classes[6].

6. Single flight leg with no consideration of network effects.

[5] This is a common assumption in many of the earlier papers (e.g. Belobaba (1987) [4], Belobaba (1989) [6], Wollmer (1992) [63], Brumelle and McGill (1993) [14] and Robinson (1995) [42]).

[6] No information on the actual demand process of one fare can be derived from the actual demand process of another fare.

7. No demand recapturing which implies that every customer has got a strict preference for a certain fare class and that a denied request is lost forever.
8. No batch booking which justifies looking at one booking request at a time.

Solution Technique:

Under some assumptions, create the total expected revenue function or the total expected cost function, then try to find the optimal booking limit for each fare class to maximize the revenue function or to minimize the cost function.

Demand Data Needed:

We need the estimation of the probability distribution of the total demand for each fare class.

Dynamic Overbooking Problem

Focus:

Whether to accept a particular reservation request at its particular arrival time, considering the dynamic characteristics of the cancellation and no-show behaviors (Time-dependent Booking Limits).

Assumptions:

1. Requests for each fare class can arrive throughout the reservation horizon, no assumptions are made on the arrival order of the fare classes.
2. The demand for each fare class is modelled as a time-dependent process.
3. Customers may cancel their reservations at any time up to the departure of the flight or simply do not utilize their reservation (no-shows), getting full, partial or no refund, which is time and class dependent/independent.

(Same as Assumptions 5-8 in Static Overbooking Problem.)

Solution Technique:

Dynamic programming, using the time remaining until departure (suitably divided into periods or stages) as the index. In order to decide whether or not to accept the booking request, at its particular arrival time, the opportunity cost of losing this seat taken up by the booking has to be evaluated and compared to the revenue generated by accepting the booking request. As the number of periods grows to infinitely, the distribution of total arrivals will converge to a nonhomogeneous Poisson distribution.

Demand Data Needed:
We need the distribution of the customer arrival times.

[**Remark 1:**] The early control systems (Static Models) are based on booking limits, which are typically determined at the beginning of the booking process based on demand estimates. Most carriers which actively control seat inventories have developed or invested in some type of statistical data management and decision support system. These systems collect and store historical reservations data and estimate demand based on historical patterns and forecasting models. This allows airline to respond to changes in booking patterns to update these booking limits as departure time approaches, although it is practically undesirable to recalculate them every time a booking request is made.

[**Remark 2:**] In dynamic models, the demand is modelled as a stochastic process and decision making is performed under uncertainty. At each point of time, the optimal decision should be determined. However, the booking policy of static models is fixed throughout the booking period and does not adapt to unexpected developments in the demand. Due to the intractable computation of the dynamic programming solutions, the static models are often used to approximate dynamic policies, by solving the model at several fixed times during the booking process.

Chapter 2

Static Overbooking Problem (Single Leg)

2.1 Introduction

The *static overbooking problems* focus on setting the booking limits for each fare class to maximize the expected profit for the airline. In detail, for each fare class on Single-Leg, statistical models are applied to historical booking data to forecast the expected demand for all future departure. We use these demand forecasts to determine the booking limits for each fare class at the start of the booking process, incorporating the overbooking factors, to minimize the lost revenue due to empty seats plus the costs of oversales, or to maximize the expected revenue function. Once these limits are set, they will not be changed until the flight takes off. The booking limits for each class give the number of seats available for that class.

In this Chapter, we explore two simple models (single-fare-class) in Section 2.2 and generate 3 Rules corresponding to different policies to determine the optimal booking limit. Rule One is generated and proved completely. The model for rule one is similar to the one by Beckmann (1958) [3]. The model for Rule Two and Three is same as Bodily and Pfeifer's (1992) in [10]. That is a general and practical Static model on Single-leg, which is worthy of further research. In Section 2.3, we concern more about the Static multi-fare-class problem on Single-Leg. We will revised Littlewood's rule (1972) [39] including overbooking for a Single-Leg, two-fare-class problem first. Similar to what Belobaba (1987) [5] did, we extend our result to get a revised EMSR method for multi-fare-class case allowing overbooking. Finally, we will propose one solution incorporating overbooking in the nested reservation system to find out optimal booking limit for each fare class. The papers relevant to our work in this Chapter are [5], [10], [62], [24], [39] and [49].

2.2 Single-fare-class Model

The Single-fare-class on Single-leg problem is considered in most of the earlier works. That is, only one fare class is considered in the reservation system. A passenger comes for a request which can be either a booking request or a cancellation. The airline should find some decision policies to determine whether to accept the booking request or reject it when it arrives. The key point for such static problem is to determine the optimal overbooking level B (i.e. booking limit for this fare class), which is a one-time setting and will not be changed until the flight takes off. The booking requests can be accepted if only if the optimal overbooking level hasn't been exceeded. So, we should find a decision rule to determine how much to overbook the flight in order to minimize the sum of oversale and spoilage costs, or to maximize the expected net revenue which is equal to the total revenue minus expected oversale cost.

Define:

k	The overbooking rate such that $(1+k)\cdot C = B$.
R^o	The overbooking penalty/Denied boarding cost.
R^{sp}	Spoilage cost per passenger, which is the revenue lost by not being able to fill the capacity due to show up falling short of capacity.
α	The show up rate.

I. Rule One

Assumptions:

1. Reservations are accepted on a first-come, first-served basis. No additional reservations will be accepted once bookings have been stopped.

2. The amount of passengers that will show up at the gate is $B * \alpha$, where the show-up rate α is not deterministic, with a probability density function, $h(\alpha)$, and $0 \leq \alpha \leq 1$.

3. The survivals are stochastically independent of one another.

[**Rule 1:**] We suppose α (show-up rate) is not deterministic here, with a probability density function $h(\alpha)$ and $0 \leq \alpha \leq 1$. The optimal overbooking level $B = (1+k) * C$, where k can be obtained by:

$$\frac{R^o}{R^o + R^{sp}} = \frac{\int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha}{\int_0^1 h(\alpha)\alpha d\alpha}$$

[**Proof:**]

In the simplest form, assuming all values are deterministic, the overbooking level may be computed as given below. Assuming reservations book to the overbooking level B, the number of passengers that will show up at the gate is $B \cdot \alpha$. Ideally, one would like to set B such that $B \cdot \alpha = C$. When $0 \leq \alpha < \frac{C}{B}$, spoilage occurs and the amount of spoilage is equal to $C - B \cdot \alpha$; the corresponding spoilage[1] cost is $R^{sp} \cdot (C - B \cdot \alpha)$. When $\alpha > \frac{C}{B}$, oversale occurs and the amount of oversold seats is equal to $B \cdot \alpha - C$; the corresponding oversale cost is equal to $R^o \cdot (B \cdot \alpha - C)$. Now, we relax the assumption that the show up rate is deterministic.

Hence, the expected spoilage cost is:

$$\begin{aligned} E[SC] &= R^{sp} \int_0^1 h(\alpha)[C - (1+k)C\alpha]^+ d\alpha \\ &= R^{sp} \int_0^{\frac{1}{1+k}} h(\alpha)[C - (1+k)C\alpha] d\alpha \end{aligned}$$

[1]Spoilage cost per passenger, which is the revenue lost by not being able to fill the capacity due to show up falling short of capacity.

The expected oversale cost is:

$$\begin{aligned} E[OC] &= R^o \int_0^1 h(\alpha)[(1+k)C\alpha - C]^+ d\alpha \\ &= R^o \int_{\frac{1}{1+k}}^1 h(\alpha)[(1+k)C\alpha - C] d\alpha \end{aligned}$$

So, the expected total cost is:

$$\begin{aligned} E[TC] &= E[SC] + E[OC] \\ &= R^{sp} \int_0^{\frac{1}{1+k}} h(\alpha)[C - (1+k)C\alpha] d\alpha \\ &\quad + R^o \int_{\frac{1}{1+k}}^1 h(\alpha)[(1+k)C\alpha - C] d\alpha \end{aligned} \tag{2.1}$$

Differentiating with respect to k, we obtain:

$$\begin{aligned} \frac{dE[TC]}{dk} &= \frac{d}{dk}[R^{sp} \int_0^{\frac{1}{1+k}} h(\alpha) C d\alpha - R^{sp} \int_0^{\frac{1}{1+k}} h(\alpha)(1+k)C\alpha d\alpha] \\ &\quad + \frac{d}{dk}[R^o \int_{\frac{1}{1+k}}^1 h(\alpha)(1+k)C\alpha d\alpha - R^o \int_{\frac{1}{1+k}}^1 C h(\alpha) d\alpha] \\ &= -\frac{1}{(1+k)^2} R^{sp} h(\frac{1}{1+k})C - R^{sp} \int_0^{\frac{1}{1+k}} C\alpha h(\alpha) d\alpha \\ &\quad + R^{sp}(1+k)\frac{1}{(1+k)^2} C(\frac{1}{1+k}) h(\frac{1}{1+k}) \\ &\quad + R^o \int_{\frac{1}{1+k}}^1 C h(\alpha)\alpha d\alpha + R^o \frac{1}{(1+k)^2}(1+k) C h(\frac{1}{1+k})(\frac{1}{1+k}) \\ &\quad + R^o C(-\frac{1}{(1+k)^2}) h(\frac{1}{1+k}) \\ &= R^o \int_{\frac{1}{1+k}}^1 C h(\alpha)\alpha d\alpha - R^{sp} \int_0^{\frac{1}{1+k}} C\alpha h(\alpha) d\alpha \\ \frac{d^2E[TC]}{dk^2} &= -R^o(-\frac{1}{(1+k)^2}) C h(\frac{1}{1+k})(\frac{1}{1+k}) - R^{sp}(-\frac{1}{(1+k)^2}) C(\frac{1}{1+k}) h(\frac{1}{1+k}) \\ &= \frac{1}{(1+k)^3} h(\frac{1}{1+k}) C(R^o + R^{sp}) \geq 0 \end{aligned}$$

So, $E[TC]$ is a convex function and we can obtain the global minimum value.

Let $\frac{dE[TC]}{dk} = 0$ and $\mu_\alpha = \int_0^1 h(\alpha)\alpha d\alpha$, we get:

$$\begin{aligned}
\frac{dE[TC]}{dk} &= R^o C(\mu_\alpha - \int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha) - R^{sp} C \int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha = 0 \\
R^o C \mu_\alpha &= (R^o C + R^{sp} C) \int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha \\
\frac{R^o}{R^o + R^{sp}} &= \frac{\int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha}{\mu_\alpha} = \frac{\int_0^{\frac{1}{1+k}} h(\alpha)\alpha d\alpha}{\int_0^1 h(\alpha)\alpha d\alpha} \qquad (2.2)
\end{aligned}$$

The k satisfying Eq.(2.2) is the overbooking rate to get the minimum expected cost for the airline, hence, optimal overbooking level can be obtained by $B = (1+k)*C$. □

Here the solution to the formulation depends upon the underlying probability distribution. Almost all the airlines keep a record of past-departure data for all flights. These data can be used to study the form and parameters of the underlying distribution, which may be uniform, normal or beta.

[**An Example**]

Here is an example to illustrate the probabilistic overbooking model formulations presented in this section. We assume the underlying distribution is uniform.

We assume that show-up rate α is distributed uniformly over $[0.8, 1]$,
then $h(\alpha) = \frac{1}{1-0.8} = 5, \qquad (0.8 \leq \alpha \leq 1)$
so $\mu_\alpha = \int_{0.8}^1 \alpha h(\alpha) d\alpha = \int_{0.8}^1 5\alpha d\alpha = 0.9$.
We suppose that: $R^o = R^{sp} = 2f$ per seat of the capacity,
From Eq.(2.2), we have

$$\begin{aligned}
\frac{1}{2} &= \frac{\int_{0.8}^{\frac{1}{1+k}} 5\alpha d\alpha}{0.9} \\
\int_{0.8}^{\frac{1}{1+k}} 5\alpha d\alpha &= 0.45 \\
(\frac{1}{1+k})^2 &= 0.82 \\
(1+k) &\doteq 1.1043.
\end{aligned}$$

The overbooking level $B = (1+k) \cdot C = 1.1043C$.

[*Remark*] The solution to the formulation Eq.(2.2) depends upon the underlying probability distributions. In real life, the case is not as simple as Eq.(2.2), since the formulation will be more complex for the decision support system. We assume the amount of passengers that will show up at the gate is $B * \alpha$, where the show up rate α is a random variable. We use the expected number instead of the distribution, to get one decision rule. This rule is attractive for its simpleness. However, it may not be optimal by using the mean number. Together with the assumption that α is random, one can try to extend the Rule One to use the compound distribution to establish the optimal decision rule. Rule Two and Three as follows are treated as classical models for Static Airline Overbooking Problem. Instead, in Rule Two and Three, we assume that the show up rate α is constant and establish the optimal decision rule using the different distributions.

II. Rule Two

We assume that:

1. The number of the survivals is the *binomial process* with constant α,

   i.e. $S(B) \backsim Bin(B, \alpha)$.

2. The survivals are stochastically independent of one another.

3. Group bookings are not allowed here or treated as individual bookings.

4. Cancellations that occur after bookings which have been stopped are treated as no-shows.

5. Customers can be rejected boarding and get back $(f + R)$, where R is the refund to customer.

[**Rule 2:**] The decision rule with binomial survivals would suggest that the $(B + 1)th$ booking be accepted as long as:

$$\sum_{k=0}^{C-1} \binom{B}{k} \alpha^k (1-\alpha)^{B-k} > \frac{R}{R+f}. \tag{2.3}$$

[**Proof:**]

Let $U(B)$ be the total revenue to be maximized if we curtail bookings at B:

$$U(B) = f \cdot S(B) - (R+f) \cdot [S(B) - C]^{+}$$

$E[U(B)]$ is the total expected revenue over the random variable S. We assume that the decision maker's goal is to find the value B^* that maximize $E[U(B)]$. If $E[U(B)]$ is concave function of B, we will book $(B+1)th$ customer as long as:

$$E[U(B+1)] - E[U(B)] > 0$$

Thus the optimal overbooking level B^* is the largest $(B+1)$ such that $E[U(B+1)] - E[U(B)] > 0$.

Note that $E[S(B) - C]^{+} = 0$ for $B \leq C$, $E[U(B)]$ will be a non-decreasing function of $B \leq C$. We now prove that $E[U(B)]$ is concave function for $B \geq C$.

$$\begin{aligned}
& E[U(B+1)] - E[U(B)] \\
= \; & f \cdot (E[S(B+1)] - E[S(B)]) - (R+f) \cdot (E[S(B+1) - C]^{+} - E[S(B) - C]^{+}) \\
= \; & f \cdot [\alpha \cdot (B+1) - \alpha \cdot B] - (R+f) \cdot [\sum_{i=C+1}^{B+1} \binom{B+1}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B+1-i} \cdot (i-C) \\
& - \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B-i} \cdot (i-C)]
\end{aligned}$$

We know that: $\binom{B+1}{i} = \binom{B}{i} + \binom{B}{i-1}$.

$$\begin{aligned}
= \; & f \cdot \alpha - (R+f) \cdot [\alpha^{B+1}(B+1-C) + \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B+1-i} \cdot (i-C) \\
+ \; & \sum_{i=C+1}^{B} \binom{B}{i-1} \cdot \alpha^{i} \cdot (1-\alpha)^{B+1-i} \cdot (i-C) - \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B-i} \cdot (i-C)] \\
= \; & f \cdot \alpha - (R+f) \cdot [\alpha^{B+1}(B+1-C) + \sum_{i=C+1}^{B} \binom{B}{i-1} \cdot \alpha^{i} \cdot (1-\alpha)^{B+1-i} \cdot (i-C) \\
& -\alpha \cdot \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B-i} \cdot (i-C)]
\end{aligned}$$

Let $k = i - 1$.

$$
\begin{aligned}
&= f \cdot \alpha - (R+f) \cdot [\alpha^{B+1}(B+1-C) + \sum_{k=C}^{B-1} \binom{B}{k} \cdot \alpha^{k+1} \cdot (1-\alpha)^{B-k} \cdot (k+1-C) \\
&\quad -\alpha \cdot \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i} \cdot (1-\alpha)^{B-i} \cdot (i-C)] \\
&= f \cdot \alpha - (R+f) \cdot [\sum_{k=C}^{B} \binom{B}{k} \cdot \alpha^{k+1} \cdot (1-\alpha)^{B-k} \cdot (k+1-C) \\
&\quad - \sum_{i=C+1}^{B} \binom{B}{i} \cdot \alpha^{i+1} \cdot (1-\alpha)^{B-i} \cdot (i-C)] \\
&= f \cdot \alpha - (R+f) \cdot \sum_{k=C}^{B} \binom{B}{k} \alpha^{k+1} (1-\alpha)^{B-k}
\end{aligned}
$$

$$
\begin{aligned}
&\implies (E[U(B+1)] - E[U(B)]) - (E[U(B)] - E[U(B-1)]) \\
&= -(R+f) \cdot [\sum_{k=C}^{B} \binom{B}{k} \alpha^{k+1}(1-\alpha)^{B-k} - \sum_{k=C}^{B-1} \binom{B-1}{k} \alpha^{k+1}(1-\alpha)^{B-1-k}] \\
&= -(R+f) \cdot [\alpha^{B+1} + \sum_{k=C}^{B-1} \binom{B-1}{k} \alpha^{k+1}(1-\alpha)^{B-k} \\
&\quad + \sum_{k=C}^{B-1} \binom{B-1}{k-1} \alpha^{k+1}(1-\alpha)^{B-k} - \sum_{k=C}^{B-1} \binom{B-1}{k} \alpha^{k+1}(1-\alpha)^{B-1-k}] \\
&= -(R+f) \cdot [\alpha^{B+1} + \sum_{k=C}^{B-1} \binom{B-1}{k-1} \alpha^{k+1}(1-\alpha)^{B-k} \\
&\quad - \sum_{k=C}^{B-1} \binom{B-1}{k} \alpha^{k+2}(1-\alpha)^{B-1-k}]
\end{aligned}
$$

Let $z = k + 1$.

$$
\begin{aligned}
&(E[U(B+1)] - E[U(B)]) - (E[U(B)] - E[U(B-1)]) \\
&= -(R+f) \cdot [\alpha^{B+1} - \sum_{z=C+1}^{B} \binom{B-1}{z-1} \alpha^{z+1}(1-\alpha)^{B-z} + \sum_{k=C}^{B-1} \binom{B-1}{k-1} \alpha^{k+1}(1-\alpha)^{B-k}] \\
&= -(R+f) \cdot [\sum_{k=C}^{B} \binom{B-1}{k-1} \alpha^{k+1}(1-\alpha)^{B-k} - \sum_{z=C+1}^{B} \binom{B-1}{z-1} \alpha^{z+1}(1-\alpha)^{B-z}] \\
&= -(R+f) \cdot [\binom{B-1}{C-1} \alpha^{C+1}(1-\alpha)^{B-C}] \\
&< 0.
\end{aligned}
$$

So $E[U(B)]$ is concave function of $B \geq C$, because $(E[U(B+1)] - E[U(B)])$ is strictly decreasing in $B \geq C$. Hence, we will book $(B+1)th$ customer as long as:

$$\begin{aligned} & E[U(B+1)] - E[U(B)] > 0 \\ \Leftrightarrow \quad & f \cdot \alpha - (R+f) \cdot \sum_{k=C}^{B} \binom{B}{k} \alpha^{k+1} (1-\alpha)^{B-k} > 0 \\ \Leftrightarrow \quad & f > (R+f) \cdot [1 - \sum_{k=0}^{C-1} \binom{B}{k} \alpha^{k} (1-\alpha)^{B-k}] \\ \Leftrightarrow \quad & \sum_{k=0}^{C-1} \binom{B}{k} \alpha^{k} (1-\alpha)^{B-k} > \frac{R}{R+f} \end{aligned}$$

We proved that the $(B+1)th$ booking can be accepted as long as:

$$\sum_{k=0}^{C-1} \binom{B}{k} \alpha^{k} (1-\alpha)^{B-k} > \frac{R}{R+f}. \qquad \square$$

[*Remark 1:*] The survival probability which is a constant here must be estimated subjectively or from past data.

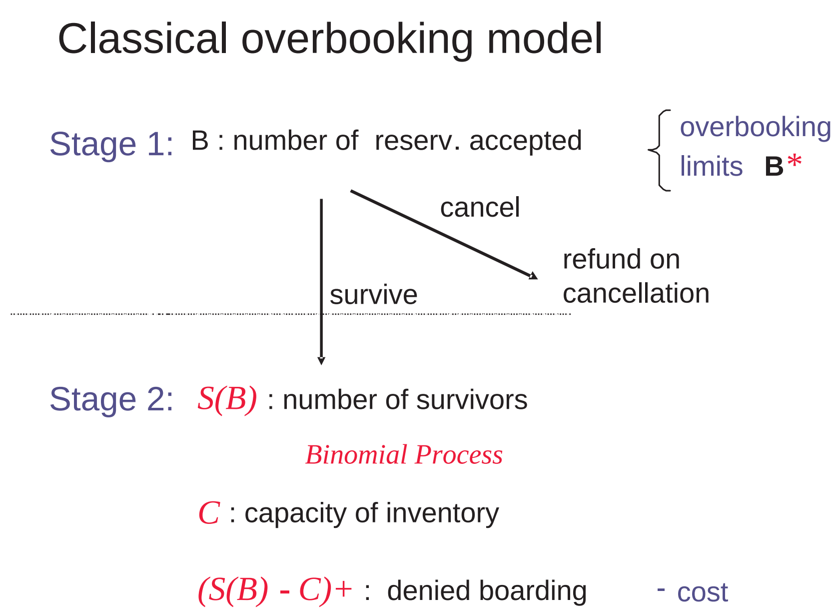
[*Remark 2:*] Eq.(2.3) can be stated: curtail bookings when the probability of spoilage has decreased to the ratio $\frac{R}{f+R}$. It can be treated as a primary decision rule and can be extended to increasingly more practical assumptions about the survival process.

