FLIGHT TRANSPORTATION LABORATORY
REPORT R96-6

COMPETITIVE IMPACTS OF YIELD MANAGEMENT SYSTEM COMPONENTS: FORECASTING AND SELL-UP MODELS

BY: DANIEL K. SKWAREK

Competitive Impacts of Yield Management System Components: Forecasting and Sell-Up Models

by

Daniel Kew Skwarek

Submitted to the Department of Aeronautics and Astronautics on August 9, 1996 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Transportation

ABSTRACT

The focus of revenue management research efforts has historically been on the development of seat optimizers which find revenue maximizing booking limits by fare class in a nested fare class structure. Significantly less attention has been devoted to the input methodologies which provide information to the seat optimization algorithm, which is then used to calculate booking limits.

Among these inputs are a forecasting method and detruncation method. The forecaster provides the seat optimizer with estimated mean unconstrained bookings and standard deviation by fare class for a forecast flight. The detruncator adjusts data from historical flights used by the forecaster which have constrained booking information because they have reached booking limits. A third optional input methodology is an adjustment within the seat inventory control process (either to booking data or booking limits provided by the seat optimization algorithm) to account for the possibility of passenger sell-up to a higher fare class when the initially-desired class has been closed. This adjustment has the effect of inducing more sell-up.

Using PODS (a comprehensive simulator of passenger behavior and seat inventory control in a fully competitive framework), this thesis compares pickup, regression, and "efficient" forecasting on a revenue basis. Similar comparisons are performed for no detruncation, booking curve detruncation with and without scaling, projection detruncation, and pickup detruncation. Finally, a modified booking limit strategy to induce sell-up introduced by Belobaba and Weatherford is tested. All tests are performed under a variety of environmental conditions.

Forecasting results indicate that the efficient forecaster is nearly always revenue inferior to pickup forecasting. Neither regression nor pickup forecasting were unambiguously superior: The relative performance of these two forecasters is dependent on detruncation method choice and environmental conditions. Among detruncation methods, not detruncating or pickup detruncation is inferior. Scaling the booking curve used for detruncation yielded superior revenue results over not scaling, and projection detruncation always performed at least as well as booking curve detruncation without scaling. Sell-up tests indicate significant revenue gains to estimating sell-up probabilities. Revenue gains are limited if competitors cannot collude, many alternative flights exist, or passengers have low willingness to pay for higher-valued fare classes.

Thesis Supervisor: Dr. Peter P. Belobaba
Title: Associate Professor of Aeronautics and Astronautics

Table of Contents

List of Tables

List of Figures

I. Introduction

1.1. A brief introduction to revenue management and forecasting

1.1.1. Why revenue manage?

Suppose that there are but two types of consumers who purchase a product. The first type values the product highly and so is willing to pay much more than the other type. This first type also values superior product characteristics and will pay more for them. Any sensible businessperson would try to segment the two, offer the appropriate level of services, and charge accordingly.

Airlines are no exception. The first problem of airline revenue management is to try to identify what passengers value, and how they fall into types ranked by willingness to pay (WTP). If passengers all arrived at once, revenue management's task would simply be to identify whether a passenger was high-value or not and then charge accordingly (this raises the issue of price discrimination, which I will address later). If a direct signal of WTP is unavailable, an airline could resort to measures that are closely correlated. One example would be how "discretionary" is the passenger's trip, i.e., how flexible an individual is making a reservation on this flight.

But passengers do not arrive at once. The *booking process* before a flight over which reservations may be made is up to a year in length. Neither discretionary nor nondiscretionary passengers make reservations randomly throughout the booking process. Instead, as shown in Figure 1.1, the greater proportion of discretionary bookings for a typical U.S. market are made *early* in the process, while most nondiscretionary bookings are made shortly before departure.

This creates a vexing problem: If low-value passengers arrive *early* and high-value passengers *late*, how do we ensure that enough seats are saved for later-arriving passengers without unnecessarily turning away low-value early arrivals? Much thought and not a few careers are dedicated to finding the *optimal* solution to this, airline revenue management's second problem. The industry is continuously regaled by claimants who are sure they have found the *optimal* revenue solution -- a *seat optimization algorithm* which strikes the correct, revenue maximizing balance in allocating seats between high and low value passengers.

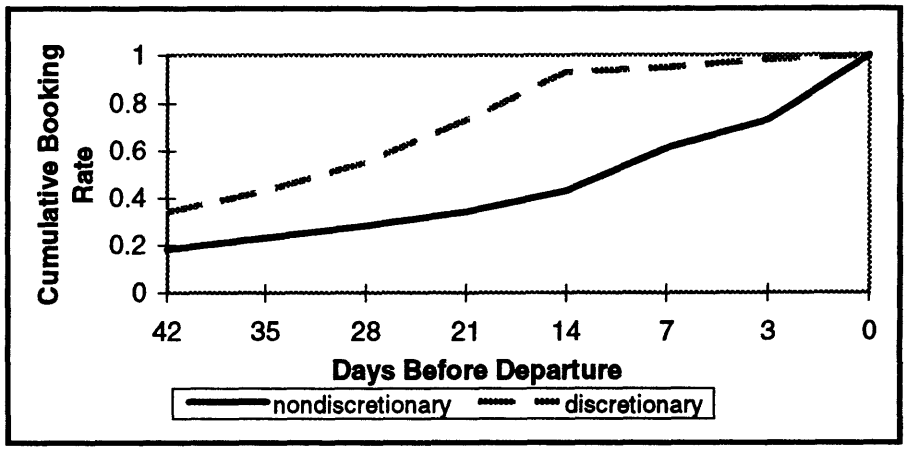


Figure 1.1. Nondiscretionary versus Discretionary Booking Curves

As shown in Figure 1.1, the booking process is divided into N time intervals of unequal length (the duration of each interval decreases as the departure date nears). Between each interval, an updated forecast of final demand is made, usually based in part on actualized bookings until that point, and the seat optimization algorithm is rerun. A *booking interval* is thus defined as a length of time within the booking process in which passengers make reservations but no reoptimization occurs. A large number of intervals is computationally impractical, while too few allows no adjustment for differences between forecast and actual bookings as the booking history for that flight develops. For most U.S. carriers, N is typically 10-25.

Presently, airlines address the two revenue management problems by (1) offering several fare classes on a given flight, each with a different set of restrictions designed to target the fare product to the appropriate passenger type, and (2) using a seat optimization algorithm to appropriately limit the seats available to lower-valued fare classes. It is an involved and complex process, but worth the trouble: Revenue management has repeatedly been shown to yield significant revenue benefits over not offering different fare products at all, or offering different products without limiting availability using a seat optimization algorithm[2].

1.1.2. Where forecasting fits in

[2] Actual trials of an early leg-based seat optimizer at Western Airlines are explored in Belobaba (1987); revenue simulations of more advanced network and quasi-network optimizers are given in Williamson (1992), Tan (1994), and Ferea (1996). Leg-based revenue simulations utilizing PODS, a comprehensive simulator of the passenger demand generation and allocation process, are given in Wilson (1995).

The two processes of constructing fare products and limiting seats to lower-valued fare classes are not independent. As shown in a simplified, idealized diagram of the seat inventory control process (Figure 1.2), such procedures involve only steps 1b, 3, and 4 of the process. This description of seat inventory control is *idealized*: Some airlines will adopt additional procedures, while others will not complete all the steps described here.

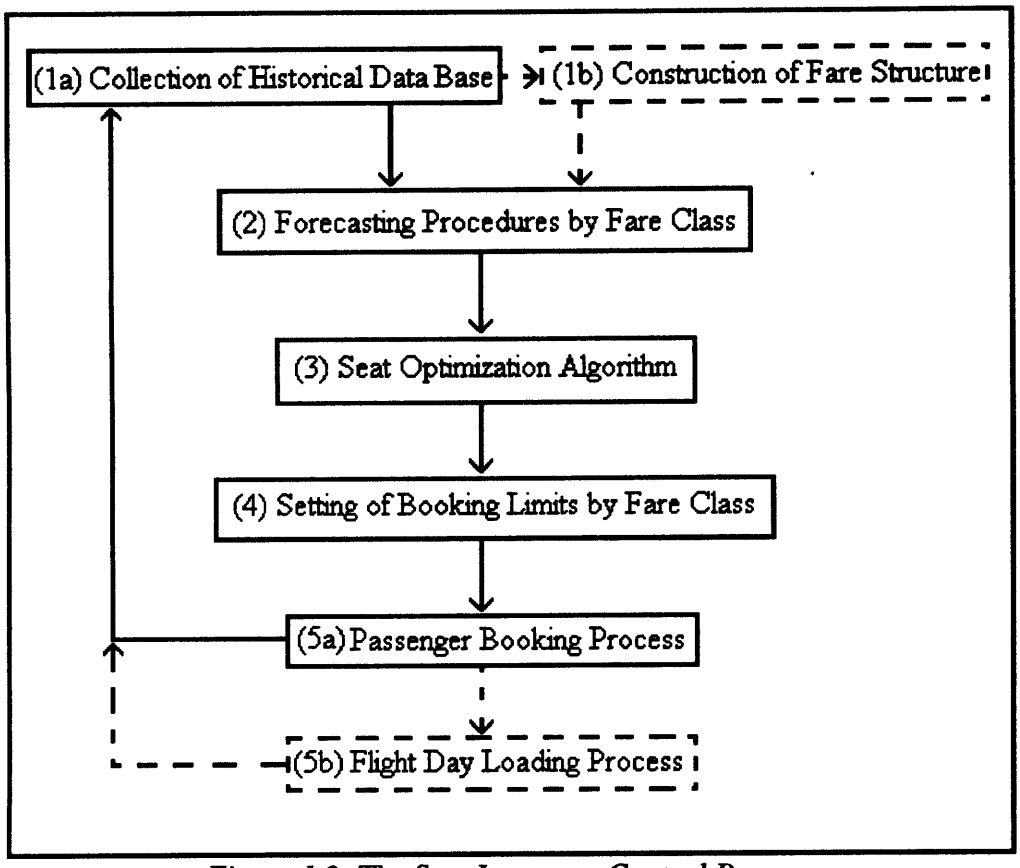


Figure 1.2 The Seat Inventory Control Process

Between each interval in the booking process, Steps 1a through 4 are performed, except Step 1b. This step is represented with a dash because realignments to the existing fare structure (e.g., major fare changes, adjustment of the number and/or restrictions on fare products) are performed less often and require extensive analysis of expected changes of the allocation of bookings to each fare class under the new structure. Careful examination of historical booking data under the existing fare structure provides clues about expected changes. Once these expected changes have been estimated, the historical database is adjusted by the results of this analysis, which is then input into the forecaster. Step 1b is usually considered to be a *pricing*

function; I include it in Figure 1.2 because of its extensive interaction with seat inventory control. Step 5b is dashed because it occurs only on flight day, after the booking process is complete. I will examine the seat inventory control process in more detail under Section 1.3.

Because seat optimization algorithms in Step 3 calculate allocations of seats between fare classes, their success depends upon forecasting tools used in Step 2, which give an accurate forecast of expected unconstrained demand by fare class for each leg[3] or origin and destination (O/D) pair the airline serves. The requirement of leg versus O/D forecasts depends on the allocative level of the seat optimization algorithm: Most present optimization techniques allocate seats at the level of the leg[4], but this is theoretically suboptimal to control on an O/D basis[5] because passenger travel itineraries may involve several legs. Regardless of the level of forecasting required, inaccurate forecasts result in distortions of seat allocations to fare classes, leading to suboptimal revenue performance.

Estimation of the Step 2 forecasts turns out to be a non-trivial problem itself, and there are many different methods to construct these inputs. It is curious but true that *much less attention* has been paid to the revenue effects of the different input methodologies than the seat optimization algorithms themselves. Presentations of novel seat optimization techniques routinely ignore difficulties in forecasting for their particular requirements, assuming instead that an accurate forecasting methodology is readily available.

1.2. Objective of Thesis

This thesis is an attempt to address the traditional neglect of forecasting and other input methodologies by testing their revenue effects in a competitive airline framework. I utilize a comprehensive simulation of the passenger demand generation and flow process originally developed by Boeing, called PODS (Passenger Origin and Destination Simulator). This simulator has several advantages which recommend it for realistically predicting revenue effects according to airline choices about input methodologies -- including barriers in the simulation between

[3] A "leg" is defined to be a flight *stage* involving one takeoff and one landing, or a *nonstop* flight.
[4] See, e.g., the EMSRa algorithm developed in Belobaba (1987), or the EMSRb algorithm in Belobaba (1992).
[5] See Curry (1994) for an example of an O/D-based seat optimization scheme. Severe small number and run-time problems prevent actual use of O/D schemes at present.

forecasting methods and passenger generation processes, and a passenger choice framework which allows for realistic selection of airlines and flights under competition.

Besides testing the revenue effect of alternative *forecasting models*, I will use PODS to compare *detruncation* methods that adjust data from flights on which total demand is not known because booking limits set by the seat optimizer were reached. Without this step, the seat optimizer will not allocate enough seats to high-value passengers who could not book on historical flights because seats were already filled with low-value passengers. Finally, I will test *sell-up models* which adjust for the willingness of some passengers to buy a more expensive fare should their originally requested fare be unavailable.

1.3. The Seat Inventory Control Process

1.3.1. The Internal Airline Perspective

In this section I discuss the seat inventory control process in greater detail, first emphasizing the process which occurs internally at the airline, and then the parts of seat inventory control involving interaction between the airline and passengers. Close examination of the relationships between the processes I will test (i.e., forecasting, detruncation, and sell-up) and seat inventory control provides a systematic understanding of the influences on these processes. Again, this depiction is *idealized* and does not describe the practice at any particular airline. Step 1b (governing changes in the fare class structure) will not be discussed.

The process begins at Step 1a in Figure 1.3, when previous departures of the flight to be forecast are initially selected from an airline's data on historical flights to form the Historical Data Base (HDB) for the forecast. Selection processes are designed to exclude previous flights which might have systematic differences with the flight in question. Thus, departures on different days of the week, with different aircraft sizes, or under differing competitive circumstances might be excluded from the data set.

The "unclean" data from the flights selected for inclusion are analyzed to remove and/or adjust entry errors, outliers, and other anomalous patterns in the data. Each previous flight in the database will have at least three characteristics: the total bookings *BIH(0)* received on each flight before the end of the last booking interval, which ends on the day of departure; no-shows (NS) who book but do not show up on departure day; and denied boardings (DB) -- would-be

passengers who book and show up but cannot board because of capacity restrictions on the aircraft . If there is strong seasonal variation over this dataset used for forecasting, it is adjusted to the season of the flight to be forecasted. Additionally, passenger loads are "*detruncated*" if previous load data are constrained by booking limits, i.e., passenger loads in a fare class reach the limits established for that fare class at any time during the booking process.