[*Remark 3:*] In Rule Two, we assume that the survival process is binomial. If passengers book in small groups or passengers cancel in small groups, then cancellation patterns are not independent any more, and the distribution is not binomial. So group bookings are not allowed here or treated as individual bookings.

[*Remark 4:*] Actually, this model (See the below chart for illustration) is the classical model in the Static Airline Overbooking Problem and can be used as a basic step for us to go further in this problem.

III. Rule Three

It is convenient and practical to treat the overbooking level and the survival quantities as continuous variables. Empirical studies have shown that the normal



probability distribution gives a good continuous approximation to airline demand distributions. Applying the normal approximation to the binomial distribution, we get the following decision Rule 3. Under the assumption of binomial survivals with constant α, the implementation of the optimal decision rule prescribes a fixed limit on the number of reservations to be booked (based on f, R^o, C, α that can be put in prior to accepting reservations) to maximize the expected net revenue which is equal to the total gross revenue minus expected oversale cost.

[**Rule 3:*Continuous normal decision rule***]

Book reservations up to B which is determined by:

$$\Phi = \frac{R^o}{f + R^o} + \frac{\phi \cdot \sqrt{1-\alpha}}{2\sqrt{B\alpha}}$$

where ϕ is the unit normal probability density function and Φ is the left-tail unit normal cumulative distribution function, both evaluated at $\frac{(C-\alpha B)}{[\alpha B(1-\alpha)]^{\frac{1}{2}}}$.

[**Proof:**]

Assume that the distribution for survivals S given B books is normal, with mean

$\mu = \alpha B$ and variance $\sigma^2 = \alpha(1-\alpha)B$. The value to be maximized is:

$$U(B,S) = \begin{cases} fS, & \text{if} \quad 0 \le S \le C \\ fC - R^o(S-C), & \text{if} \quad S > C \end{cases}$$

The expected value of $U(B,S)$ over the random variable S is then

$$E[U(B,S)] = \int_0^C fS \cdot \theta(S|B)dS + \int_C^\infty [fC - R^o(S-C)]\theta(S|B)dS$$

where $\theta(S|B)$ is the normal probability density (given B): $\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(S-\mu)^2}{2\sigma^2}}$.
Let $\Theta(C) = \int_{-\infty}^C \theta(S|B)dS$ which is the left-tail cumulative distribution function. So, we have:

$$\begin{aligned} \Theta(C) &= \int_{-\infty}^C \theta(S|B)dS \\ &= \int_{-\infty}^C \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(S-\mu)^2}{2\sigma^2}} dS \\ &= \int_{-\infty}^{\frac{C-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx \\ &= \Phi(\frac{C-\mu}{\sigma}) \\ &= \Phi(z) \end{aligned}$$

where, $z = \frac{C-\mu}{\sigma}$ and $\Phi(z) = \int_{-\infty}^z \phi(x)dx$, where $\phi(x)$ is the unit normal probability density function.

Hence, the expression may be retreated as:

$$\begin{aligned} E[U(B,S)] &= \int_{-\infty}^{+\infty} fS \cdot \theta(S|B)dS - \int_C^\infty fS \cdot \theta(S|B)dS - \int_{-\infty}^0 fS \cdot \theta(S|B)dS \\ &\quad + \int_C^\infty [fC - R^o(S-C)]\theta(S|B)dS \\ &= f\mu - (f+R^o)\int_C^\infty S\theta(S|B)dS + (f+R^o)\int_C^\infty C\theta(S|B)dS \\ &= f\mu - (f+R^o)\int_C^\infty S\theta(S|B)dS + (f+R^o)C[1-\Theta(C)] \end{aligned}$$

$$\begin{aligned}
\int_{C}^{\infty} S\theta(S|B)dS &= \int_{-\infty}^{+\infty} S\theta(S|B)dS - \int_{-\infty}^{C} S\theta(S|B)dS \\
&= \mu - \int_{-\infty}^{C} S \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(S-\mu)^2}{2\sigma^2}} dS \\
&= \mu + \int_{0}^{e^{-\frac{(C-\mu)^2}{2\sigma^2}}} \sigma^2 \frac{1}{\sqrt{2\pi}\sigma} d(e^{-\frac{(S-\mu)^2}{2\sigma^2}}) \\
&\qquad -\mu \int_{-\infty}^{C} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(S-\mu)^2}{2\sigma^2}} dS \\
&= \mu + \sigma^2 \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(C-\mu)^2}{2\sigma^2}} - \mu \cdot \Theta[C] \\
&= \mu[1 - \Theta(C)] + \sigma^2 \cdot \theta(C) \\
&= \mu[1 - \Phi(z)] + \sigma \cdot \phi(z) \qquad (2.4)
\end{aligned}$$

where, $\mu = \alpha B$, $\sigma^2 = \alpha B(1-\alpha)$.

So, we get

$$\begin{aligned}
E[U(B,S)] &= f\mu - (f+R^o)[\mu(1-\Phi) + \sigma \cdot \phi] + (f+R^o)C(1-\Phi) \\
&= f\mu - (f+R^o)[\sigma \cdot \phi - (C-\mu)(1-\Phi)] \qquad (2.5)
\end{aligned}$$

$$\frac{\partial E[U(B,S)]}{\partial B} = f\alpha - (f+R^o)[\phi\frac{\sigma}{2B} + \sigma\frac{\partial \phi}{\partial B} + \alpha(1-\Phi) + (C-\mu)(\frac{\partial \Phi}{\partial B})] \qquad (2.6)$$

Set $\frac{\partial E[U(B,S)]}{\partial B} = 0$.

$$\alpha(1-\Phi) = \frac{f\alpha}{f+R^o} - [\phi\frac{\sigma}{2B} + \sigma\frac{\partial \phi}{\partial B} + (C-\mu)(\frac{\partial \Phi}{\partial B})]$$

It is easy to get that $\sigma\frac{\partial \phi}{\partial B} + (C-\mu)(\frac{\partial \Phi}{\partial B}) = 0$.

$$\begin{aligned}
\Longrightarrow \alpha(1-\Phi) &= \frac{f\alpha}{f+R^o} - \phi\frac{\sigma}{2B} \\
\Longrightarrow \Phi &= \frac{R^o}{f+R^o} + \phi\frac{\sigma}{2\mu} \\
\Longrightarrow \Phi &= \frac{R^o}{f+R^o} + \frac{\phi \cdot \sqrt{1-\alpha}}{2\sqrt{B\alpha}}
\end{aligned}$$

The B satisfying the above equation is the optimal overbooking level. □

[**Example**:] We consider a hypothetical airline, Bell Air. On one of its flights, previous records show that about 15% of people who had tickets for the flight did not take the flight. There are $C = 120$ seats on the plane. Assuming that any particular individual on the flight has a probability of 0.15 of not showing up, and that whether each individual shows up is independent from whether any other individual shows up. This is a binomial situation with B trials and $\alpha = 0.85$. Since B will be at least 120, it is appropriate to use normal approximation to binomial distribution since $\alpha B \geq 120(0.85) = 112 > 5$ and $B(1-\alpha) \geq 120(0.15) = 18 > 5$. We use a normal approximation, $\mu = \alpha B = B(0.85)$ and $\sigma = \sqrt{B\alpha(1-\alpha)} = \sqrt{B(0.15)(0.85)}$. Suppose $f = 450$ and $R^o = 200$, so, we have $\frac{R^o}{f+R^o} = 0.31$. From the normal distribution table, the corresponding $z = -0.50$. So by Rule Three, we get 143 as the optimal booking level.

[**Remark 1:**] The model for Rule Three is same in Bodily and Pfeifer (92)[10]. They got the first result Eq. (2.4) without proof, saying that it could be obtained by using the expansion for a right-tail normal integral from Raiffa and Schlaifer (1961) [41]. We apply the same model from [10], analyze the problem carefully and complete the proof here.

[**Remark 2:**] A useful extension of the Rule Three is to consider survival probabilities that depend on the time the reservation is made, which is more practical, for reservations made only a few periods ahead may produce survivals with more likelihood than reservations made many periods ahead. These reservations with time-varying probabilities can be treated as different fare classes. In some sense, this problem can be treated as a multi-fare-class, single-leg problem.

[**Remark 3:**] Rule One, Two, Three are all simplest models with one fare class on Single Leg. We gave out different decision rules to determine the overbooking level. We ignored the demand factor which is reasonable only for one fare class. In the case with more than two fare classes, it is more complicated and we should definitely consider about the demand distribution.

[**Claim:**] The demand factor will not affect our decision rule for Static, Single-leg, Single fare class problem.

[**Proof:**]

Define $U(B,D,S)$ be the total revenue to be maximized if we curtail bookings at B, as defined previously, but it includes the demand factor.

$$U(B,D,S) = f \cdot S(min\{B,D\}) - (R+f) \cdot [S(min\{B,D\}) - C]^+$$

where D is the demand and $S(min\{B,D\})$ is the number of the survivals, where S can be any distribution. $E[U(B,D,S)]$ is the total expected revenue over the random variable S. Let x as the demand random variable, with a probability density function, $g(x)$.

$$\begin{aligned}
E[U(B,D,S)] &= \int_0^\infty f \cdot E[S(min\{B,x\})] \cdot g(x)dx \\
&\quad - \int_0^\infty (R+f) \cdot E[S(min\{B,x\}) - C]^+ \cdot g(x)dx \\
&= \int_0^B f \cdot E[S(x)] \cdot g(x)dx + \int_B^\infty f \cdot E[S(B)] \cdot g(x)dx \\
&\quad - \int_0^B (R+f) \cdot E[S(x) - C]^+ \cdot g(x)dx \\
&\quad - \int_B^\infty (R+f) \cdot E[S(B) - C]^+ \cdot g(x)dx
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E[U(B,D,S)]}{\partial B} &= f \cdot E[S(B)] \cdot g(B) + \int_B^\infty f \cdot \frac{\partial E[S(B)]}{\partial B} \cdot g(x)dx - f \cdot E[S(B)] \cdot g(B) \\
&\quad -(R+f) \cdot E[S(B) - C]^+ \cdot g(B) + (R+f) \cdot E[S(B) - C]^+ \cdot g(B) \\
&\quad - \int_B^\infty (R+f) \cdot \frac{\partial E[S(B) - C]^+}{\partial B} \cdot g(x)dx \\
&= f \cdot \frac{\partial E[S(x)]}{\partial B} \cdot Pr\{D \geq B\} - (R+f) \cdot \frac{\partial E[S(B) - C]^+}{\partial B} \cdot Pr\{D \geq B\}
\end{aligned}$$

Let $\frac{\partial E[U(B,D,S)]}{\partial B} = 0$. We have:

$$f \cdot \frac{\partial E[S(x)]}{\partial B} = (R+f) \cdot \frac{\partial E[S(B) - C]^+}{\partial B}$$

We proved that the demand factor will not affect our decision rule.

2.3 Multi-fare-class Model

Nested Reservation System:

Let's consider a Single-leg overbooking control problem in which bookings are accepted into two fare classes in the *nested* reservation system [Figure 2.1]. Nesting is preferable in airline seat management. In the partitioned structure [Figure 2.2], the booking requests for the higher fare class will not be accepted even the limit for the lower fare class will never be exceeded. So the airlines without nested reservation systems are denying themselves the flexibility of accommodating unexpectedly high demand levels in high-fare classes and, in turn, are losing potential revenues. Williamson (1992) [62] presented that the expected revenue from a nested structure is equal to or greater than that of a partitioned structure.

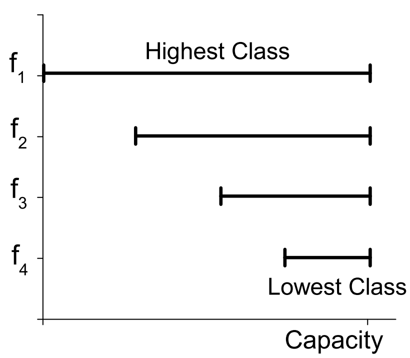


Figure 2.1: Nested Structure.

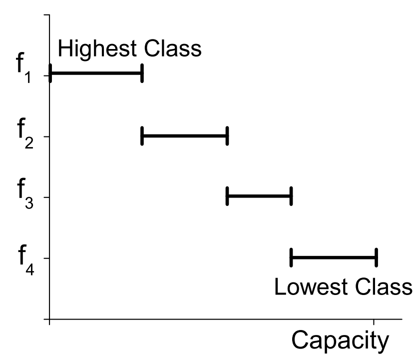


Figure 2.2: Partitioned Structure.

Littlewood (1972) [39] was the first to propose a solution to a booking control problem for a Single-leg flight with two fare classes in Nested Reservation System. The idea of his scheme is to equate the marginal revenues in each of the two fare classes, closing down the lower fare class when the certain revenue from selling another lower fare seat is exceeded by the expected revenue of selling the same seat to a higher fare.

[***Note:***] The assumption 1 of Rule One in previous section: Sell the tickets on a first-come, first-served basis, is not practical for a multi-fare-class problem. If we still use this assumption, let's say in a two-fare-problem (Leisure travellers and Business travellers), its capacity is likely to fill up early with leisure travellers, who are eager to fix their holiday trip. Later bookers, typically business travellers willing to pay a higher fare, will then find that there are no seats left, and these sales will be lost. So here we will ideally assume that lower fare comes before the higher fare, and try to find limit on the lower fare class to protect some seats for the later higher fare class.

Under this assumption, Littlewood suggested that total revenue on flight leg would be maximized if additional low fare bookings were accepted based on the condition that the certain revenue obtained from each incremental low fare passenger exceeded the expected marginal revenue of saving the seat for a potential high fare passenger. That is: $f_2 \geq f_1 \cdot \int_{\Pi_2^1}^{\infty} p_1(D_1)dD_1$. Hence, this rule can determine when to stop accepting bookings from Class 2 and how many seats should be protected for Class 1, so as to maximize the total expected revenue without considering overbooking, where,

f_i	The fare for the class i and $f_1 > f_2 > 0$.
D_1	The total demand for seats in the class 1.
D_2	The total demand for seats in the class 2.
$p_1(D_1)$	The probability distribution for the total number of requests for reservations, D_1 received by the airline for seats in class 1.
Π_2^1	The number of seats protected for Class 1 from Class 2 bookings (protection level without considering overbooking).
B_1	The maximum number of seats available for the fare class 1 (Booking limit for the higher fare class).
B_2	The maximum number of seats available for the fare class 2 (Booking limit for the lower fare class).

By determining a protection level for the high fare class, they also set the booking limits for both classes, i.e. the maximum number of seats available for higher fare class is: C and for lower fare class is: $C - \Pi_2^1$. The reservation system is nested. If class 2 booking requests never reach $C - \Pi_2^1$, the unsold seats will be available for unexpectedly higher class 1 demand.

Belobaba (1987) [5] extended Littlewood's rule to multiple fare classes and introduced the term expected marginal seat revenue (EMSR) for the general approach. Assume that: we have m fare classes and $f_1 > f_2 > \cdots > f_m$.

$$\begin{aligned}
\overline{\mathbf{P}_i(\Pi_j^i)} &= \sum_{D_i=\Pi_j^i}^{\infty} p_i(D_i) \\
EMSR_i(\Pi_j^i) &= f_i \cdot \overline{\mathbf{P}_i(\Pi_j^i)} = f_j \quad (i < j \quad j = 2, 3, \cdots, m)
\end{aligned}$$

where Π_j^i is the protection level for class i from class j without considering overbooking.

His method produced nested protection levels, i.e. set the transparent booking limits for the higher fare class.

$$B_j = \max\{0, C - \sum_{i<j} \Pi_j^i\}$$

Although the EMSR method is not optimal for more than two fare classes, it does provide good booking limits in practice.

However, Littlewood didn't consider cancellations, no-shows and overbooking in his model. It can be of interest to revise the Littlewood's rule incorporating overbooking. The demand inputs required are still estimates of the densities of requests for each fare class. The difference here is that each accepted booking request in a fare class cannot be treated as if the revenue associated with that class will always be realized.

We assume that each passenger in class 1 has an independent probability α_1 of showing up, with $0 < \alpha_1 < 1$ and each passenger in class 2 has an independent probability α_2 of showing up, with $0 < \alpha_2 < 1$.

We assume that lower valued fare classes book before higher valued fare classes. Suppose a request comes to book one more seat in Class 2, we have to make a decision whether the request should be accepted by closing down one seat in Class 1 or reject it. We use π_2^1 as the number of seats left for Class 1 from Class 2 bookings, which takes no-shows and overbooking into considerations. If we accept one more booking request for class 2, we will have $\pi_2^1 - 1$ tickets to class 1. Otherwise, we will keep all the left tickets (π_2^1) to class 1.

If we accept this booking request and no more request will be accepted in Class 2, airline will get:

$$U(1) = f_1 \cdot \alpha_1 \cdot \min\{\pi_2^1 - 1, D_1\} + f_2 \cdot \alpha_2$$

If we reject it:

$$U(0) = f_1 \cdot \alpha_1 \cdot \min\{\pi_2^1, D_1\}$$

The expected values are:

$$\begin{aligned} E[U(1)] &= f_1 \cdot \alpha_1 \cdot [\sum_{D_1=0}^{\pi_2^1-1} D_1 p_1(D_1) + \sum_{D_1=\pi_2^1}^{\infty} (\pi_2^1 - 1) p_1(D_1)] + f_2 \cdot \alpha_2 \\ E[U(0)] &= f_1 \cdot \alpha_1 \cdot [\sum_{D_1=0}^{\pi_2^1-1} D_1 p_1(D_1) + \sum_{D_1=\pi_2^1}^{\infty} \pi_2^1 p_1(D_1)] \end{aligned}$$

The optimal solution must be such that it gives a higher expected value, hence, it is optimal to accept the booking only if $E[U(1)] \geq E[U(0)]$.

$$\begin{aligned} & f_1 \cdot \alpha_1 \cdot [\sum_{D_1=0}^{\pi_2^1-1} D_1 p_1(D_1) + \sum_{D_1=\pi_2^1}^{\infty} (\pi_2^1 - 1) p_1(D_1)] + f_2 \cdot \alpha_2 \\ \geq \quad & f_1 \cdot \alpha_1 \cdot [\sum_{D_1=0}^{\pi_2^1-1} D_1 p_1(D_1) + \sum_{D_1=\pi_2^1}^{\infty} \pi_2^1 p_1(D_1)] \end{aligned}$$

That is, this Class 2 request should be accepted only if :

$$f_2 \cdot \alpha_2 \geq f_1 \cdot \alpha_1 \cdot \sum_{D_1=\pi_2^1}^{\infty} p_1(D_1).$$

From the above analysis, we will get the revised Littlewood's rule as follows:

Revised Littlewood's rule:

We accept an additional low fare bookings based on the condition that the certain *revenue obtained from each incremental low fare passenger exceeded the expected marginal revenue of saving the seat for a potential high fare passenger. That is:*

$$f_2 \cdot \alpha_2 \geq f_1 \cdot \alpha_1 \cdot \sum_{D_1=\pi_2^1}^{\infty} p_1(D_1). \tag{2.7}$$

The smallest value of π_2^1 that satisfies Eq.(2.7) is the number of seats protected for the higher fare class, which we call the optimal protection level.

[**Remark 1:**] We get the Eq.(2.7), then we can get the optimal protection level for class 1 from class 2. However, by determining a protection level for the high fare class, we still cannot set the booking limits for both classes. [2]

Define:

B^*	The optimal overbooking level (Booking limit for the higher fare class).

[**Remark 2:**] The function depends on the probability distribution of D_1, which is a discrete random variable. Most of the time, it is quite painful process to find a representation of this probability distribution when the range of all possible values for demand is large. Hence, continuous random variable can be considered to approximate the discrete random variable D_1.

Similar to Belobaba's way, we also extend the revised Littlewood's rule mentioned above to multiple fare classes on Single-leg. We call it **revised EMSR (rEMSR)**.

We accept an additional lower fare bookings based on the condition that the certain revenue obtained from each incremental lower fare passenger exceeded the expected marginal revenue of saving the seat for a potential higher fare passenger.

[2] We need to get an overbooking level, or sometimes called 'Pseudo-capacity' (which is equal to Capacity + Overbooking Pad) since the overbooking level B^* is treated as if it were the seating capacity of the airplane. As the Pseudo-capacity approach decomposes the cancellation of the total overbooking limit from calculation of the booking limit for each fare class, it is probably the most widely used approach to overbooking in real world revenue management implementations. i.e. If we can find: B^*= Capacity + Overbooking Pad, then we can set $B_1 = B^*$ and $B_2 = B^* - \pi_2^1$.

That is:

$$f_j \cdot \alpha_j \geq f_i \cdot \alpha_i \cdot \overline{\mathbf{P}_i(\pi_j^i)} = rEMSR_i(\pi_j^i), \qquad i < j, \quad j = 2, 3, \cdots, m. \qquad (2.8)$$

where the smallest value π_j^i satisfying the above condition is the protection level for Class i from Class j $(i < j)$.

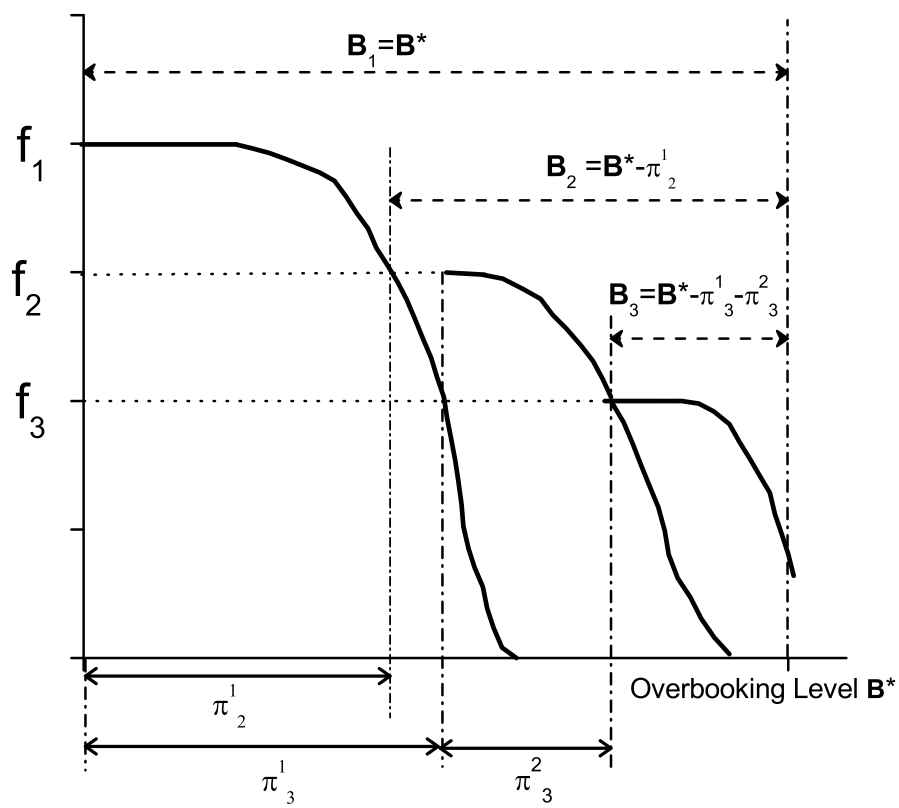


Figure 2.3: The rEMSR solution for the nested 3-class example.

However, the derivation of overbooking limits, B_i, from the revised EMSR decision rule is still complicated. The simplest case is when show up rates α_i across fare classes are following the equal, so that the revised EMSR formulation is reduced to the original EMSR. And we assume that all the oversold tickets from the lowest fare class f_m with penalty to the airline R. We set D_i as the totally demand for ith fare class and $D = D_1 + \cdots + D_m$. If we can find the optimal overbooking level $B^* = B_1$, which is available for all the fare class, we can obtain all the overbooking limits simply by: $B_j = B_1 - \sum_{i<j} \pi_j^i, \quad (j = 2, 3, \cdots, m)$. The net result is that each fare class may be overbooked by the same percentage and B^* will be the same regardless of the fare class mix actually booked for any particular flight.

We can find the optimal overbooking level using one approach by a weighted average fare as follows:

$$\overline{f} = \frac{\sum_1^m f_i \cdot E[D_i]}{\sum_1^m E[D_i]}.$$

The total revenue to be maximized for the airline can be defined as:

$$\overline{f} \cdot S(min\{B_1, D\}) - R \cdot E[S(min\{B_1, D\}) - C]^+.$$

From the claim in Section 2.2, we have known that the demand factor will not affect the Single-leg, Single-fare-class problem. So we have the total revenue function as follows to be maximized:

$$\overline{f} \cdot S(B_1) - R \cdot E[S(B_1) - C]^+.$$

Hence, we can find an optimal $B_1 = B^*$ referring to Rule Two or Rule Three when the S follows the binomial or normal distribution. Subsequently, we can find all the B_i.

[**Conclusion:**]

We have explored the multiple fare classes, Single-leg problem in nested structure in this section. In a nested reservations system, fare classes can be structured as such that a request will always be accepted as long as seats are available in the respective or any of lower fare classes. Such system is binding in its booking limits on the lower fare, but the limits are 'transparent' from the higher fare classes. Williamson (1992) [62] shows that the nested reservation system can take equal or more expected revenue for the airline than the partitioned system. We have revised EMSR method to find the protection levels π_j^i for Class i from Class j, where $i < j$, referring to Eq.(2.8). Given these protection levels, we have discussed one example to find out the optimal booking limit for each fare class under some assumptions in this section. At present, further research is still needed to make the model easier to implement and more efficient.

[**Evaluation:**] The methods in this Chapter are all static. We determine booking control policies at the start of the booking period, i.e. set the optimal booking limits for all the fare classes to maximize the expected revenue for the airline. The results generated by these solutions are optimal under the sequential arrival assumption as long as no change in the probability distributions of demand is foreseen.

Actually, the real demand process will change, and the airline has the opportunity to change the overbooking limits as departure approaches. If we can get the information on the actual demand process, then we can reduce the uncertainty associated with the estimates of demand. Hence, we can use such Static method repetitively over the booking period based on the most recent demand and capacity information, which is the general way to approximate the dynamic overbooking problem. In the next Chapter, we will emphasize on the Dynamic Overbooking Problem on the Single-leg.

Chapter 3

Dynamic Overbooking Problem

3.1 Introduction

In Chapter Two, we try to find the optimal booking limits for different fare classes at the beginning of the booking period. These booking limits will not be changed until the flight departures. The booking control policy there is to accept the booking request within the booking limit. The solutions in Chapter Two are optimal under the sequential arrival assumption as long as no change in the probability distributions of demand is foreseen.

In fact, the overbooking problem is intrinsically dynamic — an airline has the opportunity to change the limits as departure time approaches. In this Chapter, the multiple fare classes are booked concomitantly without the assumption on the arrival patterns for various fare classes. We will consider the booking process as a discrete-time dynamic programming model(DP), where demand for each fare class is modelled by a non-homogeneous Poisson process. Using a Poisson process gives rise to the use of a Markov Decision model. The states of the such model are dependent on the time until the departure of the flight and on the reserved capacity. The stochastic process of cancellations will also be considered. To overcome revenue losses resulting from no-shows, airlines will rationally adopt the overbooking policies as well. The booking period is divided into a number of decision periods.

These decision periods are sufficiently small such that no more than one request (either a booking request or a cancellation) arrives within such a period. The state of the process changes every time a decision period elapses or the number of seats previously accepted changes.

For the dynamic inherency, the control policy in this chapter is: *focus on whether to accept or reject a particular booking request at its particular arrival time, by comparing the revenue generated by accepting this request and the opportunity cost of losing this seat taken up by the booking.* So the solution for this problem is more concerned with evaluating these opportunity costs and incorporating them in booking control policy such that the expected future revenue is maximized.

The dynamic control proposed by Janakiram, Stidham, and Shaykevich (1999) [54] will be discussed in this Chapter. However, the characterization of the optimal policy will be restructured for more clear interpretation. The Model 1, a simple extension from Lee and Hersh (1993) [38] that incorporates cancellation, no-shows and overbooking will be described. The concavity of the associated optimal value function, which is the key point to get the optimality of such control policy will be proved completely in this report. Model 2 considers the more general case which allows class-dependent cancellation and no-show probabilities and refund amounts, with refunds at the time of cancellation and no-show, resulting in multidimensional state variable. We will discuss the way in [54] how to reduce the problem to a one-dimensional MDP and avoiding the curse of dimensionality. We will represent the steps more clearly and complete some proofs which are not provided in [54]. Finally, we summarize some results from a small numerical example quoted from [54], highlighting the effects of allowing cancellations and suggesting that the full multidimensional model may not always be necessary. The discussion in this chapter are mainly based on the references: [38], [54], [37] and [26].

3.2 Model Description

We will describe some common natures and assumptions for Model 1 and Model 2 first. Consider the models on single flight leg with capacity C (e.g. a one-way flight from RDU to LAX). Assuming that each passenger belongs to one of m fare classes, with class 1 corresponding to the highest fare and class m to the lowest, each of the m fare classes may arrive throughout the reservations horizon. Each passenger requests at most one seat. Multiple seat requests, where an arriving customer attempts to book more than one reservation, are not permitted. At the moment a booking request arrives, the decision to accept or reject involves three factors: the number of seats previously accepted, the time remaining in the reservations horizon, and the fare class of the request.

The models 1 and 2 share the assumptions as follows:

(1) Single-leg flight with known capacity C.

(2) m fare classes (1-highest fare, m-lowest fare) and independent demand between the booking classes.

(3) Booking requests in each fare class are time-dependent processes.

(4) Passengers may cancel the reservations until the departure time.

(5) Passengers can be no-shows at the time of departure.

(6) Overbooking is allowed with a penalty-cost function.

(7) Denied requests are considered revenue lost, i.e. passengers whose requests denied will not upgrade or take another flight on the same airline.

x	The current number of reserved seats.
$\mathbf{x} = (x_1, \cdots, x_m)$	The reservation vector, where x_i denotes the number of seats currently reserved in class i.
$U_t(x)/\widehat{U}_t(\mathbf{x})$	The maximum expected net revenue given the reserved seats and with t remaining decision periods before the departure of the flight.

3.3 Model 1

1. Assumption and Description

Formulate the problem as a finite-horizon, discrete-time Markov decision process (MDP) in which the state variable is the total number of seats already accepted.

Besides the common assumptions in Section 3.2, we assume that:

• Cancellations and no-shows have class-independent rates but are still time-dependent process(one dimensional state variable).

• No refund for cancellations and no-shows.

• At each stage, only one of the following events occurs:

(1) a booking request of a customer in fare class i with the probability p_{it};

(2) a cancellation by a customer with the probability $q_t(x)$, which is a non-decreasing and concave function of x;

(3) a null event with the probability $p_{0t}(x)$;

i.e. $\sum_{i=1}^{m} p_{it} + q_t(x) + p_{0t}(x) = 1$ for all x and $t \geqslant 1$.

• At the time of departure, each customer holding a seat reservation is a no-show with probability β. Hence, $1-\beta$ is the probability of showing-up for the flight. As a result, we can see that $Y(x) \sim Bin(x, 1-\beta)$ which is a binomial distribution.

$y = Y(x)$	The number of people who show up for the flight.
$\pi(Y(x))$	Overbooking penalty, assuming it is non-negative, convex and non-decreasing in $Y(x) \geq 0$, with $\pi(Y(x)) = 0$ for $Y(x) \leq C$.

The objective is to maximize the expected total net revenue over the horizon from period T to period 0, the departure time of the flight. At each stage t, the following transitions possible from state x are shown in Figure (3.1).

Since the Model 1 starts with no seats booked at stage T and at most one seat request can be accepted at each stage, it follows that $x \leq T - t$, at each stage t.

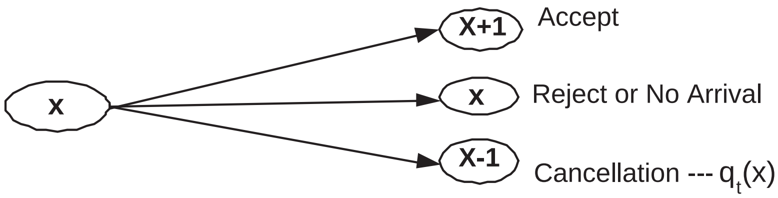


Figure 3.1: Transitions Possible.

So, the optimal value functions, U_t, are determined recursively by:

$$\begin{aligned}
U_t(x) = \quad & \sum_{i=1}^{m} p_{it} \max\{f_{it} + U_{t-1}(x+1), U_{t-1}(x)\} \\
& + q_t(x)U_{t-1}(x-1) + p_{0t}(x)U_{t-1}(x), \qquad 0 \le x \le T-t, t \ge 1 \\
U_0(x) = \quad & E[-\pi(Y(x))], \qquad 0 \le x \le T
\end{aligned} \tag{3.1}$$

[**Note 1:**] The requests for seats are independent of the number of seats already booked x, whereas the cancellation probabilities depend on x, which is the sum of the seats booked in all fare classes.

[**Note 2:**] The model here is discrete. Direct implementation of the obvious dynamic programming solution techniques leads to algorithms that are computationally intractable when applied to problems of practical size. We give out *One Step of Calculation* as follows. Particularly, we can assume that $q_t(x) = x \cdot q_t$, where q_t is the average cancellation rate at t. Passengers cancel their booked seats independently of one another.

At stage t, given the following data, we can calculate $U_t(x)$ given x.

Fare Classes	**Probability of Arrival**	**Probability of Cancellation**
$f_{1t} = 200$	$p_{1t} = 0.2$	$q_t = 0.02$
$f_{2t} = 150$	$p_{2t} = 0.4$	$q_t = 0.02$
$f_{3t} = 75$	$p_{3t} = 0.2$	$q_t = 0.02$

$$U_{t-1}(4) = 1600$$
$$U_{t-1}(5) = 1500$$
$$U_{t-1}(6) = 1400$$

Calculating $U_t(5)$ using the objective value function above:

$$\begin{aligned} U_t(5) = \quad & \sum_{i=1}^{m} p_{it} \max\{f_{it} + U_{t-1}(5+1), U_{t-1}(5)\} \\ & +5q_t U_{t-1}(5-1) + (1 - \sum_{i=1}^{m} p_{it} - 5q_t) U_{t-1}(5) \\ = \quad & 0.2 * \max\{(200+1400), 1500\} + 0.4 * \max\{(150+1400), 1500\} \\ & +0.2 * \max\{(75+1400), 1500\} + 5 * 0.02 * 1600 + 0.1 * 1500 \\ = \quad & 1550 \end{aligned}$$

2. Optimality of the Booking Control Policy

If a customer in class i arrives for a booking request in period t, the airline should determine whether to accept the request and get f_{it}, or reject it and leave this seat to a later request. Let's look at Eq.(3.1), the strategy here is to compare $f_{it} + U_{t-1}(x+1)$ and $U_{t-1}(x)$. So, we will accept this booking request for class i iff:

$$f_{it} + U_{t-1}(x+1) \geq U_{t-1}(x) \Longleftrightarrow f_{it} \geq U_{t-1}(x) - U_{t-1}(x+1)$$

We call $U_{t-1}(x) - U_{t-1}(x+1)$ is the opportunity cost[1] of accepting the booking request for class i at period t. Comparing the revenue generated by accepting the booking request and the opportunity cost of losing this seat taken up by the booking, if $f_{it} \geq U_{t-1}(x) - U_{t-1}(x+1)$, then we accept the request.

Definition 1: For each stage t and each fare class i, define the optimal booking limit B_{it} as:

$$B_{it} = \min\{x \geq 0 : U_{t-1}(x) - U_{t-1}(x+1) > f_{it}\}$$

If $f_{it} \geq U_t(x) - U_t(x+1)$ for all x, the revenue generated is always higher than the opportunity cost, then, we should always accept the booking request, for all x. That means, there is no limit for the bookings at all, we can set $B_{it} = \infty$.

[1]The opportunity cost plays the same role as the expected marginal seat revenue (EMSR) of Belobaba (1989) [6](See also Brumelle and McGill (1993) [13] and Wollmer (1992) [63]). It is also the optimal bid price for Single-leg problem, in the sense of Williamson (1992)[62].

Nevertheless, if the opportunity cost $U_t(x) - U_t(x+1)$ is non-decreasing in x, as x increases, and because f_{it} is fixed, we can find one B_{it} satisfying:

$$\begin{aligned} f_{it} &\geq U_{t-1}(x) - U_{t-1}(x+1) \qquad 0 \leq x < B_{it} \\ f_{it} &< U_{t-1}(x) - U_{t-1}(x+1) \qquad x \geq B_{it} \end{aligned} \tag{3.2}$$

If we can find more than one B_{it} satisfying the conditions as above, we will select the first one. Hence, B_{it} is well-defined.

[**Booking Policy:**]

We will accept a class i request in period t if and only if $0 \leq x < B_{it}$.

Class i booking will be accepted in period t provided the number of previously accepted reservations lies strictly below the optimal booking limit, B_{it}.

So, if we want to find a well-defined finite optimal booking limit B_{it} for the fare class i at t, we have to prove that the opportunity cost function $U_t(x) - U_t(x+1)$ is non-decreasing in x which holds iff the optimal value function $U_t(x)$ is concave in integral values x (Refer to the definition as follows).

[**Definition 2**:] A function $f : \mathcal{Z} \longrightarrow \mathcal{R}$ is concave if $f(s) - f(s+1)$ is non-decreasing in s. (From Stidham (1978) [53])

[**Theorem 1:**] For $t = 0, 1, \cdots, T$, $U_t(x)$ is concave and non-increasing in x. (It will be proved later.)

Corollary 1: For a given t, the opportunity cost: $U_t(x) - U_t(x+1)$ is non-decreasing in $x = 0, 1, \cdots, T - t - 1$.[2]

Hence, the key point for the optimality of this booking control policy is the concavity of optimal value function $U_t(x)$. The following lemmas 1, 2 and 3 are

[2] [38] obtained this main theorem without cancellations, overbooking through an ad-hoc inductive argument that appears to rely on the specific structure of that problem, whereas it is not readily clear how to similarly extend the technique of [38] here. In contrast, it is the presence of the maximization term, and thus, the applicability of Lemma 1 and 2, that induces concavity of the value functions, and not the specific details of the equations themselves.

fundamental to the proof of the concavity of $U_t(x)$. All of them will be proved. Lemma 1 is originally proved in Stidham (1978) [53] and it will be explained in greater details for easy understanding below. Lemma 2 is the result from the theory of stochastic ordering.

Lemma 1: Suppose $g : \mathcal{Z}^+ \longrightarrow \mathcal{R}$. is concave in $s \geq 0$. Let $f : \mathcal{Z}^+ \longrightarrow \mathcal{R}$ be defined by:

$$f(s) = \max_{a=0,1,\cdots,m} \{a \cdot r + g(s+a)\}, \quad s \geq 0 \tag{3.3}$$

for a given real number $r \geq 0$ and nonnegative integer m, then $f(s)$ is concave in $s \geq 0$.

[Proof:]

First, note that $f(s) = \widetilde{f}(s) - s \cdot r$, where

$$\widetilde{f}(s) = \max_{s \leq t \leq s+m} \{t \cdot r + g(t)\}, \qquad s \geq 0$$

Let: $t^* = \arg\max_{t \geq 0}\{t \cdot r + g(t)\}$

Then, from the concavity of g,

$$\widetilde{f}(s) = \begin{cases} (s+m) \cdot r + g(s+m), & s+m \leq t^* \\ t^* \cdot r + g(t^*), & s < t^* < s+m \\ s \cdot r + g(s), & t^* \leq s. \end{cases}$$

For $\quad 1 \leq s < t^* - m$,

$$\begin{aligned} \widetilde{f}(s-1) - \widetilde{f}(s) &= g(s-1+m) - g(s+m) - r \\ &\leq g(s+m) - g(s+1+m) - r \\ &= \widetilde{f}(s) - \widetilde{f}(s+1) \end{aligned}$$

For $\quad t^* - m \leq s \leq t^*$, it follows from the definition of t^* that:

$$\widetilde{f}(s-1) \leq t^* \cdot r + g(t^*) = \widetilde{f}(s) \geq \widetilde{f}(s+1)$$

so that

$$\widetilde{f}(s-1) - \widetilde{f}(s) \leq 0 \leq \widetilde{f}(s) - \widetilde{f}(s+1)$$

Finally, for $s > t^*$,

$$\begin{aligned}\widetilde{f}(s-1) - \widetilde{f}(s) &= g(s-1) - g(s) - r \\ &\leq g(s) - g(s+1) - r \\ &= \widetilde{f}(s) - \widetilde{f}(s+1)\end{aligned}$$

Thus,$\widetilde{f}(s) - \widetilde{f}(s+1)$ is nondecreasing in s, and $\widetilde{f}$ is concave for $s \geq 0$.Therefore, $f(s) = \widetilde{f}(s) - s \cdot r$ is also concave for $s \geq 0$. □

Lemma 2: Let f(y), $y \geq 0$, be a nondecreasing convex function. For each non-negative integer x, let $Y(x) \sim Bin(x, \gamma)$, random variable $(0 < \gamma < 1)$ and let $h(x) := E[f(Y(x))]$. Then $h(x)$ is nondecreasing convex in $x \in (0, 1, \cdots)$. (This result is from the Example 6.A.2 in Shaked and Shanthikumar (1994) [48]

Lemma 3: If $H(x) = g(x) \cdot f(x-1) + (\omega - g(x)) \cdot f(x)$, where $\omega \geq 0$ and $\omega - g(x) \geq 0$, g is concave, non-decreasing function in x and f is concave, non-increasing function in x, then H is a non-increasing concave function in x.

[Proof:]

Let: $\zeta(x) = f(x) - f(x+1)$.

• For f is concave and non-increasing function in x, $\Rightarrow \zeta(x) \geq 0$ and $\zeta(x)$ is non-decreasing in x.

$H(x) = g(x) \cdot f(x-1) + (\omega - g(x)) \cdot f(x)$

$H(x+1) = g(x+1) \cdot f(x) + (\omega - g(x+1)) \cdot f(x+1)$

$\Rightarrow H(x) - H(x+1) = g(x) \cdot \zeta(x-1) + (\omega - g(x+1)) \cdot \zeta(x) \geq 0.$

$\Longrightarrow H$ is non-increasing in x.

$$\begin{aligned} & [H(x) - H(x+1)] - [H(x+1) - H(x+2)] \\ = \; & g(x)(\zeta(x-1) - \zeta(x)) + (\omega - g(x+2))(\zeta(x) - \zeta(x+1)) \\ & + \zeta(x) \cdot [g(x+2) - g(x+1) - (g(x+1) - g(x))] \end{aligned}$$

• $\zeta(x)$ is non-decreasing in x. $\Rightarrow$ The first two terms are non-positive.

• $g(x)$ is non-decreasing and concave function in x.

$\Rightarrow$ $g(x+1) - g(x)$ is non-increasing in x and the third term above is also non-positive.

So we have: $[H(x) - H(x+1)] - [H(x+1) - H(x+2)] \leq 0$ in x.

$\Longrightarrow$ $H(x) - H(x+1)$ is non-decreasing in x.

$\Longrightarrow$ By definition 2, we obtain that H is concave.

$\Longrightarrow$ H is non-increasing and concave in x. $\square$

[*Remark:*] Using Lemmas 1, 2 quoted and Lemma 3 which is proved by us, we will complete the proof of the concavity for the optimal value function $U_t(x)$.

[Proof of Theorem 1:] (which is incomplete in [54])

We will prove Theorem 1 by Induction on t.

We first need to verify that $U_0(x) = E[-\pi(Y(x))]$ is concave and non-increasing in x to start the induction.