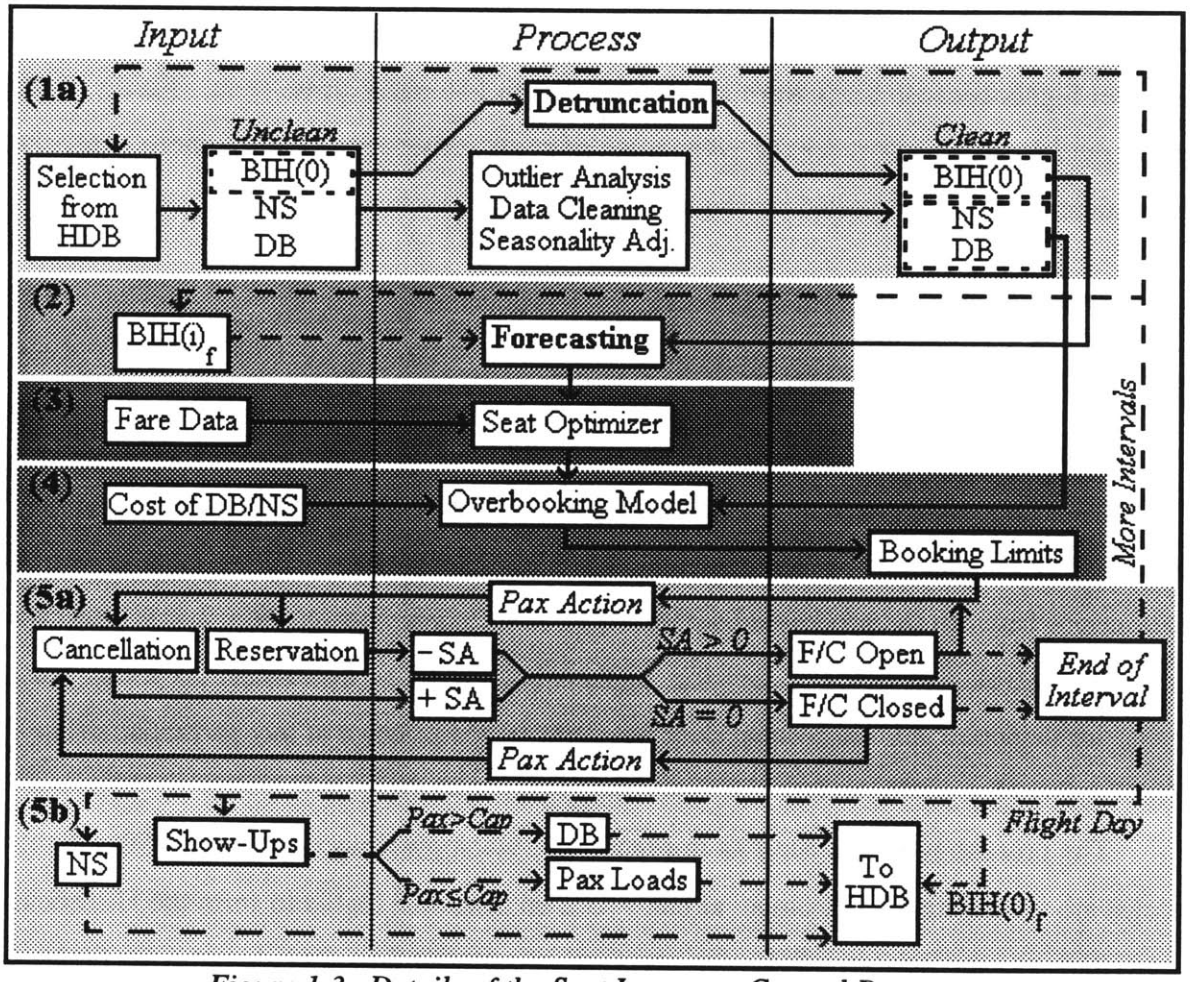


Figure 1.3. Details of the Seat Inventory Control Process

After the data have been appropriately adjusted, bookings information goes into the *forecaster* (Step 2). In the first booking interval (before any bookings on the flight have been taken) the forecaster estimates expected unconstrained total bookings $\hat{BIH}(0)_f$ or demand for the flight f singly on the basis of historical booking data. This is input into the seat optimizing algorithm (Step 3), which also takes fare values by class and sets seat booking limits on each fare class that maximize expected revenues.

In Step 4, information about NS and DB from previous flights, unadjusted booking limits provided by the seat optimizer, and the airline's estimation of the monetary cost of a denied boarding and no-show are analyzed by the overbooking model. It adjusts booking limits by trading off the expected benefit of accounting for no-shows with *overbooking* against the increased probability that more passengers will show up than space is available for[6].

Once adjusted booking limits BL_i for the ith interval have been provided by class, the booking process for the flight opens (Step 5a) and passengers may begin making reservations and (subsequently) cancellations. The seat availability SA or number of additional requests which will be accepted for each class is initially BL_i. Once a passenger makes a reservation, SA on this fare class is decremented (BL_i is set by the seat optimizer and therefore unchanged within booking intervals). Other fare classes' SA are also decremented. Many nested seat inventory control processes decrement SA in all *higher-valued* fare classes; others decrement SA in *every* fare class[7]. The rationale for the first policy is that the seat protection algorithm has *already* limited bookings in low-value classes via BL_i, *given* expected bookings in higher-valued classes. Decrementing SA in these classes therefore doubly impacts their availability -- once when BL are set, another when bookings are made in higher fare classes. At present, PODS decrements *all* fare classes. Because this approach decrements low value classes' SA more often, it closes low-value classes *earlier*.

This issue gives rise to a second measure of capacity on a fare class. For any given time within an interval i, I define the maximum allowable bookings Mx on a fare class to be the booking limit BL_i established for that fare class, less *net* bookings received in i on *other* fare classes which affect this class' SA[8]. Unlike BL but like SA, Mx reflects the fact that the maximum allowable bookings in a given fare class *within an interval* is continually adjusted for bookings made in other fare classes.

In any case, if after the reservation there is still space available in that fare class ($SA > 0$), the fare class remains open for more reservations. If not ($SA = 0$), the fare class is closed, and no more reservations will be taken. However, if a passenger acts to cancel his or her reservation,

---

[6] An interesting recent analysis of how best to do this is given by Holm (1995).

[7] Conceptually, these SA decrement strategies are heuristic compromises for the fact that adjustment of BL via reoptimization cannot feasibly occur after each booking/cancellation.

[8] *Net bookings* in these classes are total bookings less cancellations within interval i up until the time period of interest. Generally, $Mx \neq SA$ in a given fare class, because the latter includes bookings which occur during i *within* that fare class.

seat availability is incremented (+ SA) and the fare class is opened again. The same booking limit BL_i for each fare class remains as long as the same booking interval (as discussed in Section 1.1 above) is in effect. When a booking interval ends and *more intervals* remain until departure, we keep account of the total *bookings-in-hand* which have been received up to this interval i on this flight f, $BIH(i)_f$. Next, we rerun Step 1a (see dashed lines), taking advantage of the most recent information provided by flights which have departed since the last run through the seat inventory control process. Then, the forecaster (Step 2) combines the revised information from the historical data base with present bookings $BIH(i)_f$ to provide an updated forecast of total demand for the flight, by fare class[9]. The seat optimizer (Step 3) and overbooking models (Step 4) are rerun, leading to booking limits BL_{i-1} for the $(i-1)$th interval[10]. Seat availability by class at the start of interval i is then $SA = BL_i - BIH(i+1)_f$[11].

This process is repeated until no more booking intervals exist, i.e., *flight day* has arrived. On flight day, total bookings realized $BIH(0)_f$ for the flight f go into the HDB, and the passenger boarding process begins. Those passengers who do not show (NS) are recorded into the historical data base for this flight. Among those who do show up, if the overbooking model has correctly adjusted booking limits, there will be zero or few denied boardings (DB), which are also input into the HDB. Boarded passengers become passenger loads for this flight, which completes the HDB data set for this flight. The seat inventory control process for this flight is now complete.

1.3.2. The Airline/Passenger Interactional Perspective

Now I discuss a subsection of the seat inventory control process dealing with interaction between the passenger and airline. Steps 5a and 5b in Figure 1.3 above described the activity space over which passengers make requests for service, and airlines respond. Figure 1.4 expands these processes.

---

[9] Instead of reestimating final demand during every reoptimization, some forecasters estimate the *bookings to come* from the present forecast interval until departure (see Section 5.2.1).

[10] Many seat optimizers set booking limits on the basis of the remaining seats available on the flight. In this case, an estimate of bookings to come (BTC) would be necessary (see Footnote 9). BTC may be derived from forecasts estimating final demands by the simple formula $BTC = \hat{BIH}(0)_f - BIH(i)_f$.

[11] It is therefore possible that $SA < 0$, if $BIH(i+1) << BIH(i)$ and $BL(i+1) >> BL(i)$ -- say, because of unexpectedly high bookings in high-value classes. In this case no bookings in the affected class are allowed in interval i unless many cancellations occur.

- *Reservations Phase*

When a potential passenger makes a request for a reservation on a particular O/D itinerary, either a travel agent (approximately 70 percent of the time) or an airline ticket/reservations agent (about 25 percent of the time) inquires the computerized reservations system (CRS) about space availability for the itinerary. A passenger can make a reservation in a particular class if seat availability *SA* is greater than zero, after all decrements and increments up to the present time. If so, a reservation is made and *SA* is decremented by one.

Otherwise, the potential passenger's initial request is denied, and (usually) an alternative will be offered. Those who accept the alternative -- either on the same flight in a more expensive fare class or on an alternative flight on the same airline -- are "recaptured" by the airline. *Sell-up* occurs in the former case (see Section 3.3), or if a potential passenger buys a more expensive fare product on an *alternative* flight. This behavior *defines* sell-up, but does not describe where adjusting for this possibility fits into the seat inventory control process. As I will show later in Section 3.3, there are several methods to induce sell-up, each of which modifies a *different* part of the process. If the consumer refuses the offered alternatives and chooses a competitor or decides not to travel at all, the customer is *lost* to the airline.

- *Confirmation Phase*

One a passenger has made a reservation, he or she will receive a ticket upon appropriate payment to the airline. Before or after ticketing, the passenger may change plans and thereby *explicitly cancel* the reservation, in which case seat availability *SA* for the fare class in which the passenger was booked increase by one. The airline may also unilaterally cancel the reservation (resulting in an "*implicit cancellation*") if the passenger fails to meet some restriction associated with the ticket, e.g., purchase within one day of reservation (typical of low-value fare products) or reconfirmation of a reservation (typical on some international flights).

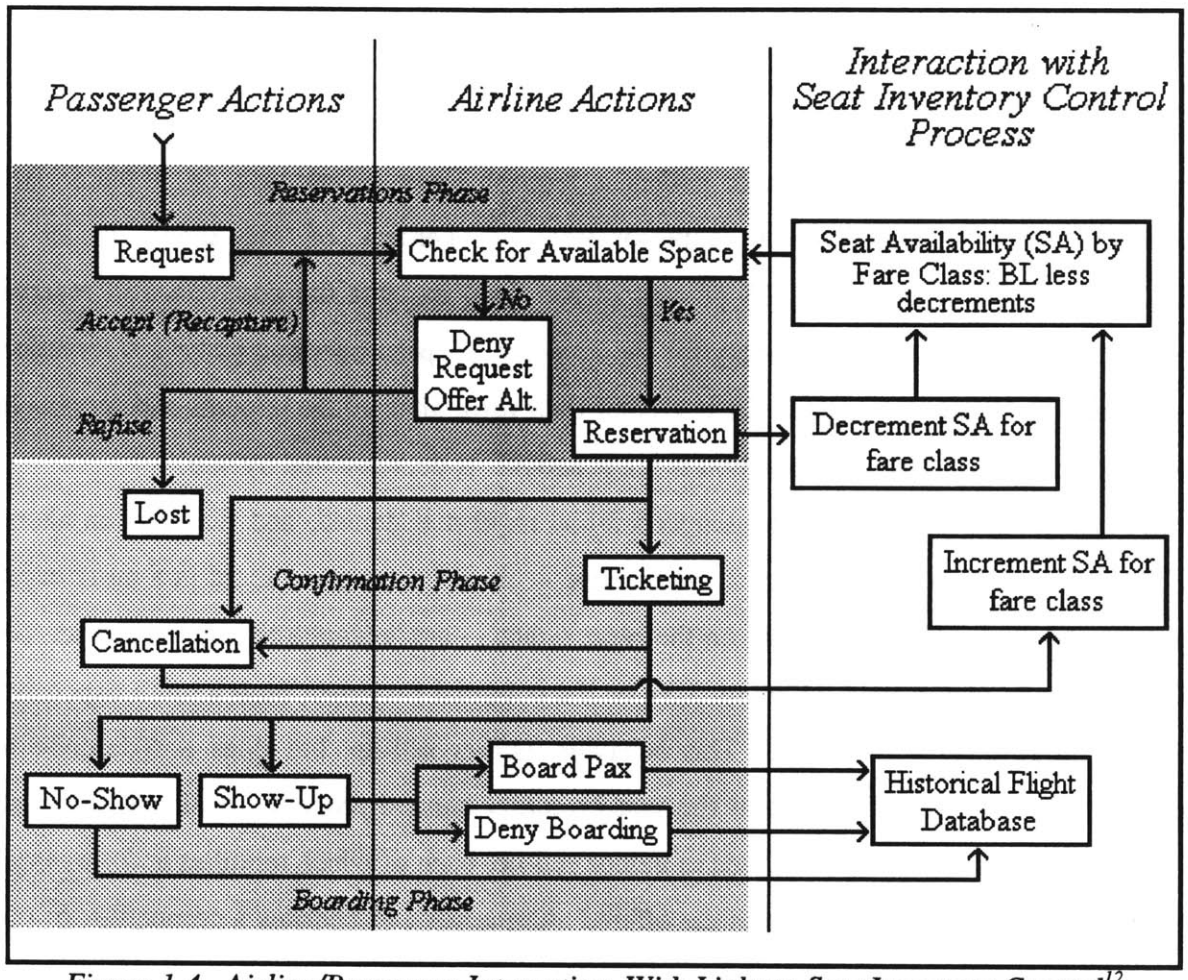


Figure 1.4. Airline/Passenger Interaction, With Links to Seat Inventory Control[12]

- *Boarding Phase*

Finally, on the day of departure we enter the *boarding phase*, and the passenger can either show up for departure or may *no-show,* i.e., not show up despite having a reservation and/or a ticket. No show probabilities are very specific to market and flight, and are generally higher on short-haul, high frequency, and business markets.

If the passenger does show up, space may not be available on the flight because of the overbooking. Airlines have established various procedures to compensate passengers *denied boarding* -- i.e., those who arrive in expectation of traveling but find no seats available for them. After the flight has departed, total bookings received, loads, no-show and denied boarding

[12] Adapted from Lee (1990), pp. 18-24.

information are recorded to this flight's HDB, which will be later be used by the forecaster to predict total demand for future departures of the same flight.

1.4. Structure of Thesis

Chapter 2 provides a brief background of basic economic theory of supply and demand, and the types and profit potential of price discrimination. This chapter provides a theoretical explanation for the provision of multiple fare products with associated levels of restrictions. Whether or not airlines price discriminate is considered by examining current fare structures and their ability to "fence" passengers into desired groups. Finally, I discuss the arbitrary nature of this demand division for forecasting purposes.

Chapter 3 includes a theoretical motivation and review of the present literature on the three input methodologies this thesis focuses on: flight-level passenger forecasting, detruncation, and sell-up. I discuss models which have been advanced in the literature, and comparative studies using simulation or analysis of accuracy. Shortcomings of the various models and comparisons are discussed.

Chapter 4 gives a background in the architecture and theoretical assumptions used by PODS, a simulator of passenger flow and revenues I use for comparing forecasting, detruncation, and sell-up schemes. This chapter summarizes the salient points of PODS without extensive details, for which the reader is referred to Wilson (1995) -- the first installment of results from the joint MIT/Boeing collaborative research project on PODS. I also pinpoint some different assumptions made by the PODS methodology and previous comparative studies of forecasters, detruncation methods, and sell-up mechanisms.

Chapter 5 describes the subset of models and methods from Chapter 3 I have chosen to compare using PODS. The assumptions and techniques of each of the tested methods are described. Chapter 6 is a discussion of simulation results for each of the chosen models and methods. Each comparison is run under a variety of conditions (e.g., demand conditions, demand stochasticity, frequencies in market, passenger characteristics) to examine the sensitivity of our conclusions about the revenue ranking of the described models to underlying market conditions. I present theoretical explanations of persistent revenue differences. These revenue results reinforce

the importance forecasting, detruncation, and sell-up play in determining the performance of the revenue management system.

Finally, Chapter 7 concludes with a detailed summary of these results and a short discussion of the projected future research areas for the Boeing/MIT PODS research project.