We assume that $\pi(\cdot)$ is non-negative, convex, and non-decreasing.

$Y(x)$ has a *binomial* $\sim (x, 1-\beta)$ distribution, $0 < 1-\beta < 1$

By Lemma 2, we get that:

$\Longrightarrow$ $E[\pi(Y(x))]$ is non-decreasing, convex.

$\Longrightarrow$ $U_0(x)$ is thus non-increasing and concave.

We assume that $U_{t-1}(x)$ is concave and non-increasing, and let:

$$g_{it}(x) \doteq \max\{f_{it} + U_{t-1}(x+1), U_{t-1}(x)\}$$

$$U_t(x) = \sum_{i=1}^{m} p_{it} g_{it}(x) + q_t(x) U_{t-1}(x-1) + p_{0t}(x) U_{t-1}(x), \quad 0 \leq x \leq T-t, t \geq 1.$$

By Lemma 1, $g_{it}(x)$ is concave in x. (Note that, this is only the special case of Lemma 1 with $m = 1$). Moreover, $g_{it}(x)$ is non-increasing as the maximum of two non-increasing functions is non-increasing.

Define: $H_{t-1}(x) \doteq q_t(x) U_{t-1}(x-1) + p_{0t}(x) U_{t-1}(x)$.

$p_{0t}(x) = 1 - \sum_{i=1}^{m} p_{it} - q_t(x) = \omega - q_t(x) \geq 0$, where $\omega \geq 0$ (not dependent on x). We assume that $U_{t-1}(x)$ is concave and non-increasing in x. We also know that $q_t(x)$ is concave and non-decreasing in x.

$\Rightarrow H_{t-1}(x)$ is concave and non-increasing in x by Lemma 3.

$$U_t(x) = \sum_{i=1}^{m} p_{it} g_{it}(x) + H_{t-1}(x), \quad 0 \leq x \leq T - t, t \geq 1.$$

[Fact:] If f and g are concave, non-increasing functions, then $f + g$ is a concave, non-increasing function, and so is $\alpha \cdot f$ for a non-negative constant α.

$U_t(x)$ is the sum of two concave, non-increasing functions. So $U_t(x)$ is concave, non-increasing in $x = 0, 1, \cdots$. $\square$

3.4 General Model (Model 2)

In this section, a more general model will be discussed.

1. Assumption and Description

- Class-dependent cancellation and no-show probabilities.
- Refunds at the time of cancellation and no-show are class-dependent. Actually, most airlines refund cancelled seats based on different fare classes.

Let $p_{it}(\mathbf{x})$, $q_{it}(\mathbf{x})$ and $p_{0t}(\mathbf{x})$ respectively, denote the probability of a booking request in fare class i, the probability of a cancellation by customer in class i, and the probability of null event in period t, given the reservation vector $\mathbf{x}$ ($i = 1, 2, \cdots, m$). We also assume that:

$\sum_{i=1}^{m} p_{it}(\mathbf{x}) + \sum_{i=1}^{m} q_{it}(\mathbf{x}) + p_{0t}(\mathbf{x}) = 1$ for all $\mathbf{x}$ and $t \geqslant 1$.

At the time of departure, each customer of class i has a probability β_i of being a no-show, only dependent on the class i. Corresponding to each β_i, $Y_i(x_i)$ denotes the number of people in class i showing up for departure, where $Y_i(x_i) \sim$

$Bin(x_i, 1-\beta_i)$. Let $y = Y(\mathbf{x}) := \sum_{i=1}^m Y_i(x_i)$ denote the total number of customers who show up for the flight. $\pi(y)$ is the overbooking penalty. As in Model 1, we assume that $\pi(y)$ is non-negative, convex and non-decreasing in $y \geq 0$, with $\pi(y) = 0$ for $y \leq C$.

At each stage t, the following transitions possible from state $\mathbf{x}$ are shown in Figure 3.2.

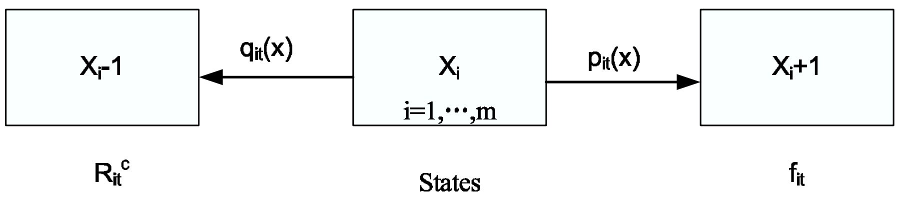


Figure 3.2: Transitions Possible.

As in Model 1, the objective now is to maximize the expected total net revenue over the horizon from period T to period 0, starting from state $\mathbf{x} = (0, \cdots, 0)$, that is, there are no booked seats at the beginning of period T. A customer who requests for a seat in fare class i at stage t pays fare $\widehat{f}_{it}$.

Let $\chi_t := \{\mathbf{x} = (x_1, x_2, \cdots, x_m) : x_i \geq 0, i = 1 : m; \sum_i x_i \leq (T-t)\}$. As a function of the state $\mathbf{x} \in \chi_t$, in period t, let $\widehat{U}_t(\mathbf{x})$ denote the maximal expected net benefit at stage $\mathbf{x}$, which are determined recursively by:

$$\begin{aligned}
\widehat{U}_t(\mathbf{x}) = \quad & \sum_{i=1}^m p_{it}(\mathbf{x}) \max\{\widehat{f}_{it} + \widehat{U}_{t-1}(\mathbf{x}+e_i), \widehat{U}_{t-1}(\mathbf{x})\} \\
& + \sum_{i=1}^m q_{it}(\mathbf{x})(-R_{it}^c + \widehat{U}_{t-1}(\mathbf{x}-e_i)) + p_{0t}(\mathbf{x})\widehat{U}_{t-1}(\mathbf{x}) \quad \mathbf{x} \in \chi_t, t \geq 1 \\
\widehat{U}_0(\mathbf{x}) = \quad & E[-\pi(Y(\mathbf{x})) - \sum_{i=1}^m (x_i - Y_i(x_i))R_i^{ns}], \qquad \mathbf{x} \in \chi_0
\end{aligned} \tag{3.4}$$

where e_i is the ith unit m-vector.

From the Eq.(3.4), we will accept the class i booking if and only if:

$$\begin{aligned}
& \widehat{f}_{it} + \widehat{U}_{t-1}(\mathbf{x}+e_i) \geq \widehat{U}_{t-1}(\mathbf{x}) \\
\Longleftrightarrow \quad & \widehat{f}_{it} \geq \widehat{U}_{t-1}(\mathbf{x}) - \widehat{U}_{t-1}(\mathbf{x}+e_i)
\end{aligned}$$

[**Booking Policy:**] We accept a class i request in stage t with reservation vector $\mathbf{x}$ iff $\widehat{f}_{it} \geq \widehat{U}_{t-1}(\mathbf{x}) - \widehat{U}_{t-1}(\mathbf{x}+e_i)$.

As the different classes of customers have different cancellation rates and refunds, it is necessary to keep track of the number of customers in each class, rather than just the total number. As a result, this model has a multi-dimensional state space.

Model 2 includes the refund for cancellations and no-shows which are ignored by model 1. Due to the computational disadvantage of the multi-dimensional state space, we should consider to reduce this general model to a one-dimensional MDP in certain cases, which can give us a good approximation to the optimal solution. In the numerical example section, we quote an example from [54], which will show that with careful use of the one-dimensional heuristic approximation, the revenue difference between a one-dimensional model and a multi-dimensional model is not much, which suggests that a multi-dimensional model may not always be necessary.

2. The Refund of the Cancellations and No-shows

We consider the refund of cancellations and no-shows in Model 2. We assume that we will reject all additional arrivals, starting from state $\mathbf{x}$ at stage t (not affected by future arrivals). Let $H_t(\mathbf{x})$ as: Total expected loss of revenue over periods t to 0 caused by cancellations and no-shows which is an unavoidable loss of revenue.

H_t is given as follows:

$$\begin{aligned} H_t(\mathbf{x}) = &\quad \sum_{i=0}^{m} p_{it}(\mathbf{x}) H_{t-1}(\mathbf{x}) + \sum_{i=1}^{m} q_{it}(\mathbf{x})(R^c_{it} + H_{t-1}(\mathbf{x}-e_i)) && \mathbf{x} \in \chi_t \\ H_0(\mathbf{x}) = &\quad E[\sum_{i=1}^{m}(x_i - Y_i(x_i))R^{ns}_i] = \sum_{i=1}^{m} \beta_i x_i R^{ns}_i. && \mathbf{x} \in \chi_0 \end{aligned} \quad (3.5)$$

We define: $U_t(\mathbf{x}) := \widehat{U}_t(\mathbf{x}) + H_t(\mathbf{x}), t \geq 0$, which represents the maximal expected controllable net revenue over periods t to 0.

By Eq. (3.4) and Eq. (3.5),

$$\begin{aligned}
U_t(\mathbf{x}) = \quad & \sum_{i=1}^{m} p_{it}(\mathbf{x}) \max\{\widehat{f}_{it} - [H_{t-1}(\mathbf{x}+e_i) - H_{t-1}(\mathbf{x})] + U_{t-1}(\mathbf{x}+e_i), U_{t-1}(\mathbf{x})\} \\
& + \sum_{i=1}^{m} q_{it}(\mathbf{x}) U_{t-1}(\mathbf{x}-e_i) + p_{0t}(\mathbf{x}) U_{t-1}(\mathbf{x}), \qquad \mathbf{x} \in \chi_t, t \geq 1 \\
U_0(\mathbf{x}) = \quad & E[-\pi(Y(\mathbf{x})], \qquad \mathbf{x} \in \chi_0 \qquad (3.6)
\end{aligned}$$

where, $H_{t-1}(\mathbf{x}+e_i) - H_{t-1}(\mathbf{x})$ is the marginal expected cost associated directly with the accepted customer, namely, the expected amount that will be refunded to that customer resulting from cancellation or no-show.

[**Booking Policy**:] We accept a class i request in stage t with reservation vector $\mathbf{x} \Leftrightarrow \widehat{f}_{it} - [H_{t-1}(\mathbf{x}+e_i) - H_{t-1}(\mathbf{x})] \geq U_{t-1}(\mathbf{x}) - U_{t-1}(\mathbf{x}+e_i)$.

3. Reduction to one-dimensional problem

In this part, we will show that: when cancellations and no-show probabilities are fare-independent, the state-space can be reduced, the multidimensional problem can be converted into a one-dimensional problem.

▲**Assumption 1**. $q_{it}(\mathbf{x}) = q_{it}(x_i)$, for all $\mathbf{x} = (x_1, x_2, \cdots, x_m), t = T, T-1, \cdots, 1$. (i.e. Probability of a class i cancellation in a period depends only on x_i).

Lemma 4: Under Assumption 1,

$$H_t(\mathbf{x}) = \sum_{i=1}^{m} H_{i,t}(x_i), \qquad t \geq 0 \qquad (3.7)$$

where the functions $H_{i,t}$ satisfy the recursive equations ($i = 1, \cdots, m$).

$$\begin{cases} H_{i,t}(x_i) = (1 - q_{it}(x_i)) H_{i,t-1}(x_i) + q_{it}(x_i)(R_{it}^{c} + H_{i,t-1}(x_i - 1)), & x_i \geq 0, t \geq 1 \\ H_{i,0}(x_i) = \beta_i x_i R_i^{ns}, & x_i \geq 0 \end{cases} \qquad (3.8)$$

(It can be proved easily by induction on t.)

▲ **Assumption 1'**: $q_{it}(\mathbf{x}) = x_i q_{it}$ for all $\mathbf{x}$, where $q_{it} > 0, i = 1 : m, t = T, T-1, \cdots, 1$ (i.e. each customer cancels independently of all other customers, with a cancellation rate solely dependent on the customers class).

We define: $G_{i,t}(x_i) := H_{i,t-1}(x_i + 1) - H_{i,t-1}(x_i)$ be the marginal expected cancellation cost associated with fare-class i booking in state $\mathbf{x}$ at t.

Lemma 5: Under Assumption 1',

$$G_{i,t}(x_i) = G_{i,t}(x_i - 1) \qquad \text{(which is not shown in [54].)}$$

where $G_{i,t}(x_i)$ is the marginal expected cancellation cost of accepting a request for a seat in fare class i in t. Furthermore, we define $G_t(i)$ satisfies the recursive equations:

$$\begin{aligned} G_t(i) = &\quad q_{i,t-1} R^c_{i,t-1} + (1 - q_{i,t-1}) G_{t-1}(i), \quad t \geq 2 \\ G_1(i) = &\quad \beta_i R^{ns}_i, \qquad\qquad\qquad i = 1, \cdots, m \end{aligned}$$

where $G_t(i)$ is the expected cancellation cost attributable to that customer which is independent of x_i. Then we have:

$$G_{i,t}(x_i) = G_t(i)$$

[**Proof:**] (By induction on t).

$$\begin{aligned} G_{i,1}(x_i) = &\quad H_{i,0}(x_i + 1) - H_{i,0}(x_i) = \beta_i R^{ns}_i \\ G_{i,1}(x_i - 1) = &\quad H_{i,0}(x_i) - H_{i,0}(x_i - 1) = \beta_i R^{ns}_i \end{aligned}$$

Hence, $G_{i,1}(x_i) = G_{i,1}(x_i - 1)$ holds.
Next we will prove that $G_{i,t}(x_i) = G_t(i)$ for $t \geq 2$.

Suppose $G_{i,t-1}(x_i) = G_{i,t-1}(x_i - 1)$ holds for $t \geq 2$.

$$\begin{aligned}
G_{i,t}(x_i) &= H_{i,t-1}(x_i+1) - H_{i,t-1}(x_i) \\
&= (q_{i,t-1}(x_i+1) - q_{i,t-1}(x_i))R^c_{i,t-1} + (1 - q_{i,t-1}(x_i+1))G_{i,t-1}(x_i) \\
&+ q_{i,t-1}(x_i)G_{i,t-1}(x_i-1), \qquad x_i \geq 0, t \geq 2 \\
&= ((x_i+1)q_{i,t-1} - x_i q_{i,t-1})R^c_{i,t-1} + (1 - (x_i+1)q_{i,t-1})G_{i,t-1}(x_i) \\
&+ x_i q_{i,t-1} G_{i,t-1}(x_i-1) \\
&= q_{i,t-1}R^c_{i,t-1} + (1 - q_{i,t-1})G_{i,t-1}(x_i) - x_i q_{i,t-1} G_{i,t-1}(x_i) \\
&+ x_i q_{i,t-1} G_{i,t-1}(x_i-1) \\
&= q_{i,t-1}R^c_{i,t-1} + (1 - q_{i,t-1})G_{i,t-1}(x_i) - x_i q_{i,t-1}[G_{i,t-1}(x_i) - G_{i,t-1}(x_i-1)] \\
&= q_{i,t-1}R^c_{i,t-1} + (1 - q_{i,t-1})G_{i,t-1}(x_i)
\end{aligned}$$

$$\begin{aligned}
G_{i,t}(x_i-1) &= H_{i,t-1}(x_i) - H_{i,t-1}(x_i-1) \\
&= q_{i,t-1}R^c_{i,t-1} + (1 - x_i q_{i,t-1})G_{i,t-1}(x_i) + (x_i-1)q_{i,t-1}G_{i,t-1}(x_i-1) \\
&= q_{i,t-1}R^c_{i,t-1} + G_{i,t-1}(x_i) - q_{i,t-1}G_{i,t-1}(x_i-1) \\
&- x_i q_{i,t-1}(G_{i,t-1}(x_i) - G_{i,t-1}(x_i-1) \\
&= q_{i,t-1}R^c_{i,t-1} + G_{i,t-1}(x_i) - q_{i,t-1}G_{i,t-1}(x_i-1)
\end{aligned}$$

$$\begin{aligned}
\because G_{i,t-1}(x_i) &= G_{i,t-1}(x_i-1) \\
\therefore G_{i,t}(x_i) &= G_{i,t}(x_i-1)
\end{aligned}$$

And we get:

$$\begin{aligned}
G_{i,t}(x_i) &= q_{i,t-1}R^c_{i,t-1} + (1 - q_{i,t-1})G_{i,t-1}(x_i), \quad t \geq 2 \\
G_{i,1}(x_i) &= \beta_i R^{ns}_i. \qquad i = 1, \cdots, m
\end{aligned}$$

We know that the functions $G_t(i)$ satisfy the recursive equations:

$$\begin{aligned}
G_t(i) = &\quad q_{i,t-1}R^c_{i,t-1} + (1 - q_{i,t-1})G_{t-1}(i), \quad t \geq 2 \\
G_1(i) = &\quad \beta_i R^{ns}_i, \qquad i = 1, \cdots, m
\end{aligned}$$

$G_{i,t}(x_i)$ and $G_t(i)$ have the same recursive functions, so we get: $G_{i,t}(x_i) = G_t(i)$. We prove the Lemma 4. $\Box$.

▲ In this case, the recursive optimality equations, which now take the form:

$$\begin{aligned} U_t(\mathbf{x}) = & \sum_{i=1}^{m} p_{it}(\mathbf{x}) \max\{\widehat{f}_{it} - G_t(i) + U_{t-1}(\mathbf{x} + e_i), U_{t-1}(\mathbf{x})\} \\ & + \sum_{i=1}^{m} x_i q_{it} U_{t-1}(\mathbf{x} - e_i) + p_{0t}(\mathbf{x}) U_{t-1}(\mathbf{x}), \quad \mathbf{x} \in \chi_t, t \geq 1 \\ U_0(\mathbf{x}) = & E[-\pi(Y(\mathbf{x}))], \qquad \mathbf{x} \in \chi_0 \end{aligned} \tag{3.9}$$

[**Booking Policy:**] We will accept a class i request in stage t with reservation vector $\mathbf{x}$ iff $\widehat{f}_{it} - G_t(i) \geq U_{t-1}(\mathbf{x}) - U_{t-1}(\mathbf{x} + e_i)$.

We know that $U_t(\mathbf{x})$ are recursive, dynamic which depend on $\mathbf{x}$ which are not one-dimensional. So, the curse of dimensionality is still there.

Assumption 1". $q_{it}(\mathbf{x}) = x_i q_t$, for all $\mathbf{x} \in \chi_t, i = 1 : m$, where $q_t > 0, t = T, T-1, \cdots, 1$ (cancellation rates are same for all fare class $i = 1 : m$ in period t).

Assumption 2. $p_{it}(\mathbf{x}) = p_{it}$, for all $\mathbf{x} \in \chi_t, i = 1 : m, t = T, T-1, \cdots, 1$, i.e. arrivals of booking requests are independent of the number of seats already booked.

Assumption 3. $\beta_i = \beta, i = 1 : m$, the no-show probabilities are independent of the fare class.

[**Theorem 2:**] Under Assumptions 1",2,3, the optimal value functions, $U_t(\mathbf{x})$, depend on $\mathbf{x}$ only through $x = \sum_{i=1}^{m} x_i$, and are determined by the recursive optimality equations:

$$\begin{aligned} U_t(\mathbf{x}) = U_t(x) = & \sum_{i=1}^{m} p_{it} \max\{\widehat{f}_{it} - G_t(i) + U_{t-1}(x+1), U_{t-1}(x)\} + x q_t U_{t-1}(x-1) \\ & + (1 - \sum_{i=1}^{m} p_{it} - x q_t) U_{t-1}(x), \qquad 0 \leq x \leq T - t, t \geq 1 \\ U_0(\mathbf{x}) = U_0(x) = & E[-\pi(Y(x))], \qquad 0 \leq x \leq T \end{aligned} \tag{3.10}$$

where $Y(x) \sim Bin(x, 1-\beta)$.

[Proof:] (The proof is by induction on t.)

Suppose $U_0(\mathbf{x}) = U_0(x) = E[-\pi(Y(x))]$ depending on $\mathbf{x}$ only through $x = \sum_{i=1}^{m} x_i$. Let $t \geq 1$ and suppose $U_{t-1}(\mathbf{x}) = U_{t-1}(x)$ for all $\mathbf{x}$. Then, it follows from Eq.(3.9) and Assumptions 1",2, and 3 that

$$\begin{aligned} U_t(\mathbf{x}) = \quad & \sum_{i=1}^{m} p_{it} \max\{\widehat{f}_{it} - G_t(i) + U_{t-1}(x+1), U_{t-1}(x)\} + xq_t U_{t-1}(x-1) \\ & +(1 - \sum_{i=1}^{m} p_{it} - xq_t)U_{t-1}(x), \qquad 0 \leq x \leq T-t, t \geq 1 \end{aligned} \tag{3.11}$$

where $x = \sum_{i=1}^{m} x_i$ and we use the fact $\sum_{i=1}^{m} p_{it} + \sum_i x_i q_t + p_{0t}(\mathbf{x}) = 1$. It follows from Eq.(3.11) that $U_t(\mathbf{x}) = U_t(x)$ only depending on the total number of seats booked $x = \sum_{i=1}^{m} x_i$, so Eq.(3.11) holds. This completes the induction and the proof of the theorem 2. □

So, when cancellation and no-show probabilities are independent of the fare class, the Model 2 is converted to an equivalent form with one-dimensional state variable: the total number of seats booked $x = \sum_{i=1}^{m} x_i$.

[Optimality:] Eq.(3.10) has exactly the form of the one-dimensional problem (Model 1) discussed in the previous section, with $f_{it} = \widehat{f}_{it} - G_t(i)$, and $q_t(x) = xq_t$ (a concave function of x), so all the monotonicity results for Model 1 apply, in particular, the optimality of the booking control policy.

3.5 Numerical Example

To reduce the computational burden, it is advantageous to introduce a maximum overbooking pad Q, resulting in an additional state constraint $0 \leq x \leq C + Q$, at each period t.