II. Revenue Management, Price Discrimination, and Product Differentiation

2.1. Uniform Pricing

2.1.1. Basic Supply and Demand Theory

Basic neoclassical economic theory posits upward-sloping supply curves and downward-sloping demand curves in most markets[13]. The latter results from the *differential willingness to pay* property of the simple model of competitive markets: Some potential consumers are very willing and able to purchase a product, but others are willing to purchase only if the price is lower. As price is successively lowered, more potential consumers are coaxed into purchasing the product. An opposite relationship obtains for producers: As more quantity is supplied, production costs rise as less efficient resources are employed to the production of the product. These relationships are depicted in Figure 2.1 where $S(q)$ and $D(q)$ represent the supply and demand curves, respectively.

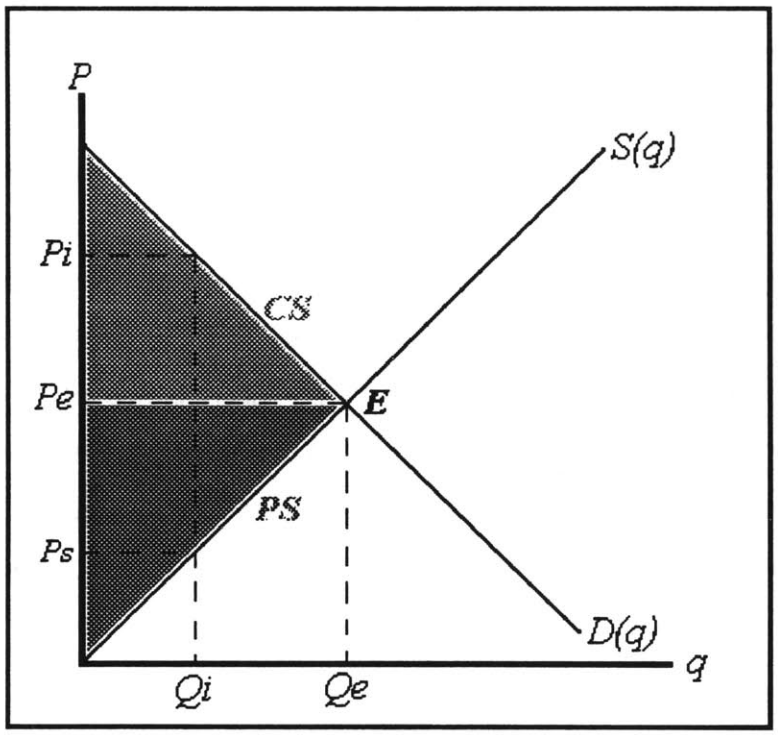


Figure 2.1, Basic Supply and Demand Graph.

[13] Hirschleifer and Glazer (1992), pp. 22-39 gives one of the innumerable treatments of supply and demand relationships.

Under these conditions an equilibrium at point *E* results where the price consumers are willing to pay and the price suppliers are willing to accept is equal at *Pe*. Here *Qe* is exchanged. Any other quantity leads to a mismatch between the price consumers are willing to pay and the price suppliers are willing to accept. Negotiation between the two resolves the mismatch as price and quantity are driven to the equilibrium point.

With the uniform price *Pe,* some consumers are willing to pay more for the product but end up paying only *Pe*. For example, a customer *Qi* in Figure 2.1 is willing to pay up to *Pi* for the product but only pays *Pe*. This consumer gains *consumer surplus* in the amount *Pi - Pe*. Analogously, the supplier gains a *producer surplus* of *Pe - Ps* from trade with the *Qi*th consumer, because he or she would be willing to accept a price as low as *Ps*. The total surplus gained by consumers and suppliers as a result of exchange to *Qe* is represented by the shaded areas *CS* and *PS,* respectively.

2.1.2. Disadvantages of Uniform Pricing

From the producer's perspective, uniform pricing has some disadvantages. Namely, *CS* remains to consumers as the benefit of this transaction. If market conditions and antitrust law permit, suppliers would prefer to obtain not only *PS* but *CS*. Under most market conditions a uniform price will *always* leave some surplus to consumers[14]. Therefore, profits are not maximized.

Attempting to extract *CS* is commonly understood to be "bad" since it requires some consumers to pay more for the same product. There are three responses to this assertion. First, in an economic sense costless extraction of *CS* by producers is neutral. It does not affect the competitive equilibrium (*Qe* at price *Pe* in Figure 2.1), but is merely a transfer of funds from one to another economic actor. Second, when consumers have some degree of market power, quasi-monopsonistic conditions allow appropriation of the producer's *PS* -- but society rarely considers this to be a bad[15].

---

[14] Tirole (1988), p. 133. Of course, consumers prefer the opposite: that they not only keep *CS* but extract *PS* from producers.

[15] One example would be the market for specialized defense products that are salable only to the U.S. government.

Third, under certain cost conditions, a producer cannot charge a single price and break even (let alone profit). Consider Figure 2.2, where demand *D(q)* is downward sloping as before but producer marginal cost (the cost of producing an *additional* unit) is constant, while average costs *AC(q)* (the total cost of production averaged over all units) are continuously declining. Such a situation would obtain if there were a high fixed cost to production and low costs of producing marginal units. In the airline industry, this is certainly true with respect to the provision of the marginal seat, where fixed costs include aircraft, gates, administrative, and other expenses[16].

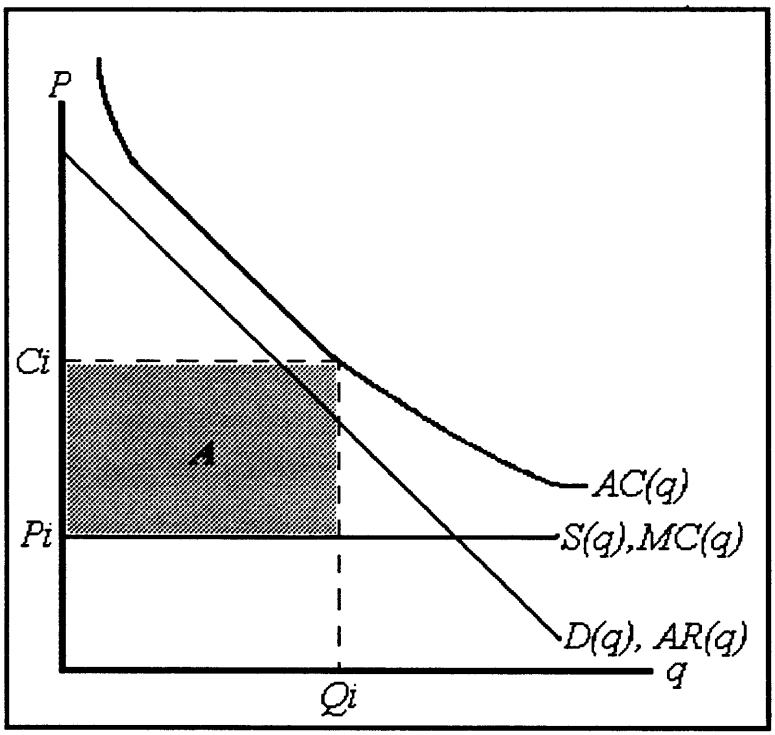


Figure 2.2, Declining Average Curve above Marginal Cost Curve.

Under moderately competitive conditions and uniform pricing, economic theory posits that suppliers will charge consumers the *marginal cost* of producing the good[17]. The supply curve

[16] It is important to note that "fixed" versus "variable" distinction is defined only with respect to the time period within which the decision to incur a particular cost is made. Thus, the aircraft fleet size decision is fixed within a decision period of a month, but variable within a year. For a sufficiently long time period (e.g., several years), all costs are variable, while all costs are fixed in the extremely short run. The time period of interest in this analysis is about a month, i.e., long enough for the *individual flight* decision to be variable while facilities and aircraft available are fixed.

[17] Hirschleifer and Glazer (1992), pp. 153-156.

$S(q)$ is therefore equivalent to the horizontal marginal cost curve $MC(q)$ in Figure 2.2. In this illustrative case average costs are high relative to demand, so $AC(q)$ is above $D(q)$ for all q. Since under competitive uniform pricing the profit per unit for the producer is defined as average revenue (the uniform price) less average cost or $S(q) - AC(q)$, the producer will always take a loss regardless of production level q. At quantity Q_i, for example, the producer loses area A since average costs are C_i at this point but the uniform price is only P_i. In the long run, this loss-making condition is not sustainable: If firms cannot at least break-even, they will eventually exit the industry[18].

A final argument against uniform pricing is the inefficiencies it creates by failing to account for peak versus off-peak demand conditions. Economic theory indicates that when a good is not storable (e.g., electricity, which must be produced on demand), demand is subject to significant fluctuations, and provision of a unit of capacity is not free, uniform pricing will result in allocative distortions[19]. That is, in peak demand conditions individuals with the highest valuations of the good should receive the good while those with low valuations should purchase only in off-peak demand conditions. Under uniform pricing, each is just as likely to get the good. This is not a problem in off-peak periods because all can be accommodated. But in high periods some low-value individuals will receive the good, denying high-value individuals.

Peak load pricing avoids this by setting a higher price in peak periods and a low price in low periods. Only high-value consumers will be accommodated in the former, while all can be accommodated in the latter. Such pricing encourages the efficient use of resources as low-value consumers turned away in peak periods switch to off-peak consumption. This behavior is exhibited in the airline industry: A major point of studies of seat optimizers under variable demand conditions is the automatic protection of more seats for high fare passengers as market demand increases, thereby restricting availability of low fare seats to low demand flights[20].

2.2. Price Discrimination: An Alternative to Uniform Pricing

---

[18] Hirschleifer and Glazer (1992), pp. 159-161.

[19] Crew and Kleindorfer (1986), pp. 33-37; and Borenstein (1983), pp. 111-113.

[20] See, e.g., Wilson (1995). Chapter 6 of this thesis will confirm these results. A significant benefit of seat optimization algorithms is the ability to identify and dynamically adjust to demand variations on a *per flight* basis as bookings develop. This is opposed to relying on generalizations about which seasons or days of week, etc. have high demand, and performing blanket adjustment of availability over presumably affected flights.

In certain circumstances, producers avoid the suboptimal condition of uniform pricing by *price discriminating* -- charging different prices to different consumers for the same product. The objective here is to somehow identify consumers by willingness to pay and charge accordingly, and thus extract as much of the *CS* of Figure 2.1 as possible. Price discrimination is only operative if: it is legal[21], the producer/s can influence prices[22] and can successfully identify consumers by willingness to pay (WTP), and resale of product among consumers is impossible (otherwise the identified low-WTP consumer buys all the product and resells to others).

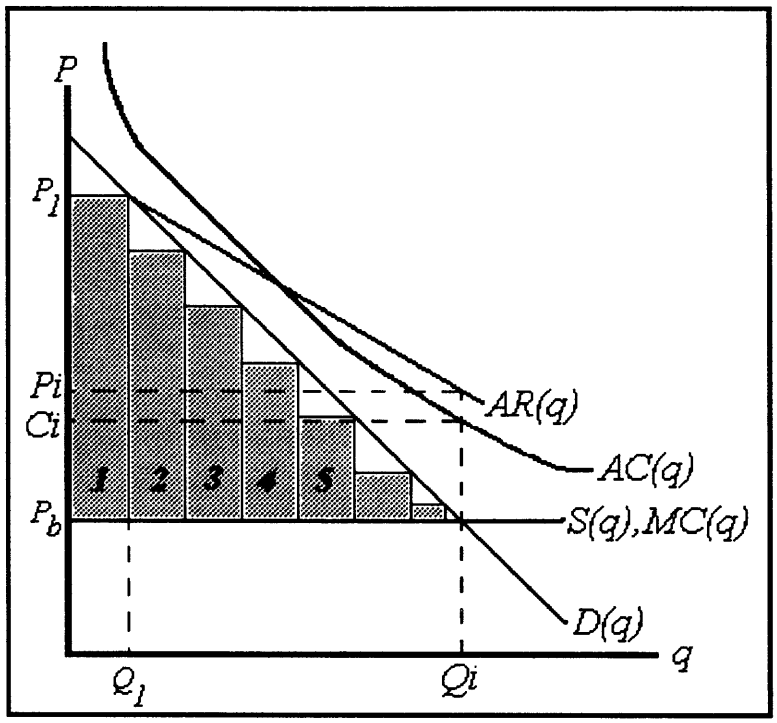


Figure 2.3, Second Degree Price Discrimination with Declining Average Cost.

[21] The Robertson-Patman Act prohibits some price discrimination, but it has never been applied to the airline industry. Several difficulties prevent such an application, including the understood exemption from the law of services and intangibles (Areeda [1981], p. 1058), the "like commodities" requirement that the products offered at varying prices be substantially similar, the cost differential by product defense (Areeda [1981], pp. 1102-1105), and the "meeting competition in good faith" defense (Areeda [1981], p. 1115). The recent antitrust lawsuits against airlines' pricing practices are a case in point. The allegations involved alleged *price fixing* attempts using CRS systems (Hunt [1994]). The Department of Justice did not raise a price discrimination issue, despite wide price ranges in fares which airlines allegedly attempted to fix.

[22] This traditionally requires the market to be either monopolistic or oligopolistic. However, Borenstein (1983) shows that price discrimination can occur in reasonably competitive circumstances if products are somewhat heterogeneous *between suppliers* and consumers have brand preferences. A spatial model of monopolistic competition is used to prove this result.

Economists categorize methods of price discrimination into several types[23]. Airlines are often accused of *second-degree* price discrimination, where a set of product "bundles" are offered. By voluntary market exchange, consumers choose the bundle most suited to them and incidentally forfeit more or less *CS* to the producer depending on their WTP. This type of price discrimination in the decreasing $AC(q)$ with constant $MC(q)$ case is shown in Figure 2.3. The producer has offered seven distinct products *1* ... *7* in decreasing order of price, and has successfully segregated consumers by decreasing WTP into the seven groups. Thus, consumers with the highest WTP purchase Q_1 at price P_1. Suppose in this way a total of Q_i consumers purchase the product. At this production level the average cost per unit is C_i, which is above the demand curve $D(q)$. However, because each of the seven groups of consumers have paid greater than the uniform price P_b at Q_i, the average revenue curve $AR(q)$ (i.e., average price paid per unit at Q_i) is no longer coterminous with the demand curve $D(q)$. Instead $AR(q)$ shifts out so P_i, the average price at Q_i under the seven-product strategy, is greater than the uniform price P_b *and* average cost C_i. A profit per unit of P_i - C_i is earned. In this case price discrimination is preferred by consumers *and* producers: It is the difference between the market existing or not.

2.3. The Discriminatory Nature of Airline Fare Structures

2.3.1. Identifying Passenger WTP and Type

Traditionally, airlines divided passengers into two types, business and leisure. Business travelers were assumed to have the higher WTP or, equivalently, lower *price sensitivity*. But consumer research in the late 1970s demonstrated that the distinction which should be drawn was between *discretionary* and *nondiscretionary* business travel, since some non-business travel is mandatory (e.g., emergencies among close relations), while some business travel is optional[24]. Second, it was noted that passengers differ on variables other than simply price sensitivity: Some operate under extreme time constraints and others have more flexible schedules.

[23] Tirole (1988), p. 135.
[24] Belobaba (1987), pp. 20-22, discussing CTC (Canadian Transport Commission) survey.

A two-axis representation of passenger type on *price* and *time sensitivity* variables was developed by Belobaba[25]. As shown in Figure 2.4, this creates four types of passengers based on price and time sensitivity.

- *Type I: Time-sensitive and price-insensitive.* This category characterizes the nondiscretionary business traveler, especially one who is not personally paying for the trip. Such passengers have extremely tight schedules (requiring nonstop travel when possible), firm up travel plans very late, and often change plans. They are sometimes willing to pay for a superior cabin class. Type I business travelers dislike spending weekends out of town.

- *Type II: Time-sensitive and price-sensitive.* These passengers are nondiscretionary but are more concerned about price. They will not pay for first or business class and exhibit limited flexibility in schedules in order to obtain cheaper fares.