For the one-dimensional model, assume that booking requests will always be rejected in state $C + Q$ at t. The recursive optimality equations in this case can be written as:

$$\begin{aligned} U_t(x) = & \sum_{i=1}^{m} p_{it} \max\{f_{it} - (U_{t-1}(x) - U_{t-1}(x+1))), 0\} \\ & + x q_t U_{t-1}(x-1) + (1 - x q_t) U_{t-1}(x) \qquad 0 \le x \le C+Q-1, t \ge 1 \\ U_t(C+Q) = & (C+Q) q_t U_{t-1}(C+Q-1) + (1 - (C+Q) q_t) U_{t-1}(C+Q), \qquad t \ge 1 \\ U_0(x) = & E[-\pi(Y(x))], \qquad 0 \le x \le C+Q \end{aligned} \tag{3.12}$$

where q_t is the average cancellation rate. Similarly, for the multi-dimensional example, we also use the corresponding special case of Eq.(3.4), with the addition of an overbooking pad Q.

[**Data:**]

The cornerstone of the approach demonstrated in this Chapter is the inclusion of customer cancellations. We quote a small example from [54] with class-dependent cancellation and no-show rates.

Available capacity is $C = 4$ with an overbooking pad of $Q = 2$.

There are 2 classes, with $f_1 = 3$ and $f_2 = 1$. Class 1 is fully refundable, $R^c_{1t} = R^{ns}_1 = 3$, whereas class 2 is non-refundable, $R^c_{2t} = R^{ns}_2 = 0$. Penalties of 2 and 6 correspond to overbooking levels of 1 and 2, respectively. The remaining parameters are summarized in Figure 3.3.

Compare the performance of four methods in order of increasing accuracy (and complexity):

Method 1: Model from [38], which completely ignore cancellations and no-shows, however, add a overbooking pad to the physical capacity and allocate the booking limits for each fare class.

Method 2: Model 1, subtracting the expected cancellations and no-shows refund from the gross fare, but ignoring the effects of cancellations (probabilities) on future seat availabilities.

Method 3: Model 1, incorporating both refunds and probabilities of cancellation and no-show, but using approximate probabilities (class-independent).

Method 4: Model 2, with both class-dependent refunds and probabilities (multi-dimensional).

Here, Method 3 is used for the purpose of comparison with three different rates, summarized in Figure 3.4.

Period t \ Parameters	p_{1t}	p_{2t}	p_{3t}	p_{4t}
16-13	0.00	0.30	0.00	0.00
12-9	0.30	0.50	0.00	0.00
8-5	0.30	0.00	0.05	0.00
4-1	0.40	0.00	0.10	0.00
0	0.00	0.00	0.20	0.00

Figure 3.3: Parameters

Period t \ Method	3a	3b	3c
16-9	0.000	0.000	0.000
8-5	0.025	0.050	0.020
4-1	0.050	0.100	0.040
0	0.100	0.200	0.080

Figure 3.4: Comparison of Class-Independent Cancellation Rates.

Method 3a—average the cancellation and no-show rates of the two classes.
Method 3b—simply use the rate corresponding to class 1.
Method 3c—use 40% of the class 1 cancellation and no-show rates.

Using Methods 1,2,3a-3c, the simplified one-dimensional problem is solved optimally, and the results are compared with those obtained using Method 4. The expected revenues obtained by each of the methods are summarized in Figure 3.5, where % Sacrificed is the additional revenue that could be gained by solving the problem optimally (Method 4), expressed as a percentage of the revenue obtained using the given method. Figure 3.6 compares the values of $U_t(0)$ for each of the four methods.

Method	1	2	3a	3b	3c	4
$U_{16}(0.0)$	5.86	5.74	6.22	5.05	6.38	6.41
%Sacrificed	9.39	11.67	3.05	26.93	0.47	-

Figure 3.5: Summary of Methods 1-4

[**Result 1:**] The approaches introduced in this Chapter are computational feasible using data from a real-life airline application.

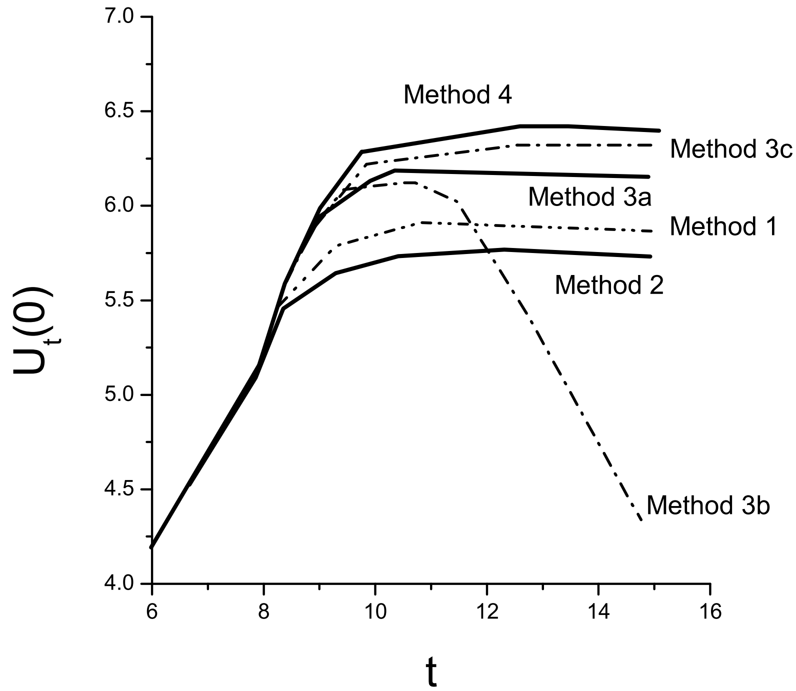


Figure 3.6: $U_t(0)$ versus t for the different methods.

[**Result 2:**] Revenue decreases when going from Method 1 to Method 2.

Note: Even though Method 2 appears to be a more accurate model in that cancellation and no-show refunds are subtracted from the gross fare, it doesn't take into account the effects of cancellations on future seat availabilities, which send the wrong signal to the algorithm for this example.

[**Result 3:**] An increase of nearly 9% can be achieved by incorporating both refunds and cancellation and no-show probabilities, with careful use of the one-dimensional heuristic approximation (Method 3c).

Note: The revenue increment from Method 3 to Method 4 is less than 1%, which suggests that the full multi-dimensional model (Method 4) may not always be necessary.

[**Result 4:**] Method 3b which simply use the cancellation rate corresponding to class 1 results in a bad approximation. We note the importance of choosing the approximation cancellation rate properly.

3.6 Evaluation

One of the advantages of dynamic models is that they can easily be extended to include cancellations, no-shows, and overbooking. The discrete time model in this chapter assumed that, in each time period, at most one booking request or one cancellation can occur. When the number of time periods is very large, this approach gives a good approximation of the continuous time model. However, for the discrete-time model, the size of the state space always poses a threat to such application to real problems when either the inventory position is high or the length of time-to-go is big. The interdependence between different states of time and inventories makes it difficult to apply the recursive solution procedure which works well in continuous-time revenue management models.

Most researchers have discovered that direct implementation of the obvious dynamic programming solution techniques leads to algorithms that are computationally intractable when applied to problems of practical size. They formulated continuous-time models whose computation effort is fairly mild and tackled the optimal booking policy with cancellations and no-shows (Refer to Feng, Lin and Xiao (2001) [26]).

Alternatively, two general approaches can be used to overcome this difficulty:

1. Make restrictive assumptions or to suppress certain elements of the problem.
2. To use heuristics that lead to suboptimal, but easily implemented rules.

Both approaches can provide approximate solutions to partially restricted or even practical versions of the problem, but we should be careful what the effects of changes in parameters of the problem are induced.

To conclude this chapter, we pose a few items which are in the queue of future research. They may be significant from theoretical and practical viewpoints:

1. One may consider some efficient heuristic approaches to approximate the optimal policy for the models in this Chapter.

2. One may consider optimal booking control for the single-leg flight models with group/batch demand. This is particularly relevant when airlines deal with travel agencies who often book a large number of tickets by pooling individuals' booking requests. Booking process with both individual and group demands can be modelled as a compound Poisson Process, which may drastically increase the degree of analytical complication.

3. A more general model that considers the disposition of refused reservations request, which might not always result in a booking loss to the airline (it can be a request to a higher fare class, same flight; a shift to a different flight, same fare class and airline or a booking loss for the airline).

4. It is also conceivable that the technique may be applied to control the entire flight network simultaneously, which will be discussed in the next Chapter.

Chapter 4

Overbooking in Network Environment

4.1 Introduction

Airlines tend to operate as a network more than as individual flights. Maximizing revenues on each flight leg individually in no way guarantees that total network revenues are being maximized (Williamson (1992) [62]). The *core problem* in this chapter for network revenue management (NRM) is determining optimal decision rules for sequentially accepting or denying Origin-Destination-Fare (ODF) itinerary requests to maximize the airline's expected revenue. The *decision rule* here is: grant the booking request only if the offered fare exceeds the opportunity cost of that ODF itinerary. The *major issues* associated with this chapter for the network overbooking controls are: the availability of data at the itinerary fare class level, demand forecasting, cancellation process at this level, the mathematical optimization tools necessary for controlling seats at the network level, etc.

Besides the emphasis on network-based rather than leg-based models, there is an increased interest in models for dynamic booking control. That is, we are more interested in investigating the design of dynamic overbooking policy for allocating inventory to correlated, stochastic demand for multiple classes on multiple legs, so as to maximize total expected revenue in a network environment at each point of time. Several optimization models addressing the NRM problem will be introduced.

Two approximate dynamic programming algorithms incorporating cancellations, no-shows and overbooking will be presented and analyzed both theoretically and computationally.

The discussions are mainly based on Bertsimas and Popescu (2001) [7], Boer et al. (2002) [11] and Williamson (1992) [62]. However, all these papers do not emphasize on overbooking, even in [7]. In this chapter, we will pay more attention on the dynamic overbooking policy in the network environment. Several major results from [7] will be explained in greater details for better understanding. We check the rationality of the two approximate dynamic programming algorithms proposed in [7], by completing the proofs. Furthermore, more structural properties of the two algorithms are dig out. Some other related papers referred to in this chapter are: [8], [56], and [57].

4.2 Problem Definition and Notations

We give an airline network composed of l legs, which are used to serve a total of m ODF itinerary demand classes.

Let:

$\mathbf{N} = (N_1, \cdots, N_l)$: total initial network capacity.

$A = (a_{ji})_{l\times m}$: leg-class incidence matrix, where A^i is the ith column of A and $a_{ji} = 1$ if itinerary i uses leg j and $a_{ji} = 0$ otherwise.

$\triangleright$ In an airline network without group discounts, A is a 0 - 1 matrix which may contain repeated columns for different fare classes on same itinerary. Assume: A^i has at least one nonzero component, i.e. all itineraries use at least one leg.

$\mathbf{f} = (f_1, \cdots, f_m)$: f_i is the fare category of ith ODF itinerary request[1] Assume:

[1]A much more serious practical obstacle is the need to forecast expected demand for each Origin-Destination-Fare itinerary in a given connecting network. We have taken the fare classes on each Origin-Destination into account, because different fare classes make different revenue contributions to the network. So, we define that different ODF itinerary requests for different fare classes, even they have identical origin and destination.

The rejected requests for itineraries are lost to the network. In particular, we don't model diversion/upgrade among itineraries.

[**Example**:]

Consider a very simple network, where there are 3 nodes A, B and C, and with $l = 2$ legs: (1)AB and (2)BC with total capacity $\mathbf{N} = (N_1, N_2)$. Suppose there is demand for all itineraries: (1)AB, (2)BC, and (3)ABC, with only one fare class for each itinerary. Furthermore, there are discounts for groups of size $k_1 = 8$ for (1)AB, at a rate of f_1^8 per group. $\Longrightarrow$ total of $m = 4$ classes.

So,

$$\begin{pmatrix} \mathbf{f} \\ A \end{pmatrix} = \begin{pmatrix} f_1 & f_2 & f_3 & f_1^8 \\ 1 & 0 & 1 & 8 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$\mathbf{s}$ = sales to date vector ($\mathbf{s} = (s_1, \cdots, s_m)$ which is a m-vector).

s_i^t = the number of seats sold to ith ODF itinerary until time t.

s_i^o = the number of seats overbooked at departure for ith ODF itinerary.

e_i = a unit vector with ith column is 1.

The time is discrete. The state of the system $\mathbf{S} = (t, \mathbf{s})$ is given by the time t (t periods to departure) and the sales-to-date record $\mathbf{s}$ for each ODF itinerary.

4.3 General Models

The problem of dynamic overbooking control for network revenue management belongs to the class of finite horizon decision problems under uncertainty. The DP approach which will be presented below can be viewed as extending the single-leg models investigated in Chapter Three and Bitran and Mondschein (1995) [9], and is similar to the early single-leg DP formulation of Rothestein (1971, 1974) [44][45]. The DP approach is based on the perfect state information and is not feasible in practice because of the enormous size of the state space.

The LP formulation introduced later is similar to the one proposed by Williamson (1992) [62]. The LP formulations belong to the Static models, which are easy to implement. Furthermore, the LP approach can handle multiple classes and group bookings, and incorporate cancellations, no shows and overbooking. In the dynamic network setting, the static model can be used to approximate DP by solving the model at several fixed times during the booking process.

1. Dynamic Programming Model Allowing Overbooking

In the case of perfect state information, this model provides the optimal control policy. Since we allow cancellations, we cannot use the vector of remaining capacities as the state variable. It is necessary to keep track of the past sales record $\mathbf{s}$. The state space is large here. The random quantities involved are the demand, cancellation and no-show processes. We assume that no refund for no-shows here. Given the initial network inventory $\mathbf{N}$, define $DP^o(\mathbf{s}, t)$ as the maximum expected net benefit of operating the system over periods t to 0, given by:

$$\begin{aligned}
DP^o_{\mathbf{N}}(\mathbf{s}, t) &= \sum_i p_{it} \cdot \max(DP^o_{\mathbf{N}}(\mathbf{s}, t-1), f_i + DP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1)) \\
&\quad + \sum_{i|s_i \geq 1} q_{it} \cdot (DP^o_{\mathbf{N}}(\mathbf{s} - e_i, t-1) - R^{ct}_i) + p^t_0 \cdot DP^o_{\mathbf{N}}(\mathbf{s}, t-1) \\
DP^o_{\mathbf{N}}(\mathbf{s}, 0) &= \sum_{\widetilde{\mathbf{s}}} P(\widetilde{\mathbf{s}} \quad \text{bookings out of } \mathbf{s} \text{ show up}) \cdot [DP^o_{\mathbf{N}}(\widetilde{\mathbf{s}}, -1) - (\mathbf{R}^{ns})' \cdot (\mathbf{s} - \widetilde{\mathbf{s}})] \\
DP^o_{\mathbf{N}}(\widetilde{\mathbf{s}}, -1) &= \begin{cases} -\min & (\mathbf{R}^o)' \cdot \mathbf{s}^o \\ s.t. & \mathbf{A} \cdot (\widetilde{\mathbf{s}} - \mathbf{s}^o) \leq \mathbf{N} \\ & 0 \leq \mathbf{s}^o \leq \widetilde{\mathbf{s}} \end{cases}
\end{aligned} \tag{4.1}$$

where,

$\mathbf{s}^o$	The vector for the number of customers which are overbooked;
$\mathbf{R}^o$	$\mathbf{R}^o = (R^o_1, \cdots, R^o_m)$, R^o_i is the penalty to the customer who is overbooked for *ith* ODF itinerary;
R^{ct}_i	The refund to one cancellation at time t for *ith* ODF itinerary;
$\mathbf{R}^{ns}$	The refund vector for no-show.
p_{it}	Prob. of a booking request for *ith* ODF itinerary at time t;

q_{it} Prob. of cancellation for *ith* ODF itinerary at time t;

p_0^t Prob. of no request (reservation or cancellation) at time t.

We assume that $\sum_{i=1}^{m} p_{it} + \sum_{i=1}^{m} q_{it} + p_0^t = 1$ for all $t \geqslant 1$.

The boundary conditions are changed to account for no-shows (time $t = 0$) and final bumping decisions (time $t = -1$). The final bumping decision is made so as to minimize total penalties, while keeping the actual capacity restrictions satisfied.

If customer-walking penalties are paid per leg, i.e. $\mathbf{c} = (c_1, c_2, \cdots, c_l)'$, rather than per itinerary, the boundary condition at $t = -1$ is simply $\mathbf{c}'(\mathbf{As} - \mathbf{N})^+$.[2] When bumping penalties are itinerary specific (not leg-additive), then an optimization problem needs to be solved to decide which passengers should be refused boarding so as to incur least penalties. For example, in a two-leg network which is oversold by one seat on each leg, it is better to bump a connecting passenger rather than two different passengers on each leg whenever overbooking penalties are leg-subadditive.

If a customer in *ith* ODF itinerary arrives for a booking request in period t, the airline should determine whether to accept the request and get f_i or reject it and leave this seat to a later request. From Eq.(4.1), the strategy is to compare $f_i + DP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1)$ and $DP^o_{\mathbf{N}}(\mathbf{s}, t-1)$. So, we will accept iff $f_i + DP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1) \geq DP^o_{\mathbf{N}}(\mathbf{s}, t-1)$.

[**Booking Policy for DP Model:**]

At any given state $\mathbf{S}$, accept a booking request for *ith* ODF itinerary iff

$$f_i \geq OC_i(\mathbf{s}, t) = DP^o_{\mathbf{N}}(\mathbf{s}, t-1) - DP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1). \tag{4.2}$$

where, $OC_i(\mathbf{s}, t)$ is the opportunity cost of selling one booking request for *ith* ODF itinerary at time t. However, computing $DP^o_{\mathbf{N}}(\cdot, t)$ is not feasible in practice because of the enormous size of the state space[3]. Therefore, the only practical option

[2] Use the operator $(x)^+ = \max(x, 0)$, for $x \in \mathcal{R}$, which naturally extends for vectors: $(\mathbf{x})^+ = ((x_1)^+, \cdots, (x_n)^+)$ for $\mathbf{x} = (x_1, \cdots, x_n) \in \mathcal{R}^n$.

[3] DP formulations of the revenue management problem are required to properly model real

in practice is to use approximate methods.

2. Integer and Linear Programming Models Allowing Overbooking

The static model is often used to approximate dynamic model by solving the model at several fixed times during the booking process. Actually, static model is the most frequently used formulation for the NRM problem. Let's consider its most general formulation without overbooking first.

$$\begin{aligned} \max \quad & E[\sum_i f_i \cdot \min\{z_i, \overline{D}_i^t\}] \\ s.t. \quad & \sum_{i \quad \text{uses leg} j} z_i \leq N_j \quad \text{For all flight legs} j = 1, 2, \cdots, l. \\ & z_i \geq 0, \quad \text{integer for all itineraries} \quad i = 1, 2, \cdots, m \end{aligned} \tag{4.3}$$

The demand process at time t is denoted by D^t , and $\overline{D}^t$ represents the corresponding random vector of cumulative demands. That is, $\overline{D}_i^t$ is a random variable representing the number of ith ODF itinerary to come from time t to departure time. Usually, only partial information about the demand process could be obtained, which might consist of the expected demand to come $\mathbf{D}^t = E[\overline{D}^t]$. Based on expected demand information, a simple approximation to E.q.(4.3) is to substitute each $\overline{D}^t$ by $\mathbf{D}^t = E[\overline{D}^t]$, resulting in the Integer Programming(IP) model:

$$\begin{aligned} \max \quad & E[\sum_i f_i \cdot y_i] \\ s.t. \quad & \sum_{i \quad \text{uses leg} j} y_i \leq N_j \quad \text{For all flight legs} j = 1, 2, \cdots, l. \\ & y_i \leq \mathbf{D}_i^t \\ & y_i \geq 0, \quad \text{integer for all itineraries} \quad i = 1, 2, \cdots, m \end{aligned} \tag{4.4}$$

where $\mathbf{y}$ is the integer vector that decides how many seats to be reserved in the future for ODF itineraries to maximize the airline's expected revenue.

world factors like cancellations, overbooking, batch bookings, and interspersed arrivals. But, exact DP formulations, particularly stochastic ones, are well known for their unmanageable growth in size when real world implementations are attempted. However, DP can be used as a calibration tool for checking the performance of less accurate but more efficient solution methods.

This will result in an IP approximation model to the value function which is obtained by finding a tentative itinerary allocation assuming that demand to come is always equal to its mean. Such a static model is used to approximate dynamic model without considering overbooking by solving the model at several fixed times during the booking process.

Similarly, we want to get an IP approximation model to Eq.(4.1) considering cancellations, no-shows and overbooking. Here, we assume that the cancellation and no-show for each given ODF itinerary reservation are independent.

Let:

$\mathbf{s}^o$	A vector for the number of seats (itinerary requests) that the airline decides to overbook.
q_i^c	The probability that a given *ith* ODF itinerary reservation is cancelled at some point in the booking period.
γ_i^{ns}	The probability that a given *ith* ODF itinerary reservation does not show up for the flight at departure time.
R^c	The refund received by a cancellation.
R^{ns}	The refund received by a no-show.

For the cancellations and no-shows are taken into account, both kind of penalties should be considered. And we allow overbooking at the departure, so, the penalty for the denied passengers should also be included:

1: Given *ith* ODF itinerary reservation is cancelled at some point in the booking period and the customer gets refund R_i^c immediately.

2: Given *ith* ODF itinerary reservation is not cancelled at any point in the booking period. However, it doesn't show up at departure time. Customer gets refund R_i^{ns}.

3: Given *ith* ODF itinerary reservation is not cancelled at any point in the booking period. It shows up at departure time. However, it is rejected boarding in the end and the customer gets refund R_i^o.