- *Type III: Time-insensitive and price-sensitive.* The typical leisure passenger falls into this group, which finds changes in travel date and even destination acceptable if it means a lower fare. Their schedule plans are typically set far in advance of travel.

- *Type IV:* These few travelers have few constraints on travel dates, and are willing to pay for superior service cabins and flexibility of travel arrangements in case they make alternative plans.

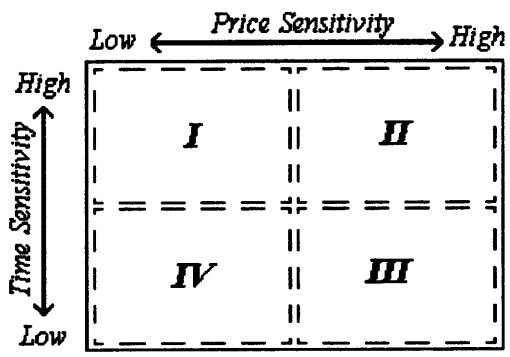


Figure 2.4. Market Demand Segmentation Model

Based on this information, the WTP ranking of the four passenger types is: $WTP_{IV} \geq WTP_{I} > WTP_{II} \geq WTP_{III}$. When a potential passenger makes a request for a reservation

[25] Belobaba (1987), pp. 24-27.

as previously described (see Section 1.3 above), the request will contain information on *at least* the following characteristics listed below. If these variables signal underlying passenger type and associated WTP, tailoring specific products accordingly allows discrimination.

- Days before departure that request is made
- One way or round trip
- If round trip, duration of stay at destination
- If round trip, does stay involve a Saturday night
- Desired time of departure/s or arrival/s
- Desired cabin of service

Table 2.1 provides hypothetical results by passenger type for each of the characteristics revealed by a potential passenger at time of request. Most characteristics appear to be good predictors of passenger type. Do airlines introduce different fares on these bases?

Pax Type	*Hypothetical Revealed Characteristics at Time of Request*					
	Days Until Departure	One Way or Round Trip	If RT, Duration of Stay	If RT, Saturday Stay?	Departure/ Arrival Time	Cabin of Service[26]
I	< 7	OW	< 1 week	No	peak hours	Y or C
II	< 14	?	≤ 1 week	No	more flexible	Y
III	≥ 14	RT	> 1 week	Yes	very flexible	Y
IV	?	OW	?	?	?	F or C

Table 2.1. Hypothetical Characteristics of a Request by Pax Type

2.3.2. Evidence of Possible Price Discrimination

An example of the array of fare classes offered on a hypothetical market is shown in Table 2.2 below. In this market, there are three cabins with varying levels of restrictions on each of the fare classes offered in the cabins. A "luxury" Type IV passenger will select to class C or F because of the superior service offered, the absence of restrictions, and full refundability of the product. Type I passengers, however, make reservations shortly in advance of travel, wish to spend weekends with family, and often travel for indeterminate lengths -- thus making round trip

[26] These are standard *cabin class* (as opposed to *fare class*) distinctions: F is first class, C is business class, and Y is coach.

travel difficult. Combined with their desire to avoid to avoid the strict nonrefundability conditions of the low-valued classes, in these circumstances they will likely select a C, Y, or B fare.

Type II passengers are more flexible in timing arrangements but still dislike travel involving a Saturday, and sometimes change travel plans (thus making nonrefundability onerous, given their higher price sensitivity). They are likely to select a B, M, or (perhaps) an H fare. Finally, the typical leisure passenger is completely flexible with travel dates and Saturday night stays, and can make reservations far in advance. This type will therefore typically purchase the cheapest Q fare class.

Airlines include fine gradations of fares *within* each fare class, which sorts between the types of passengers who might select a particular fare class. For example, successively lower fares within the M fare class will be associated with more stringent restrictions on time of day, connections vs. one-way, and day of week to sort between the Types I and II passengers who typically purchase that fare class. The percentage proportion of the base Y-Class fare for each fare class is therefore an average over fares offered in that class. The illustrative example of Table 2.2 indicates *significant* price differences between and within cabin classes: A first class fare (F) costs five times the cheapest (Q) fare on the flight, while the most expensive coach class fare (Y) is about three times as expensive as the cheapest coach fare.

Cabin	Fare Class	Avg. Percent of Y Fare	Restrictions on Fare Class
First	F	150%	None (First Class)
Business	C	120%	None (Business Class)
Coach	Y	100%	None (Full Fare Coach Class)
	B	75%	3 Day Advance Purchase (A/P)
	M	60%	7 Day A/P, Sat. Night Stay
	H	45%	14 Day A/P, Sat. Night Stay, Non Refundable, Round Trip
	Q	30%	21 Day A/P, Sat. Night Stay, Non Refundable, Round Trip

Table 2.2: Typical Airline Fare Class Structure[27]

Limited information is available on how successfully these restrictions or "fences" direct passengers to the appropriate fare product. The complication involves an imprecise ability to identify a passenger's type from the information provided at time of request. The variables we

[27] Adapted from Ferea (1996), p. 22 and Lee (1990), p. 30. Fare class structures vary slightly by airline and by market (only a limited number of markets have business class cabins, for example).

have examined are only *imperfectly correlated* with a passenger's willingness to pay, and some passengers will always be able to circumvent the designed sorting mechanisms by replanning their trip or other evasive methods[28]. Thus, it is not possible to quantify the proportion of passengers in each type.

A few independent attempts have been made at identifying the effectiveness of restrictions by passenger type. Results of a Boeing survey which asked passengers on markets of greater than 1,300 miles to classify themselves into one of three categories and then indicate whether they would on their present trip be able to meet a specified set of restrictions are given in Table 2.3 below[29]. The three categories -- nondiscretionary business (NDB), discretionary business, and leisure -- match fairly closely to Types I, II, and III, respectively.

Requirement		*% of Travelers Able to Meet Restriction, by Type*			
Minimum Stay	Advance Purchase	Non-Disc. Business	Discretionary Business	Leisure	Total
0	7	72%	86%	95%	86%
0	14	53%	77%	90%	76%
0	30	30%	61%	74%	59%
7	0	28%	32%	64%	48%
7	7	25%	28%	62%	45%
7	14	21%	27%	58%	41%
7	30	14%	23%	49%	34%
Saturday	0	47%	56%	80%	66%
Saturday	7	39%	51%	77%	61%
Saturday	14	30%	47%	72%	55%
Saturday	30	19%	40%	61%	45%

Table 2.3: Percentage of Travelers Able to Meet Fare Restrictions, By Pax Type, Flights over 1300 Miles

While the restrictions considered are limited to a minimum stay and/or advance purchase requirement, the survey indicates that offering differing prices on these signals effectively fences passengers into appropriate fare classes: Only 28 percent of NDB travelers could accept a seven-

[28] One common technique to avoid the Saturday night stay restriction is to buy *two* round trip tickets, one from one's origin and the other from the destination, with scheduled departure dates on the first leg of each ticket matching one's original preferences. This is sensible only if the price of a Saturday stay round-trip ticket is less than half of the unrestricted price.

[29] Boeing (1988).

night stay restriction, while 64 percent of leisure travelers could. Similarly, the percentage of travelers able to meet advance purchase restrictions increases as we move from NDB to leisure, while declines as the restriction is tightened are more precipitous for NDB and discretionary business. Over the markets surveyed, the advance purchase restriction is most effective in screening between NDB and discretionary business, while the minimum stay requirement segments between business and leisure passengers. Table 2.3 in concert with Table 2.2 confirm the assertion that characteristics of the reservation request are important signals of differences in WTP[30], which airlines exploit by offering restricted fare products accordingly. That is, some evidence indicates that airlines offer different prices for substantially the "same" service.

2.3.3. Is Airline Behavior Actually Price Discrimination?

However, the arguments in Section 2.3.2 are absolutely *not* conclusive about whether or not airlines actually price discriminate. In this section I examine three arguments that the pricing structure in current practice is not discrimination at all. The primary point is that the various fares offered to passengers are not similar. If the fare products airlines offer are not the same, the evidence in Section 2.3.2 indicates only acceptable *differential* pricing.