The objective is to maximize the total net benefit over the network of the flights at time t is as follows:

$$
\begin{aligned}
IP_N^o(\mathbf{s},t) = \max \quad & \sum_i [f_i \cdot y_i - q_i^c \cdot y_i \cdot R_i^c - (1-q_i^c) \cdot \gamma_i^{ns} \cdot y_i \cdot R_i^{ns}] \\
& - \sum_i [(1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot s_i^o \cdot R_i^o]
\end{aligned}
$$

There two types of constraint functions: the capacity constraints on each flight leg and the demand constraints associated with each passenger as follows:

1. The demand constraint is:

$$0 \le \mathbf{y} \le \mathbf{D}^t$$

2. The capacity constraints on each flight leg for $i = 1, \cdots m$ are:

$$0 \le (1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot (y_i + s_i) \cdot A^i - (1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot s_i^o \cdot A^i \le N_i.$$

We define:

$$
\begin{aligned}
& f_i \cdot y_i - q_i^c \cdot y_i \cdot R_i^c - (1-q_i^c) \cdot \gamma_i^{ns} \cdot y_i \cdot R_i^{ns} \\
& = [f_i - q_i^c \cdot R_i^c - (1-q_i^c) \cdot \gamma_i^{ns} \cdot R_i^{ns}] \cdot y_i \doteq \widetilde{f}_i \cdot y_i
\end{aligned}
$$

$\widetilde{f}_i$ is defined to be the expected revenue before the overbooking period for ith ODF itinerary at time period t.

$$
\begin{aligned}
& (1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot (y_i + s_i) \cdot A^i \\
& = (y_i + s_i) \cdot [(1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot A^i] \doteq (y_i + s_i) \cdot \widetilde{A}^i
\end{aligned}
$$

We denote $(y_i + s_i) \cdot \widetilde{A}^i$ as the expected capacity occupied by ith ODF itinerary reservations at the end of horizon.

$(1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot s_i^o$ is the number actually overbooked in the end for the ith itinerary. We let:

$$
\begin{aligned}
(1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot s_i^o \cdot A^i &\doteq s_i^o \cdot \widetilde{A}^i, \qquad i = 1, \cdots m. \\
(1-q_i^c) \cdot (1-\gamma_i^{ns}) \cdot s_i^o \cdot R_i^o &\doteq s_i^o \cdot \widetilde{R}_i^o
\end{aligned}
$$

$\widetilde{R}_i$ is defined as the average overbooking penalty of one *ith* ODF itinerary reservation.

Hence, the integer programming (IP) approximation model, that maximizes the expected net benefit at time t subject to expected capacity constraints is:

$$\begin{aligned} IP_N^o(\mathbf{s},t) = \max \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^o}' \cdot \mathbf{s}^o \\ s.t. \quad & 0 \le \widetilde{\mathbf{A}} \cdot (\mathbf{y} + \mathbf{s} - \mathbf{s}^o) \le \mathbf{N} \\ & 0 \le \mathbf{y} \le \mathbf{D}^t \\ & 0 \le \mathbf{s}^o \le \mathbf{y} + \mathbf{s} \\ & \mathbf{y}, \mathbf{s}^o \quad \text{integer.} \end{aligned}$$

[*Note:*] Fractions of seats cannot be sold. The seat allocations must be integral numbers. The difficulty of such a constraint is that integer solutions usually require a considerable amount of extra processing. Simple rounding of a non-integer solution often does not give the optimal integer solution and can be significantly different from it.

The corresponding LP relaxation is explored below which is appealing precisely because it is so simple and computationally efficient. And simulation studies have shown that with frequent re-optimizing, the LP approximation is quite good, producing a higher expected revenue compared to some other approximation schemes.

$$\begin{aligned} LP_N^o(\mathbf{s},t) = \max \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^o}' \cdot \mathbf{s}^o \\ s.t. \quad & 0 \le \widetilde{\mathbf{A}} \cdot (\mathbf{y} + \mathbf{s} - \mathbf{s}^o) \le \mathbf{N} \\ & 0 \le \mathbf{y} \le \mathbf{D}^t \\ & 0 \le \mathbf{s}^o \le \mathbf{y} + \mathbf{s} \end{aligned} \tag{4.5}$$

Its dual can be expressed as follows:

$$\begin{aligned}
\min \quad & \mathbf{v}' \cdot \mathbf{N} - \mathbf{u}' \cdot \mathbf{s} + (\widetilde{\mathbf{f}}' - \mathbf{u}')^{+} \cdot \mathbf{D}^{t} \\
s.t. \quad & \mathbf{u}' = \min(\widetilde{\mathbf{R}^{o}}', \mathbf{v}' \cdot \widetilde{\mathbf{A}}) \\
& \mathbf{v} \geq 0.
\end{aligned} \tag{4.6}$$

This formulation can be equivalently written as:

$$LP_N^o(\mathbf{s}, t) = \min_{\mathbf{v} \geq 0} \mathbf{v}' \cdot \mathbf{N} - (\min(\widetilde{\mathbf{R}^{o}}', \mathbf{v}' \cdot \widetilde{\mathbf{A}})) \cdot \mathbf{s} + (\widetilde{\mathbf{f}}' - \min(\widetilde{\mathbf{R}^{o}}', \mathbf{v}' \cdot \widetilde{\mathbf{A}}))^{+} \cdot \mathbf{D}^{t}$$

Thus the objective value function is piecewise linear, concave in the expected demand to come $\mathbf{D}^t$ and $\mathbf{s}$.

[**Proof:**]

$$\begin{aligned}
LP_N^o(\mathbf{s}, t) = \max \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^{o}}' \cdot \mathbf{s}^{o} & = \max \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^{o}}' \cdot \mathbf{s}^{o} \\
s.t. \quad & 0 \leq \widetilde{\mathbf{A}} \cdot (\mathbf{y} + \mathbf{s} - \mathbf{s}^{o}) \leq \mathbf{N} & s.t. \quad & \widetilde{\mathbf{A}} \cdot (\mathbf{y} + \mathbf{s} - \mathbf{s}^{o}) \leq \mathbf{N} \\
& 0 \leq \mathbf{y} \leq \mathbf{D}^{t} & & \widetilde{\mathbf{A}} \cdot (-\mathbf{y} - \mathbf{s} + \mathbf{s}^{o}) \leq 0 \\
& 0 \leq \mathbf{s}^{o} \leq \mathbf{y} + \mathbf{s} & & \mathbf{y} \leq \mathbf{D}^{t} \\
& & & \mathbf{s}^{o} \leq \mathbf{y} + \mathbf{s} \\
& & & \mathbf{y} \geq 0 \\
& & & \mathbf{s}^{o} \geq 0
\end{aligned}$$

Given a *primal problem* with the structure shown above, its *dual* is defined to be the problem shown as follows (See [8], P145):

$$\begin{aligned}
LP_N^o(\mathbf{s}, t) = \min \quad & \mathbf{v}' \cdot (\mathbf{N} - \widetilde{\mathbf{A}} \cdot \mathbf{s}) + \mathbf{w}' \cdot (\widetilde{\mathbf{A}} \cdot \mathbf{s}) + \mathbf{h}' \cdot \mathbf{D}^{t} + \mathbf{g}' \cdot \mathbf{s} \\
s.t. \quad & \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}' + \mathbf{h}' \geq \widetilde{\mathbf{f}}' \\
& -(\mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}') \geq -\widetilde{\mathbf{R}^{o}}' \\
& \mathbf{v}, \mathbf{w}, \mathbf{h}, \mathbf{g} \geq 0. \\
= \min \quad & \mathbf{v}' \cdot \mathbf{N} - (\mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}') \cdot \mathbf{s} + \mathbf{h}' \cdot \mathbf{D}^{t} \\
s.t. \quad & \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}' + \mathbf{h}' \geq \widetilde{\mathbf{f}}' \\
& -(\mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}') \geq -\widetilde{\mathbf{R}^{o}}' \\
& \mathbf{v}, \mathbf{w}, \mathbf{h}, \mathbf{g} \geq 0.
\end{aligned} \tag{4.7}$$

Let $\mathbf{u}' = \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} - \mathbf{g}'$, where $\mathbf{v}, \mathbf{w}, \mathbf{g} \geq 0$

$$\mathbf{u}' \leq \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}}$$

$$\begin{aligned} LP_N^o(\mathbf{s}, t) = \min \quad & \mathbf{v}' \cdot \mathbf{N} - \mathbf{u}' \cdot \mathbf{s} + \mathbf{h}' \cdot \mathbf{D}^t = \min \quad & \mathbf{v}' \cdot \mathbf{N} - \mathbf{u}' \cdot \mathbf{s} + \mathbf{h}' \cdot \mathbf{D}^t \\ s.t. \quad & \mathbf{u}' + \mathbf{h}' \geq \widetilde{\mathbf{f}}' \qquad\qquad\qquad s.t. \quad & \mathbf{u}' \geq \widetilde{\mathbf{f}}' - \mathbf{h}' \\ & -\mathbf{u}' \geq -\widetilde{\mathbf{R}^o}' & \mathbf{u}' \leq \min\{\widetilde{\mathbf{R}^o}', \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}}\} \\ & \mathbf{u}' \leq \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}} & \mathbf{w}, \mathbf{v}, \mathbf{h} \geq 0. \\ & \mathbf{w}, \mathbf{v}, \mathbf{h} \geq 0. & \end{aligned}$$

We know that: $\min -\mathbf{u}' \cdot \mathbf{s} \Longleftrightarrow \max \mathbf{u}' \cdot \mathbf{s}$ and $\mathbf{s} \geq 0$,

hence, $\mathbf{u}$ should be as large as possible

satisfying: $\mathbf{u}' \leq \widetilde{\mathbf{R}^o}'$ and $\mathbf{u}' \leq \mathbf{v}' \cdot \widetilde{\mathbf{A}} - \mathbf{w}' \cdot \widetilde{\mathbf{A}}$.

$\mathbf{w}' \geq 0 \Longrightarrow \mathbf{w}' \cdot \widetilde{\mathbf{A}} \geq 0$

therefore,

$$\Longrightarrow \mathbf{w}' = 0 \qquad \text{and} \qquad \mathbf{u}' = \min(\widetilde{\mathbf{R}^o}', \mathbf{v}' \cdot \widetilde{\mathbf{A}})$$

The problem is equivalent to:

$$\begin{aligned} LP_N^o(\mathbf{s}, t) = \min \quad & \mathbf{v}' \cdot \mathbf{N} - \mathbf{u}' \cdot \mathbf{s} + \mathbf{h}' \cdot \mathbf{D}^t \\ s.t. \quad & \mathbf{u}' \geq \widetilde{\mathbf{f}}' - \mathbf{h}' \\ & \mathbf{u}' = \min\{\widetilde{\mathbf{R}^o}', \mathbf{v}' \cdot \widetilde{\mathbf{A}}\} \\ & \mathbf{v}, \mathbf{h} \geq 0 \end{aligned}$$

We want $\min \mathbf{h}' \cdot \mathbf{D}^t$ and $\mathbf{D}^t \geq 0$,

so, $\mathbf{h}$ should be as small as possible satisfying:

$$\begin{aligned} \mathbf{h} &\geq 0 \\ -\widetilde{\mathbf{f}}' + \mathbf{u}' + \mathbf{h}' &\geq 0 \end{aligned}$$

then, $\quad \mathbf{h}' = (\widetilde{\mathbf{f}}' - \mathbf{u}')^+$.

Finally, its dual can be expressed as follows:

$$
\begin{aligned}
LP_N^o(\mathbf{s},t) = \min \quad & \mathbf{v}'\cdot\mathbf{N} - \mathbf{u}'\cdot\mathbf{s} + \mathbf{h}'\cdot\mathbf{D}^t & = \min \quad & \mathbf{v}'\cdot\mathbf{N} - \mathbf{u}'\cdot\mathbf{s} + (\widetilde{\mathbf{f}}' - \mathbf{u}')^+\cdot\mathbf{D}^t \\
s.t. \quad & \mathbf{h}' = (\widetilde{\mathbf{f}}' - \mathbf{u}')^+ & s.t. \quad & \mathbf{u}' = \min(\widetilde{\mathbf{R}^o}', \mathbf{v}'\cdot\widetilde{\mathbf{A}}) \\
& \mathbf{u}' = \min\{\widetilde{\mathbf{R}^o}', \mathbf{v}'\cdot\widetilde{\mathbf{A}}\} & & \mathbf{v}\geq 0. \\
& \mathbf{v}\geq 0 & &
\end{aligned}
$$

4.4 Approximate DP Algorithms

Up to this point, certain efficient mathematical programming models of the NRM problem are given. However, the IP/LP model can be treated as a computational efficient approximation to the DP Model by solving the model at several fixed time points during the booking process.

Given a certain efficient mathematical programming formulation of the NRM problem, a generic approximate DP algorithm for the NRM problem should have the following structure:

Generic Mathematical Programming (MP) Policy:

Identify an efficient formulation MP of the NRM problem.

At any current state $\mathbf{S} = (\mathbf{s}, t)$,

1. For a *ith* ODF itinerary request, compute an MP-based estimate of the opportunity cost $OC_i^{MP}(\mathbf{S})$.

2. Sell to *ith* ODF itinerary if and only if its fare f_i exceeds its opportunity cost estimate, i.e.,

$$f_i \geq OC_i^{MP}(\mathbf{S})$$

3. Go to step 1 and Iterate. ($OC_i^{MP}(\mathbf{S})$ is recomputed periodically throughout the booking horizon in response to change in $\mathbf{s}$ or demand forecast).

The difference between various algorithms comes from the approximate MP formulation and MP-based opportunity cost measure, which is from Step 1.

In this section, we will consider two approximate dynamic programming algorithms incorporating cancellations and overbooking for the NRM problem — BPC and CEC, both are derived from the LP relaxation.

1. BPC Policy incorporating Overbooking

Bid-price control (BPC) is a popular method in NRM, where the opportunity cost of an itinerary is approximated by the sum of bid-prices of the legs along that itinerary. Here, *shadow prices* are determined for each leg in the network, related to the dual ($\mathbf{v}^{\mathbf{S}}$) of the LP approximation in Eq. (4.6). We assume that the bid-prices at state equals the shadow prices obtained from the LP formulation of the problem. The leg bid-prices are computed additively. At the state $\mathbf{S} = (\mathbf{s}, t)$, the opportunity cost estimates (include the penalty vector because of the overbooking factors) are:

$$OC^{BPC}(\mathbf{s}, t) = \min\{\widetilde{\mathbf{R}^{o}}', BP^{o}(\mathbf{s}, t)\} = \min\{\widetilde{\mathbf{R}^{o}}', (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{\mathbf{A}}\}$$

[**Booking Policy for BPC:**]

At any given state $\mathbf{S} = (\mathbf{s}, t)$, accept a booking request for *ith* ODF itinerary iff:

$$\widetilde{f}_i \geq OC_i^{BPC}(\mathbf{s}, t) = \min\{\widetilde{R}_i^o, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i\}. \tag{4.8}$$

That is, the opportunity cost for *ith* ODF itinerary at the state $(\mathbf{s}, t)$ is estimated as the smaller between the sum of the bid prices of the incident legs $((\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i)$ and the overbooking penalty for that itinerary $(\widetilde{R}_i^o)$. If its adjusted fare $\widetilde{f}_i$ is higher than the opportunity cost at state $(\mathbf{s}, t)$, then the expected revenue for the airline can be increased, so we can accept this booking request in *ith* ODF itinerary at time t.

[*Note:*] From Eq. (4.6), we have: $(u_i^{\mathbf{s},t-1})' = \min\{\widetilde{R}_i^o, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i\} = OC_i^{BPC}(\mathbf{s}, t)$.

The additive bid-price approach may be the most popular technique in the current literature. However, there are several obvious drawbacks in additive bid-prices:

• The bid-prices are not well defined if there are multiple dual solutions;

• Even if the sum of the bid-prices along an itinerary are unique, they can not account for changes of dual basis which are due to large group and multi-leg itinerary requests);

• Finally, there lie the computational difficulties in reducing the amount of time required to obtain the bid-prices Hence, it is difficult for the airlines to provide near real time re-optimizations. Hence, they cannot ensure to make optimal accept/deny decisions when booking requests arrive.

[*Note:*] One useful by-product of the BPC policy is to identify flight legs which have exceptionally high bid prices. These legs correspond to bottlenecks in the airline network. Bottlenecks can constrain the flow of passengers and should be attended to. Just as in most industrial settings, the two possible treatments for bottlenecks are: 1) to increase the capacities of these legs by assigning more or larger aircraft, and 2) to maximize the use of bottlenecks by placing buffer inventories, which can be achieved with overbooking.

2. CEC Policy incorporating Overbooking

There is another different approximate estimate for the opportunity cost, which is called certainty equivalent adaptive control (CEC). The idea is to approximate the value function of the dynamic models $DP_N^o(\mathbf{s}, t)$ defined in Eq.(4.1) by the value of the linear programming problem $LP_N^o(\mathbf{s}, t)$ defined in Eq.(4.5). And we get the opportunity cost estimate in terms of LP objective values, which is uniquely determined. Then the cost estimate will not depend on the choice of dual solutions which may not be unique. So, in this way we can get a better approximation estimate.

[**Booking Control for CEC:**]

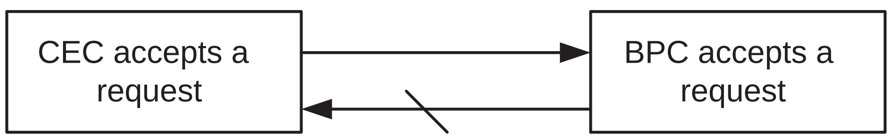
At any given state $\mathbf{S} = (\mathbf{s}, t)$, accept a booking request for *ith* ODF itinerary iff:

$$\widetilde{f}_i \geq OC_i^{CEC}(\mathbf{s}, t) = LP_{\mathbf{N}}^o(\mathbf{s}, t-1) - LP_{\mathbf{N}}^o(\mathbf{s} + e_i, t-1) \tag{4.9}$$

4.5 Structural Properties

We have introduced two approximate dynamic programming algorithms BPC and CEC in the previous section. Now let us dig out some structural properties of these two algorithms.

[***Proposition 1***]. In any state $(\mathbf{s}, t)$, for any $OC_i^{BPC}(\mathbf{s}, t)$, $OC_i^{BPC}(\mathbf{s} + e_i, t)$ the following inequalities hold: $OC_i^{BPC}(\mathbf{s}, t) \le OC_i^{CEC}(\mathbf{s}, t) \le OC_i^{BPC}(\mathbf{s} + e_i, t)$. Inequalities are strict if accepting ith ODF itinerary request must incur a change of basis in the LP dual. (We mean that any of the optimal solutions for $LP^o_{\mathbf{N}}(\mathbf{s}, t-1)$ is not the optimal solution of $LP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1)$, and vice versa.)



[**Proof:**]

$$
\begin{aligned}
OC_i^{CEC}(\mathbf{s}, t) &= LP^o_{\mathbf{N}}(\mathbf{s}, t-1) - LP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1) \\
&= (\mathbf{v}^{\mathbf{s},t-1})' \cdot \mathbf{N} - (\mathbf{u}^{\mathbf{s},t-1})' \cdot \mathbf{s} + (\widetilde{\mathbf{f}'} - (\mathbf{u}^{\mathbf{s},t-1})')^+ \cdot \mathbf{D}^{t-1} \\
&\quad -[(\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{N} - (\mathbf{u}^{\mathbf{s}+e_i,t-1})' \cdot (\mathbf{s} + e_i) + (\widetilde{\mathbf{f}'} - (\mathbf{u}^{\mathbf{s}+e_i,t-1})')^+ \cdot \mathbf{D}^{t-1}]
\end{aligned}
$$

where $(\mathbf{v}^{\mathbf{s},t-1}, \mathbf{u}^{\mathbf{s},t-1})$ and $(\mathbf{v}^{\mathbf{s}+e_i,t-1}, \mathbf{u}^{\mathbf{s}+e_i,t-1})$ are optimal dual solutions of $LP_i^o(\mathbf{s}, t-1)$ and $LP_i^o(\mathbf{s} + e_i, t-1)$, corresponding to the given bid prices: $BP_i^o(\mathbf{s}, t) = (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i$ and $BP_i^o(\mathbf{s} + e_i, t) = (\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \widetilde{A}^i$ respectively.

Since both solutions are feasible for both programs, we obtain the following upper bounds by evaluating each LP at the optimal solution of the other:

$$
LP^o_{\mathbf{N}}(\mathbf{s}, t-1) \le (\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{N} - (\mathbf{u}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{s} + (\widetilde{\mathbf{f}'} - (\mathbf{u}^{\mathbf{s}+e_i,t-1})')^+ \cdot \mathbf{D}^{t-1} \quad (4.10)
$$

$$
LP^o_{\mathbf{N}}(\mathbf{s} + e_i, t-1) \le (\mathbf{v}^{\mathbf{s},t-1})' \cdot \mathbf{N} - (\mathbf{u}^{\mathbf{s},t-1})' \cdot (\mathbf{s} + e_i) + (\widetilde{\mathbf{f}'} - (\mathbf{u}^{\mathbf{s},t-1})')^+ \cdot \mathbf{D}^{t-1} \quad (4.11)
$$

In case the two dual optimal solutions coincide, we obtain equality throughout.

By Eq. (4.10) and Eq. (4.8), we get,

$$\begin{aligned} OC_i^{CEC}(\mathbf{s},t) &\le (\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{N} - (\mathbf{u}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{s} + (\widetilde{\mathbf{f}}' - (\mathbf{u}^{\mathbf{s}+e_i,t-1})')^+ \cdot \mathbf{D}^{t-1} \\ &\quad -(\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \mathbf{N} + (\mathbf{u}^{s+e_i,t-1})' \cdot (\mathbf{s}+e_i) - (\widetilde{\mathbf{f}}' - (\mathbf{u}^{\mathbf{s}+e_i,t-1})')^+ \cdot \mathbf{D}^{t-1} \\ &= (\mathbf{u}^{\mathbf{s}+e_i,t-1})' \cdot e_i \\ &= \min\{\widetilde{R}_i^o, (\mathbf{v}^{\mathbf{s}+e_i,t-1})' \cdot \widetilde{A}^i\} = OC_i^{BPC}(\mathbf{s}+e_i,t) \end{aligned}$$

In the same way, by Eq. (4.11), we can prove that:

$$OC_i^{CEC}(\mathbf{s},t) \ge (\mathbf{u}^{\mathbf{s},t-1})' \cdot e_i = OC_i^{BPC}(\mathbf{s},t).$$

So, we prove that: $OC_i^{BPC}(\mathbf{s},t) \le OC_i^{CEC}(\mathbf{s},t) \le OC_i^{BPC}(\mathbf{s}+e_i,t)$. □

[***Proposition 2***]. (Structural Properties of the BPC Policy)

At any state $(\mathbf{s},t)$, if $LP_N^o(\mathbf{s},t-1)$ has a unique dual optimal solution, then the corresponding BPC accepts and only accept requests for which $y_i^* > 0$ in some primal optimal solution.

[**Proof:**]

At the state $(\mathbf{s},t)$, we consider the primal and the dual of $LP_N^o(\mathbf{s},t-1)$ below:

$$\begin{array}{llcll} \max & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^o}' \cdot \mathbf{s}^o & = \min & \mathbf{v}' \cdot \mathbf{N} - \mathbf{u}' \cdot \mathbf{s} + \mathbf{h}' \cdot \mathbf{D}^{t-1} \\ s.t. & \widetilde{\mathbf{A}} \cdot (\mathbf{y}+\mathbf{s}-\mathbf{s}^o) \le \mathbf{N} & s.t. & \mathbf{u}' + \mathbf{h}' \ge \widetilde{\mathbf{f}}' \\ & \widetilde{\mathbf{A}} \cdot (-\mathbf{y}-\mathbf{s}+\mathbf{s}^o) \le 0 & & \mathbf{u}' \le \widetilde{\mathbf{R}^o}' \\ & \mathbf{y} \le \mathbf{D}^{t-1} & & \mathbf{v} \ge 0. \\ & \mathbf{s}^o \le \mathbf{y}+\mathbf{s} \\ & \mathbf{y} \ge 0 \\ & \mathbf{s}^o \ge 0 \end{array} \tag{4.12}$$

we can distinguish the following situations:

• For ith ODF itinerary, $y_i^* = 0$ in all optimal LP-solutions.

We have $h_i = 0$ by complementary slackness properties.

We assume that $LP^o_N(\mathbf{s}, t-1)$ has a unique dual optimal solution at any state $(\mathbf{s}, t)$, by strict complementary slackness (See [8], P192), we know that:

$$u_i + h_i = \min(\widetilde{R}^o_i, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i) + h_i > \widetilde{f}_i$$

$$\begin{aligned} \min(\widetilde{R}^o_i, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i) &> \widetilde{f}_i \\ OC^{BPC}_i(\mathbf{s}, t) = \min(\widetilde{R}^o_i, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i) &> \widetilde{f}_i \end{aligned}$$

By the condition in Eq.(4.8), the BPC rejects this *ith* ODF itinerary request.

• For *ith* ODF itinerary, $y^*_i > 0$ in some optimal LP-solutions. We will get $0 < y^*_i \leq D^{t-1}_i$. From the complementary slackness, we have that:

$$\begin{aligned} u_i + h_i &= \widetilde{f}_i \\ \widetilde{f}_i - u_i &= h_i \geq 0 \end{aligned}$$

Clearly, we have that:

$$\widetilde{f}_i \geq u_i = \min(\widetilde{R}^o_i, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}_i) = OC^{BPC}_i(\mathbf{s}, t)$$

So, the BPC accepts this *ith* ODF itinerary request. □

[***Proposition 3***]. (Structural Properties of the CEC Policy)

Suppose that $LP^o_{\mathbf{N}}(\mathbf{s}+e_i, t-1)$ and $LP^o_{\mathbf{N}}(\mathbf{s}, t-1)$ have different optimal dual bases. And we assume that $D^{t-1}_i \geq 1$.

[*Part One:*] CEC accepts *ith* ODF itinerary request if $y^*_i \geq 1$ in some optimal solution of $LP^o_{\mathbf{N}}(\mathbf{s}, t-1)$.

[*Part Two:*] CEC rejects *ith* ODF itinerary request if $0 \leq y^*_i < 1$ in all optimal solutions of $LP^o_{\mathbf{N}}(\mathbf{s}, t-1)$.

[**Proof:**]

For Part One, we assume that: $y^*_i \geq 1$ in some optimal solution of $LP^o_{\mathbf{N}}(\mathbf{s}, t-1)$.

Hence, we have $\mathbf{y}^* - e_i \geq 0$.

$$\begin{aligned} LP_N^o(\mathbf{s}, t-1) = \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y}^* - \widetilde{\mathbf{R}^o}' \cdot (\mathbf{s}^o)^* \\ \text{where, } (\mathbf{y}^*, (\mathbf{s}^o)^*) \text{ satisfying} \quad & 0 \leq \widetilde{\mathbf{A}} \cdot (\mathbf{y}^* + \mathbf{s} - (\mathbf{s}^o)^*) \leq \mathbf{N} \\ & 0 \leq \mathbf{y}^* \leq \mathbf{D}^{t-1} \\ & \mathbf{y}^* - e_i \geq 0 \\ & 0 \leq (\mathbf{s}^o)^* \leq \mathbf{y}^* + \mathbf{s} \end{aligned}$$

$\Longrightarrow (\mathbf{y}^*, (\mathbf{s}^o)^*)$ satisfying:

$$\begin{aligned} & 0 \leq \widetilde{\mathbf{A}} \cdot (\mathbf{y}^* - e_i + \mathbf{s} + e_i - (\mathbf{s}^o)^*) \leq \mathbf{N} \\ & 0 \leq \mathbf{y}^* - e_i \leq \mathbf{D}^{t-1} \\ & 0 \leq (\mathbf{s}^o)^* \leq \mathbf{y}^* - e_i + \mathbf{s} + e_i \end{aligned}$$

So, $(\mathbf{y}^* - e_i, (\mathbf{s}^o)^*)$ is a feasible solution of $LP_{\mathbf{N}}^o(\mathbf{s} + e_i, t-1)$, where,

$$\begin{aligned} LP_N^o(\mathbf{s} + e_i, t-1) = \max \quad & \widetilde{\mathbf{f}}' \cdot \mathbf{y} - \widetilde{\mathbf{R}^o}' \cdot \mathbf{s}^o \\ s.t. \quad & 0 \leq \widetilde{\mathbf{A}} \cdot (\mathbf{y} + \mathbf{s} + e_i - \mathbf{s}^o) \leq \mathbf{N} \\ & 0 \leq \mathbf{y} \leq \mathbf{D}^{t-1} \\ & 0 \leq \mathbf{s}^o \leq \mathbf{y} + \mathbf{s} + e_i \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad LP_{\mathbf{N}}^o(\mathbf{s}, t-1) &= \widetilde{\mathbf{f}}' \cdot \mathbf{y}^* - \widetilde{\mathbf{R}^o}' \cdot (\mathbf{s}^o)^* \\ &= \widetilde{f}_i + \widetilde{\mathbf{f}}' \cdot (\mathbf{y}^* - e_i) - \widetilde{\mathbf{R}^o}' \cdot (\mathbf{s}^o)^* \\ &\leq \widetilde{f}_i + LP_{\mathbf{N}}^o(\mathbf{s} + e_i, t-1) \end{aligned}$$

$$\Longrightarrow \quad OC_i^{CEC}(\mathbf{s}, t) = LP_{\mathbf{N}}^o(\mathbf{s}, t-1) - LP_{\mathbf{N}}^o(\mathbf{s} + e_i, t-1) \leq \widetilde{f}_i$$

CEC will accept *ith* ODF itinerary request in this case.

Then we proved the part one.

For part two, $0 \leq y_i^* < 1$ in all primal optimal solutions, then $0 \leq \mathbf{y}^* < D_i^{t-1}$.

By complementary slackness, $h_i = 0$ and $u_i + h_i \geq \widetilde{f}_i$.

That is: $\min(\widetilde{R}_i^o, (\mathbf{v}^{s,t-1})' \cdot \widetilde{A}^i) \geq \widetilde{f}_i$

Under the assumption that the optimal dual basis changes, and by Proposition 1:

$$OC_i^{CEC}(\mathbf{s},t) > OC_i^{BPC}(\mathbf{s},t) = \min(\widetilde{R}_i^o, (\mathbf{v}^{\mathbf{s},t-1})' \cdot \widetilde{A}^i) \geq \widetilde{f}_i$$

So, CEC policy rejects ith ODF itinerary request. □

∗ Hence, if we have the assumptions as follows:

1. At any state $(\mathbf{s},t)$, if $LP_N^o(\mathbf{s},t-1)$ has a unique dual optimal solution.
2. $LP_{\mathbf{N}}^o(\mathbf{s}+e_i,t-1)$ and $LP_{\mathbf{N}}^o(\mathbf{s},t-1)$ have different optimal dual bases.
3. $D_i^{t-1} \geq 1$.

the behavior of the BPC and CEC policies can be characterized as a function of y_i^*.

y_i^*	$y_i^* \geq 1$ for some $\mathbf{y}^*$	$y_i^* < 1$ for all $\mathbf{y}^*$, but $y_i^* \neq 0$ in some i	$y_i^* = 0$ for all $\mathbf{y}^*$
CEC	accept	reject	reject
BPC	accept	accept	reject

4.6 Computational Performance

The aim of this section is to understand the relative performance of BPC and CEC in the environment with cancellations and overbooking. Exact calculations of the optimal expected revenue (DP) and the expected values of the proposed policies (CEC, BPC) are practically impossible. A tractable approach for measuring performance of the proposed policies, however, is provided by simulation. We consider a booking horizon of 15 periods for a hub and spoke network with 5 cities and two classes, as in the following example from [7]:

The leg-class incidence matrix, together with a high-low fare structure $f = (\mathbf{f}^h, \mathbf{f}^l)$:

$$\begin{pmatrix} f \\ A \end{pmatrix} = \begin{pmatrix} f_1^l & f_1^h & f_2^l & f_2^h & f_3^l & f_3^h & f_4^l & f_4^h & f_{12}^l & f_{12}^h & f_{13}^l & f_{13}^h & f_{23}^l & f_{23}^h & f_{24}^l & f_{24}^h \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix}$$

The arrival process for the highest fare class is nonhomogeneous Poisson with rate 0.5 for period 1-13 and 5 for periods 14 and 15. The arrival process for the lowest fare class is homogeneous Poisson with rate 3. For simplicity, keep that fare of the higher class in a single-leg itinerary equal to \$100 and of the lower class equal to \$80. Vary the fare of two leg itineraries.

After experimentation, some results are presented below:

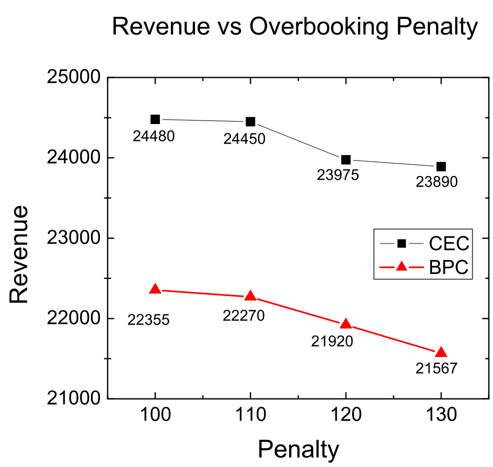


Figure 4.1: The expected revenue in $T = 200$ simulation runs as a function of the overbooking penalty. The cancellation probability was 0.01.

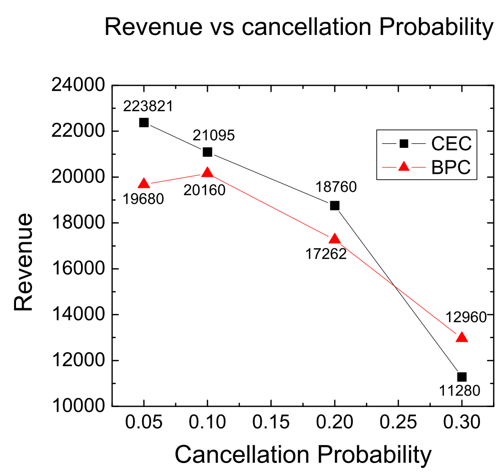


Figure 4.2: The expected revenue in $T = 200$ simulation runs as a function of the overbooking penalty. The overbooking penalty was \$130.

[**Result 1:**] By comparing the behaviors of CEC and BPC as a function of the overbooking penalty, the CEC algorithm leads to higher revenue by approximately 1%, see Figure (4.1).

[**Result 2:**] For the behaviors of CEC and BPC as a function of the cancellation probability, with the exception of very high cancellation rate (0.3), CEC outperforms BPC still, see Figure (4.2).

4.7 Conclusion

Network-based models consider booking requests for multiple legs at the same time. The booking control policy can be static, such that decisions are based on pre-calculated booking limits, or dynamic, where the decision rules constantly change during the booking period. In this chapter, we investigate dynamic policies for allocating inventory to stochastic demand of multiple fare classes, in a network environment so as to maximize total expected revenue. Although in principle it is not hard to come up with a practical, dynamic model in a network environment, the proverbial "curse of dimensionality" of dynamic programming would prohibit solving it to optimality in practice. In this chapter, we consider IP and LP Models as the approximation schemes by solving the models at several fixed times during the booking process. Both models incorporate cancellation, no-shows and overbooking, though these schemes considers only the expected demand and ignores all distributional information. Furthermore, based on the LP models, two approximate dynamic programming algorithms (BPC and CEC) are introduced. We find out and prove some structural properties and compare the behaviors of both policies. Finally, we quote one computational result to give insight into the performance of the algorithms. With the importance of overbooking control to airline profitability, emphasis is being placed on finding more effective methods and better solutions so as to maximize the expected revenue for the airline in the network environment.

Chapter 5

Conclusion and Future Work

The emphasis of this report has been on the overbooking in airline revenue management. Both leg-based and network-based overbooking control approaches have been developed and evaluated.

We provide an introduction for Airline Overbooking problem and state different models in use today. The booking control policy for these models can be static or dynamic. We present the difference between Static and Dynamic models in detail.

In this report, we present several new approaches to solve static Single-Leg problem with cancellations, no-shows and overbooking. We deduce 3 different rules for Single-fare, Single-Leg overbooking problem to determine the optimal booking limits for each fare class at the start of the booking process. For the two-fare-class on Single-Leg, Littlewood's Rule is revised to include overbooking for two nested fare classes case. The revised rule has provided the optimal protection level for the higher fare class from the lower one. We present a heuristic to the nested, Single-Leg problem with multiple fare classes (rEMSR approach). Given all these protection levels in the nested reservation system, we propose simple example to show that how the optimal booking limits are determined to maximize the total expected revenue.

For these static models on Single-Leg, the structure of the optimal booking control is such that the decision is determined by the booking limits for each fare class. The results generated by these solutions are optimal under the sequential arrival assumption as long as no change in the probability distributions of the demand is foreseen. So, it is necessary to guarantee that the fare demands are accurately forecasted. At present, further research is still needed to improve the forecasting models.

However, passenger behavior, and therefore the optimal booking limits, vary with time to departure. We discussed two models in [54] to formulate the overbooking problem as a discrete time Markov decision process and use dynamic programming to analyze it. These two dynamic models on Single-Leg are discussed in depth. We prove the concavity of the value function completely, which is the key point for the optimality of the booking control policies. The structure of the optimal booking control is such that the decision is determined by the opportunity cost of reducing one seat in capacity. The optimal decision rule can be determined at each time period before flight departure. However, the size and complexity of these theoretical models make them impractical for an airline to use routinely in its reservations control system.

We have also studied several models and algorithms for solving the dynamic Network Revenue Management problem (NRM). We find out two approximate dynamic algorithms — BPC and CEC which are discussed by Bertsimas and Popescu (2001) [7]. However, [7] focuses on the models without considering the overbooking, cancellations and no-shows. In this report, all these random and complicated processes are handled by incorporating oversales decisions in the underlying linear programming formulation. We provide and prove some structure properties that compare the behavior of the proposed CEC policy with the BPC approach. These results offer insight into the behavior of both methods.

The proverbial "curse of dimensionality" of dynamic methods would still prohibit implementing them in practice. The most frequently utilized methods to the NRM are static models. They are solved at several fixed times during the booking process. We consider the LP approximation methods, which are appealing for large-scale network revenue management because they can be very efficiently solved each time a re-optimization is required. However, we haven't considered the nested inventory structure which will generate equal or greater expected revenue for the airlines [62].

The Overbooking problem in Airline Revenue Management is greatly complicated. The airline's practices, the size of the aircrafts, the overall route structure of an airline, the demand densities, the cancellations process of the passengers, the no-show activities and even the competitive environment in which the airline operates all dictate the specific approach which is best for a particular airline. It has received a lot of attention throughout the past years and will continue to be of interest. In the near future, one can expect to see more research work done on overbooking in airline revenue management, and finally, considering the full integration of overbooking with pricing and traffic management [52], rather than treating each as separate function of the marketing process.

Appendix A

Useful Terminology

We provide here a glossary of terminologies in Overbooking in airline revenue management. The aim is to supply a separate glossary to avoid needless definitions for readers familiar with this problem while assisting others who are new to this field. Many of the terms described here have different meanings in more general contexts but are presented here with their usual meanings in Airline Overbooking problem.

For convenience, we will use some abbreviations.
We state them as follows:

AGIFORS	Airline Group of the International Federation of Operational Research Societies
BPC	Bid-price Control Algorithms
CAB	Civil Aeronautics Board
DINAMO	Dynamic Inventory and Maintenance Optimizer
EMSR	Expected Marginal Seat Revenue
IP	Integer Programming
LP	Linear Programming
MP	Mathematical Programming Formulation
NRM	Network Revenue Management

OBL	Optimal Booking Limits
ODF	Origin-Destination-Fare
PARM	Perishable Asset Revenue Management
SABRE	Semi-Automated Business Research Environment

Arrival pattern: The pattern of arrivals of booking requests. In the airline context, some possible arrival patterns are: sequential booking classes, low-before-high fares, or interspersed arrivals.

Batch booking: A booking request that arrives through normal reservation channels for two or more seats to be booked for the same itinerary. Contrast with group bookings.(Also multiple booking, or bulk arrival)

Bid Price: A net value (bid-price) for an incremental seat on a particular flight leg in the airline network. Also referred to as minimum acceptable fare, hurdle price, probabilistic shadow price, displacement cost, or probabilistic dual cost.

Bid price control: A method of network seat inventory control that assesses the value of an ODF itinerary as the sum of the bid-prices assigned to individual legs in the itinerary. Typically, an ODF request is accepted if its fare exceeds the total bid-prices.

Booking class: A category of bookings that share common features (e.g., similar revenue values or restrictions) and are controlled as one class. This term is often used interchangeably with fare class or bucket.

Booking limit: The maximum number of seats that can be sold to a particular booking class. In nested booking systems, booking limits apply to the total number of seats sold to a particular booking class and any lower fare booking classes.

Booking Policy: A booking policy is a set of rules that specify at any point during the booking process whether a booking class should be open. In general, such policies may depend on the pattern of prior demands or be randomized in some manner and must be generated dynamically as the booking process unfolds for each flight. In some circumstances, optimal or approximately optimal booking policies can be defined by a set of fixed *protection levels* or *threshold curves.*

Cancellations: Returns or changes in bookings that occur early enough in the booking period to permit subsequent re-booking through the reservations system.

Compensation costs or Cost of Oversales: The total value of money and other incentives given to bumped passengers by airlines. It may consist of compensation for the inconvenience (in the form of vouchers that can be redeemed on a future flight), hotel and meal accommodations if necessary, and accommodation on a later flight, either on same airline or some other airline. The oversale cost is not constant (nonlinear with a positive slope as the number of oversales increase).

Denied boarding: Turning away ticketed passengers when more passengers show-up at flight time than there are seats available on the flight, usually as a result of overbooking practices. Denied boarding can be either voluntary, when passengers accept compensation for waiting for a later flight, or involuntary, when an insufficient number of passengers agree to accept compensation. In the latter case, the airline will be required to provide compensation in a form mandated by civil aviation law.

Dynamic models: Models that take into account future possible booking decisions in assessing current decisions. Most revenue management problems are properly modelled as dynamic programming problems.

Dual prices: The marginal value of one additional unit of a constrained resource, as determined by a mathematical programming solution to an optimization model. Dual prices are one source of the marginal seat values used in bid-price control.

Expected marginal seat revenue (EMSR): The expected revenue of an incremental seat if held open. This is a similar concept to that of bid-price but generally used in a simpler context.

Flight leg: A section of a flight involving a single takeoff and landing (or no boarding or deplaning of passengers at any intermediate stops).

Flight Capacity: The total number of physical seats on a given flight.

Goodwill costs: An airline's rejection of a booking request can affect a customer's propensity to seek future bookings from that airline. This cost is difficult to assess but is considered particularly acute in competitive markets and with customers who are frequent air travellers.

Go-show: Passengers who appear at the time of flight departure with a valid ticket for the flight but for whom there is no record in the reservation system. This no-record situation can occur when there are significant time lags in transferring booking information form reservations sources (e.g., travel agent's offices) to the CRS or when there are transmission breakdowns.

Group bookings: Bookings for groups of passengers that are negotiated with sales representatives of airlines; for example, for a large group from one company travelling to a trade show. These should be distinguished from batch bookings.

Hub-and-spoke network: A configuration of an airline's network around one or more major hubs that serve as switching points in passengers' itineraries to spokes connected to smaller centers. The proliferation of these networks has greatly increased the number of passenger itineraries that include connections to different flights.

Independence of demands: The assumption that demands in one customer category (e.g., booking class or ODF) are statistically independent of demands in other categories. It is widely believed that this assumption is not satisfied in practice. See, for example, Hopperstad (1994).

Itinerary: For purposes of this report, an itinerary is a trip from an origin to a destination across one or more airline networks. A complete specification of an itinerary includes departure and arrival times, flight numbers, and booking classes. The term is used ambiguously to include both one-way and round-trip travel. That is, used in the first way, a round-trip involves two itineraries and, in the second way, one itinerary.

Leg based control: An old, but still common method of reservations control and revenue management in which limits are set at the flight leg level on the number of passengers flying in each booking class. Such systems are unable to properly control multi-leg traffic, although virtual nesting provides a partial solution.

Load factor: The ratio of seats filled on a flight to the total number of seats available.

Low-before-high fares: (Also called monotonic fares or sequential fares) The sequential booking class assumption is often augmented by the additional assumption that booking requests arrive in strict fare sequence, generally from lowest to highest as flight departure approaches.

Multi-leg: A section of an itinerary or network involving more than one leg.

Nested booking: In fully nested (also called serially nested) booking systems, seats that are available for sale to a particular booking class are also available to bookings in any higher fare booking class, but not the reverse. Thus, a booking limit B for a discount booking class defines an upper bound on bookings in that class and any lower valued classes and a corresponding protection level for all higher classes. This should be contrasted with the partitioned booking system.

No-shows: Booked passengers who fail to show up at the time of flight departure, thus allowing no time for their seat to be booked through normal reservations processes. No-shows are particularly common among full fare passengers whose tickets are fully refundable in the event of cancellation or no-show.

Opportunity cost: In revenue management, the opportunity cost of a booking includes all future revenues that may be lost if the booking is accepted. Taken to the extreme, these include the revenue value of potential displaced future bookings anywhere in the airline network and goodwill costs from those displacements. Assessment of the costs and probabilities of such displacements should allow for the dynamics of cancellations and overbooking and the expected costs of oversold conditions.

Oversold: An ambiguous term sometimes used when more passengers show up for a flight than there are seats available. Such situations must be resolved with denied boarding.

Overbooking: The practice of ticketing seats beyond the capacity of an aircraft to offset the effects of passenger cancellations and no-shows.

Protection levels: The total number of protected seats for a booking class. In fully nested booking systems the protection level for a fare class applies to that class and all higher fare classes.

Recapture: The booking of a passenger who is unable to obtain a reservation for a particular flight or set of flights with an airline onto alternative flights with the same airline. High recapture probabilities imply that less oversale risk should be taken, so that the overbooking level will be lower.

Revenue Management: The practice of controlling the availability and pricing of the seats in different booking classes with the goal of maximizing expected revenues or profits. This term has largely replaced the original term *yield management*.

Show-ups: Passengers who appear for boarding at the time of flight departure. The total number of show-ups is = final bookings + go-shows + standbys - no-shows.

Spoilage: Seats that travel empty despite the presence of sufficient demand to fill them. This should be distinguished from excess capacity — seats that are empty because of insufficient total demand. Spoilage therefore represents a lost-opportunity cost to the airline.

Standby fares: Some airlines will sell last minute discount seats to certain categories of travellers (e.g., youth or military service personnel) who are willing to wait for a flight that would otherwise depart with empty seats. In other words, standbys are customers who buy tickets at (possibly reduced) rates with the restriction that they may travel on the next flight with available seats only after all reservations for that flight have been honored.

Static models: Models that set current booking limits without consideration of the possibility of adjustments to the limits later in the booking process. (Compare with dynamic models.)

Ticket holders or Ticketed passengers: People who have purchased a ticket and whose individual ticket revenue has already been received by the airline.

Threshold times: They are points in time during the booking horizon before which requests are rejected, and after which requests are accepted.

Upgrade: This term is used in two ways. Firstly, it refers to an offer to a passenger to fly in a higher service class without additional charge (e.g., in exchange for frequent flyer points, or to avoid a denied boarding). Secondly, it refers to a decision by a customer to book in a higher fare class than originally intended when he or she is advised that no seats are available at their preferred fare (**Sell-ups**).

Virtual nesting/virtual classes: This is one approach to incorporate origin-destination information into leg or segment based control systems. Multiple ODFs are grouped into virtual buckets on the basis of similar revenue characteristics (e.g., comparable total fare classes.) The buckets are then nested and assigned to traditional booking classes for control in a leg based reservation system.

Yield management: The early term used for what is now more commonly called revenue management. Cross (1995) attributes the original term to Robert L. Crandal when he was Senior Vice President for Marketing (Later CEO) at American Airlines.

Appendix B

Literature Review

Airline Revenue Management research has been reduced to three distinct smaller problems: Overbooking, Discount allocation, and Traffic management (Smith and etc. [52]). Overbooking has the longest research history of any of the components of the revenue management problem. For convenience to the readers, we collect and outline some important results as follows. You can also check the references.[1]

Year	**Reference**	**Main Contributions**
1958	Beckmann [3]	An early, non-dynamic optimization model is formed for overbooking, employing a static one-period model with reservation requests, booking and finally cancellations (in term of τ distributions) that balances the lost revenue of empty seats with the costs to the airline of passengers denied boarding.

[1] These papers are important in this field, we collect them to get a rough idea about what is going on for the Airline Overbooking problem. As I have said, overbooking is a strategic tool to increase corporate profitability and most airlines generally do not publish their yield management approaches, models and implementation aspects due to their proprietary nature. So we tried my best to find those papers published as possible as we can. Some incorporate overbooking, some do not, while they are very crucial for the development of this problem.

Year	Reference	Main Contributions
1960	Kosten [31]	A continuous time approach is developed. However, this approach requires solution of a set of simultaneous differential equations which make the implementation impractical. He provided the interspersion of reservations and cancellations which Beckmann ignored and thus yielded a booking level depending upon the number of days yet to transpire before flight.
1961	Thompson [59]	A model is developed to provide booking levels that constrained the probability of denied boarding. Different from Beckmann's and Kosten's models, this model omits the costs and passenger reservation demands, and only describes the cancellation patterns of any fixed number of reserved passengers.
1962	Taylor [58]	He formed a statistical model(Adapted Thompson's approach).
1964	Deetman [20]	Taylor's model was studied to test its behavior and implementability at KLM.
1967	Rothstein and Stone [47]	They formed a model for Single-leg flight carrying a single type of passengers, developed a computer system for booking levels using a slightly simplified version of the Taylor model and capitalizing on the copious cancellation statistics available from SABRE.
1968	Rothstein [43]	He described the first dynamic programming(DP) model for overbooking and reviewed the results of test runs of the model at American Airlines.
1971	Rothstein [44]	The procedure of reservations was viewed as a Markovian sequential decision process. He first proposed a mathematic model to analyze the overbooking policy.

Year	Reference	Main Contributions
1971	Howard [27]	The airline overbooking problem is set up for a single fare class as a Markov decision problem. Howard proposed the use of the value iteration method to obtain the optimal policy for the problem of overbooking. However, only very small problems can be solved with this approach because of the computational limitations of value iteration.
1972	Vickrey [60]	He claimed that oversold conditions could be resolved with auctions and he describe a concept of the multiple fare classes reservations system.
1972	Littlewood [39]	He proposed a rule for the two-period, two-fare-class problem in which low-fare customers book prior to high-fare customers. It is an important paper even though it ignores cancellations, no-shows and overbooking. His rule was shown to be optimal by Bhatia and Parekh(1973) of TWA, and later by Richter(1982) of Lufthansa.
1974	Etschmaier and Rothstein [23]	They formulated the airline and hotel overbooking problem as a non-homogeneous markovian sequential decision process. Solutions to the formulations were obtained with the aid of dynamic programming.
1975	Shlifer and Vardi [49]	An overbooking model is extended to allow for two fare classes and a two-leg problem is described. A model was presented to determine overbooking levels under three different criteria assuming deterministic capacity of the three criteria chosen.

Year	Reference	Main Contributions
1977	Simon and Visvabhanathy [50]	They came up with a remarkable proposal for solving the overbooking problem: if too many reserved passengers show up at flight time, the airline agents should conduct an auction among them.
1978	Hersh and Ladany [36]	They considered a flight with one class and one intermediate stop. In both effects, a sequential decision process was developed which incorporated the time distribution at which reservations and cancellations were actually made, as well as effects due to waitlisted and standby passengers and overbooking.
1985	Rothestein [46]	He presented a survey of the application of operations research to airline overbooking. The article analyzed the issues that motivated overbooking and discussed the relevant practices of the air carriers.
1986	Alstrup et al. [2]	A DP treatment of overbooking for a two-class, non-stop flight was described. The model treats the airline booking process as a Markovian non-homogeneous sequential decision process. They stated computational experience with the approach at Scandinavian Airlines(solved by two-dimensional stochastic dynamic programming).
1987a,b	Belobaba [5]	He discussed the problem of overbooking in multiple fare classes and suggested a heuristic approach to solve the problem.
1988	Dror et al. [22]	A basic network model with gains/losses on certain arcs for seats allocated to a single flight with intermediate stops is first presented. (A network flow representation of the problem incorporating both cancellations and no-shows).

Year	Reference	Main Contributions
1989	Brumelle and McGill [12]	He presented a static formulation of the overbooking problem and showed that it was a special case of a general model of the two fare class seat allocation problem.
1989	Simpson [51]	He introduced the idea of bid-price controls and proposed many of the main approximation approaches in the area.
1992	Williamson [62]	Similar to Simpson(1989) and in particular, she used extensive simulation studies to analyze a variety of approaches to network revenue management.
1992	Smith et al. [52]	American Airlines Decision Technologies developed a series of OR models and implemented the static one-period overbooking model with additional constraints to ensure that the level of service was not overly degraded. A brief discussion on overbooking was presented in the article.
1992	Bodily and Pfeifer [10]	They worked on the static single-leg overbooking problem and stated the general overbooking rule, and adapted it for specific models of the random survival process for reservations.
1993	Chatwin [15]	He dealt exclusively with the overbooking problem and provided a number of new structural results. A rigorous treatment of the multi-period overbooking problem that relates to a single flight leg with known capacity and single service class is provided. A continuous time version of the model with stationary fares and refund is also presented.

Year	Reference	Main Contributions
1993	Weatherford, Bodily and Pfeifer [61]	They investigated dynamic booking limits for two classes of passengers whose booking requests arrive concurrently, assuming that the distribution of remaining demand for each fare class was known.
1993	Lee and Hersh [38]	They considered a discrete time dynamic programming model, where demand for each fare class was modelled by a non-homogeneous Poisson process. They proposed a practical issue in airline seat inventory control (without overbooking).
1993	Curry [19]	A simple and easy to understand discussion on overbooking in revenue management is provided, and a couple of models for solving the overbooking problem are presented.
1998	Chatwin [16]	He analyzed a multi-period airline overbooking problem with non-stationary fares.
1998	Karaesman and Van Ryzin [30]	They addressed the problem of jointly setting overbooking levels when there were multiple inventory classes that could serve as substitutes for one another.
1999a,b	Chatwin [17][18]	He modelled customer cancellations, and no-shows in a dynamic framework. He was the first to take advantage of the properties that TP3(totally positive of order 3) density functions preserve quasi-concavity and concavity in order to prove results regarding the structure of optimal policies. Gave conditions that ensure the intuitive result that a booking-limit policy was optimal.

Year	Reference	Main Contributions
1999	Subranmanian, Lautenbacher and Stidham [54]	They allowed for bookings in multiple fare classes as well as cancellations and no-shows. Borrowing a result from the queueing control literature, they proved the concavity of the associated optimal value functions and subsequently, the optimality of a booking limit policy.
2000	Ignaccolo and Inturri [29]	A fuzzy approach to the overbooking problem in air transportation is considered.
2000	Zhao and Zheng [64]	They proved that a similar threshold control as Fend and Xiao (1999,2000) was optimal for a more general airline seat allocation model that allowed diversion/upgrade and no-shows. They also showed that under certain conditions, the optimal threshold may not be monotone.
2001	Feng, Lin and Xiao [26]	They formulated the airline seat control problem with cancellations into a continuous-time, stochastic revenue management model. They showed that optimal seat control was of the thresholds built upon the characteristic minimum acceptable fare.
2001	Dimitris Bertsimas and Ioana Popescu [7]	They investigated dynamic policies for allocating scarce inventory to stochastic demand from multiple fare classes, in a network environment so as to maximize total expected revenues. They proposed and analyzed a new algorithm, based on approximate dynamic programming and extended that to handle cancellations and no-shows by incorporating oversales decisions in the underlying linear programming formulation.
2002	Suzuki [55]	The behaviors of the denied-boarding passengers after they were bumped are investigated.

Bibliography

[1] American Airlines, *Overbooking and the Public Interest: An American Airlines Perspective*, 1987.

[2] J. Alstrup, S. Boas, O. Madsen, and R. Vidal, *Booking Policy for Flights with Two Types of Passengers*, Eur. J. Opnl. Res. **27** (1986), 274–288.

[3] M. J. Beckmann, *Decision and Team Problem in Airline Reservations*, Econometrica **26** (1958), 134–145.

[4] P. P. Belababa, *Air Travel Demand and Airline Seat Inventory Management*, Ph.D. thesis, MIT, Cambridge, Mass, 1987.

[5] ———, *Application Yield Management: An Overview of Seat Inventory Control*, Trans. Sci. **21** (1987), 63–73.

[6] ———, *Application of a Probabilistic Decision Model to Airline Seat Inventory Control*, Opns. Res. **37** (1989), 183–197.

[7] D. Bertsimas and I. Popescu, *Revenue Management in a Dynamic Network Environment*, 2001.

[8] D. Bertsimas and J. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA, 1999.

[9] G. R. Bitran and S. V. Mondschein, *An Application of Yield Management to the Hotel Industry Considering Multiple Day Stays*, Opns. Res. **43(3)** (1995), 427–443.

[10] S. E. Bodily and P. E. Pfeifer, *Overbooking Decision Rules*, Omega **20** (1992), no. 1, 129–133.

[11] S. Boer, R. Freling, and N. Piersma, *Mathematical Programming for Network Revenue Management revisited*, European Journal of Operational Research **137** (2002), 72–92.

[12] S. L. Brumelle and J. I. McGill, *A General Model for Airline Overbooking and Two-class Revenue Management with Dependent Demands*, 1989.

[13] ______, *Airline Seat Allocation with Multiple Nested Fare Classes*, Opns. Res. **41** (1993), 127–137.

[14] ______, *Airline Seats Allocation with Multiple Nested Fare Classes*, Opns. Res. **41** (1993), 127–137.

[15] R. E. Chatwin, *Optimal Airline Overbooking*, Ph.D. thesis, Stanford University, Palo Alto, CA, 1993.

[16] ______, *Optimal Control of Continuous-Time Terminal-Value Birth-and-Death Processes and Airline Overbooking*, Naval Research Logistics **43** (1996), 159–168.

[17] ______, *Continuous-Time Airline Overbooking with Time Dependent Fare and Refunds*, Trans. Sci. **33** (1999a), 182–191.

[18] ______, *Multiperiod Airline Overbooking with a Single Fare Class*, Opns. Res. **46** (1999b), 805–819.

[19] R. E. Curry, *Overbooking in Revenue Management*, 1993.

[20] C. Deetman, *Booking Levels*, Proc. the Fourth AGIFORS Symposium (1964), 210–235.

[21] M. Diamond and R. Stone, *Dynamic Yield Management on a Single Leg*, 1991.

[22] M. Dror, P. Trudeau, and S. P. Ladany, *Network Models for Seat Allocation on Flights*, Trans. Res. **22(B)** (1988), 239–250.

[23] M. M. Etschmaier and M. Rothstein, *Operations Research in the Management of the Airlines*, Omega **2** (1974), 193–204.

[24] B. Eyal, C. Ziv, E. F. Gila, and G. Eitan, *Overselling with Opportunistic Cancellations*, Marketing Science **18** (1999), no. 4, 0605–0610.

[25] Y. Feng and G. Gallego, *Optimal Starting Times for Promotional Fares and Optimal Stopping Times for End-of-season Sales*, Mgmt. Sci. **41** (1995), 1371–1391.

[26] Y. Y. Feng, P. Lin, and B. C. Xiao, *An Analysis of Airline Seat Control with Cancellations*, Working Paper (2001).

[27] A. Gosavi, N. Bandla, and T. K. Das, *A Reinforcement Learning Approach to a Single Leg Airline Revenue Management Problem with Multiple Fare Classes and Overbooking*, IIE Transactions.

[28] D. Gunther, V. Chen, and E. Johnson, *Airline Yield Management: Optimal Bid Prices for Single-hub Problems without Ccancellation*, 1998.

[29] M. Ignaccolo and G. Inturri, *A Fuzzy Approach to Overbooking in Air Transportation*, J. Air Transp. World Wide **5** (2000), 19–38.

[30] I. Karaesman and G. J. van Ryzin, *Overbooking with Substitutable Inventory Classes*, 1998.

[31] L. Kosten, *Een Mathematisch Model Voor Een Reserveringsporbleem*, Statistica Neerlandica **14** (1960), 85–94.

[32] S. P. Ladany, *Dynamic Operating Rules for Motel Reservations*, Dec. Sci. **7** (1976), 829–840.

[33] ______, *Bayesian Dynamic Operating Rules for Optimal Hotel Reservations*, Zeitschrift für Opns. Res. **21** (1977), 165–176.

[34] S. P. Ladany and D. N. Bedi, *Dynamic Booking Rules for Flights with an Intermediate Stop*, Omega **5** (1977), 721–730.

[35] S. P. Ladany and M. Hersh, *Non-stop vs. One-stop flights*, Trans. Res. **11** (1977), 155–159.

[36] ______, *Optimal Seat Allocation for Flights with one Intermediate Stop*, Computers and Opns. Res. **5** (1978), 31–37.

[37] C. J. Lautenbacher and S. J. Stidham, *The Underlying Markov Decision Process in the Single-Leg Airline Yield Management Problem*, Trans. Sci. **33** (1999), 136–146.

[38] T. C. Lee and M. Hersh, *A Model for Airline Seat Inventory Control with Multiple Seat Bookings*, Transp. Sci. **27** (1993), 1252–1265.

[39] K. Littlewood, *Forcasting and Control of Passengers*, In Proceedings 12th AGIFORS symposium (1972), 95–117.

[40] J. I. McGill, *Optimization and Estimation Problems in Airline Yield Management*, Ph.D. thesis, Commerce and Business Administration Program, University of British Columbia, Canada, 1989.

[41] H. Raiffa and R. Schlaifer, *Applied Statistical Decision Theory*, 1961, p. 295.

[42] L. W. Robinson, *Optimal and Approximate Control Policies for Airline Booking with Sequential Nonmonotonic Fare Classes*, Opns. Res. **43** (1995), 252–263.

[43] M. Rothstein, *Stochastic Models for Airline Booking Policies*, Ph.D. thesis, Graduate School of Engineering and Science, New York University, New York, 1968.

[44] ———, *An Airline Overbooking Model*, Trans. Sci. **5** (1971b), 180–192.

[45] ———, *Hotel Overbooking as a Markovian Sequential Decision Process*, Dec. Sci. **5** (1974), 389–404.

[46] ———, *Operations Research and the Airline Overbooking Problem*, Opns. Res. **33** (1985), 237–248.

[47] M. Rothstein and W. Stone, *Passenger Booking Levels*, Proc. Seventh AGIFORS Symposium (1967), 292–435.

[48] M. Shaked and J. G. Shanthikumar, *Stochastic Orders and Their Applications*, 1994.

[49] R. Shlifer and Y. Vardi, *An Airline Overbooking Policy*, Trans. Sci. **9** (1975), 101–114.

[50] J. L. Simon and G. Visvabhanathy, *The Auction Solution to Overbooking - the Data Fit the Theory*, J. Transp. Econ. Policy (1977).

[51] R. W. Simpson, *Using Network Flow Techniques to Find Shadow Prices for Market and Inventory Control*, 1989.

[52] B. C. Smith, J. F. Leimkuhler, and R. M. Darrow, *Yield Management at American Airlines*, Interfaces **22** (1992), 8–31.

[53] S. Stidham, *Socially and Individually Optimal Control of Arrivals to a GI/M/1 Queue*, Mgmt. Sci. **24** (1978), 1598–1610.

[54] J. Subramanian, C. J. Lautenbacher, and S. J. Stidham, *Yield Management with Overbooking, Cancellations and No Shows*, Trans. Sci. **33** (1999), 147–167.

[55] Y. Suzuki, *An Empirical Analysis of the Optimal Overbooking Policies for US Major Airlines*, Trans. Sci. **38** (2002), no. Part E, 135–149.

[56] K. T. Talluri and G. J. van Ryzin, *A Randomized Linear Programming Method for Computing Network Bid Prices*, Trans. Sci. **33** (1999a), 207–216.

[57] ———, *An Analysis of Bid-Price Controls for Network Revenue Management*, Mgmt. Sci. **44** (1999b), 1577–1593.

[58] C. J. Taylor, *The Determination of Passenger Booking Levels*, Proceedings II AGIFORS Symposium **2** (1962), 93–116.

[59] H. R. Thompson, *Statistical Problems in Airline Reservation Control*, Opnal. Res. Quart. **12** (1961), 167–185.

[60] W. Vickrey, *Airline Ooverbooking: Some Further Solutions*, J. Transp. Econ. Policy **6** (1972), 257–270.

[61] L. R. Weatherford, S. E. Bodily, and P. E. Pfeifer, *Modeling the Customer Arrival Process and Comparing Decision Rules in Perishable Asset Revenue Management Situations*, Trans. Sci. **27** (1993), 239–251.

[62] E. L. Williamson, *Air Network Seat Inventory Control: Methodologies and Revenue Impacts*, Ph.D. thesis, MIT, Cambridge, MA, 1992.

[63] R. D. Wollmer, *An Airline Seat Management Model for a Single Leg Route When Lower Fare Classes Book First*, Opns. Res. **40** (1992), 26–37.

[64] W. Zhao and Y. S. Zheng, *Optimal Dynamic Pricing for Perishable Assets with Nonhomogeneous Demand*, Mgmt. Sci. **46** (2000), 375–388.