First, consider the fact that potential passengers do not make requests for service at the same time. Some requests arrive early, others arrive late. If airlines adopted a first-come-first-serve strategy in that late arrivals would simply be denied service, it is justifiable that all be charged the same price. However, recognizing that many passengers cannot make decisions until late in the decision process, airlines save space for late-arriving passengers -- often explicitly turning away earlier-arriving passengers to do so. The result of this, of course, is a different price. Those arriving late also pay for having space saved for them until just before the flight. Early arrivals should pay less because there is a minimal likelihood that the airline has had to turn someone else away in order to serve them. The entire purpose of seat optimization schemes, as has been noted, is to strike the correct balance between saving enough seats for the uncertain

[30] There are two caveats to Table 2.3. First, it is self-reported and therefore are subject to all the problems of surveys (e.g., correct interpretation of questions, class misidentification, and the stated versus revealed preferences problem). Second, the percentages are a function of fare differences between fare products assumed by respondents and the characteristics of the markets surveyed (e.g., a short-haul market will be more significantly affected by advance purchase restrictions).

number of late-arriving passengers without turning away too many certain early-arriving passengers, given a set of fares. *Availability of a seat before departure* becomes a product characteristic which differentiates the fares purchased early and late in the booking process.

Generally, the variety of restrictions attached to specific fares create numerous different products. A reduced-fare product (say, H in Table 2.2) requiring round trip travel and not refundable cannot plausibly be considered to be same product as a full-Y fare product which does not entail round trip travel and is refundable. This situation is exactly analogous to the division of seating in a theater according to desirability, charging the highest prices for close-in seats, and limiting of discount seats before the performance. No one argues that such a division is appropriate, since high-WTP theater patrons derive more utility from a close-in view of the performance. This is true even though the physical seats are identical, and everyone sees substantially the same performance regardless of seating location. Similarly, the physical seats on an airplane are substantially identical, and passengers in the same cabin receive approximately equivalent levels of service. However, high-WTP passengers derive more utility from the flexibility and availability of a fare product.

Finally, it has been argued that because nondiscretionary passengers place a high value on travel flexibility and options, they should compensate for cost inefficiencies inherent in catering to their demands[31]. For example, offering a given capacity in a market would be cheapest if there were but one frequency on a large aircraft. Since the economies to aircraft size are well-established, demand consolidation via reduced frequencies clearly reduces overall costs.

Discretionary leisure passengers would be satisfied with this outcome because of their time insensitivity. However, the one frequency option imposes a large disutility on time sensitive nondiscretionary passengers, who place a high cost on "schedule dislocation" between desired and available departure time/s. The offering of multiple frequencies reduces time sensitive passengers' schedule dislocation costs. Since the additional costs are almost completely borne for the benefit of the high-WTP time sensitive passengers, it is entirely equitable that they should pay more for that flexibility[32].

---

[31] Frank (1983), pp. 238-255. This depends critically on the existence of scale economies in provision of air transportation services, which is discussed in Botimer (1994), pp. 56-58.
[32] Borenstein (1983), pp. 115-116.

These considerations suggest not only that there are plausible reasons for charging different prices, but also that such prices are not necessarily discriminatory. There may be elements of both *differential* and *discriminatory* pricing in the airlines' pricing structures. It is not possible to identify from the information reviewed in this chapter the portion of observed price dispersion due to each.

2.4. Arbitrary Nature of Demand by Fare Class

Regardless of intent, the effect of this variable pricing mechanism is to distribute demand into several fare classes. As discussed above, the imperfect nature of this segmentation implies that each fare class will not be composed of a homogenous type of passengers. That is, passengers who take a particular fare product have heterogeneous demand functions. Passengers can and will switch to other fare classes if the fare originally targeted to them is not available.

Therefore, the demand for a particular fare class is not isolatable in theory -- it is not *independent* of demands in other fare classes. It is defined only with respect to and situated within the particular suite of other fare products offered in the market, by the same and competing airlines. Demand *by passenger type* is more plausibly independent than by fare class, since the characteristics which influence the respective demands in the former case differ by type. Thus, demand among Type IV passengers is most influenced by *service variables* like frequency, amenities on the flight, etc. while among Type III passengers the *price variable* is paramount. Even by passenger type, however, interdependencies exist: Types I and II are both sensitive to frequency and related service variables, and differ only according to price sensitivity.

This immediately creates a problem, because (as described in Section 1.1 above) forecasting and the seat optimizing algorithm are both done by fare class, and both assume independence among fare classes. One simple example of the revenue consequences of assuming independence indicates its naiveté: Suppose an airline has but two fare classes, full-coach and deep-discount with appropriate restrictions. The airline feels its current fare structure is inadequate, since it believes a large intermediate group of passengers unwilling to pay the full-coach price but unable to meet the deep-discount fare's restrictions is not being served.

This problem is remedied by introducing an intermediate fare product, with a moderate level of restrictions. Revenue estimates are performed assuming this is an independent market not

yet served. Ignored in the calculation are passengers previously paying full-coach but willing to make the tradeoff to the lower fare (which could significantly *dilute* revenues) and those previously purchasing the low-fare product but happier with the less onerous restrictions of the intermediate product (thus increasing revenues). Under most circumstances, the number of passengers stimulated by the intermediate fare product is insubstantial relative to the reallocations from existing passengers, so ignoring the latter in the name of independence will lead to grossly incorrect supply decisions.

In practice, recognizing and adjusting for the limitations of our independence assertion is sufficient for forecasting and seat optimizing. Generated forecasts by fare class based on historical booking data are valid, even though the demands are not independent, if the fare product context does not change between the historical and forecast period. If it does, appropriate adjustment of the historical database (Step 1b in Figure 1.2) can eliminate this difficulty.

Adjustment for the independence problem is further enhanced by the examination of passenger behavior under constraints. Forecasts of *unconstrained* demands by fare class are required by the seat optimizer. I will discuss methods to unconstrain demands in Chapter 3. This does not ensure that enough space will be allocated to serve every fare class: The seat optimizing algorithm will limit bookings on lower-valued fare classes if the forecaster indicates that many high-value passengers will arrive. Under the independence assumption, the denied low-value passengers simply do not travel. Realistically, when faced with the closure of their initially desired fare class, some will *sell-up* to an available, higher-priced fare class. This occurs precisely because the present fare structures are imperfect. Consideration of this possibility further improves revenues as this behavior may be *induced* by limiting seats available in low-value fare classes. I discuss methods to induce sell-up in Chapter 3.

III. Literature Review of Methods and Past Assessments

This chapter considers each of forecasting, detruncation, and sell-up. I detail the purpose and assumptions of each methodology, various formulation methods which have been advanced in previous work, and existing comparative studies between models. Differences with the PODS simulation approach to comparing the effects of differences in these input methodologies will be discussed in Chapter 4.

3.1. Forecasting

3.1.1. Types of and Issues in Forecasting

3.1.1.1. Forecasting Applications in the Flight Planning Process

Forecasting is defined as the use of some systematic procedure (e.g., judgment, a "rule of thumb," or mathematical technique) and historical data describing some process to predict how it will be realized in the future. Although this thesis' forecasting section focuses on applications specific to the seat inventory control process, there are many other airline-specific uses for forecasting tools.

Figure 3.1 details some of these applications and their ideal period of relevance for a single flight departure. Differences in application require the selection of the appropriate forecasting technique, which I discuss below. *Strategic planning* involves long-term assessments of the airline's objectives, e.g., what kind of route structure it should have, which specific markets it should serve, and which aircraft it should serve these routes with. At this stage, macro-forecasts describing long-term economic and social conditions are most appropriate. The results of the strategic planning exercise will decide whether or not a particular flight should be offered.

Approximately a year before initiation of the flight, *budgeting* involves the coordination of financing to meet forecast expenditures. Here micro-forecasts which describe the expected financial costs incurred with introduction of a specific flight, facilities, etc. are required. Preliminary *aircraft assignment* also occurs at this stage, and includes forecasts of expected fleet size and aircraft availability at the time of departure, and medium-term expected economic and seasonal trends affecting flight demand.

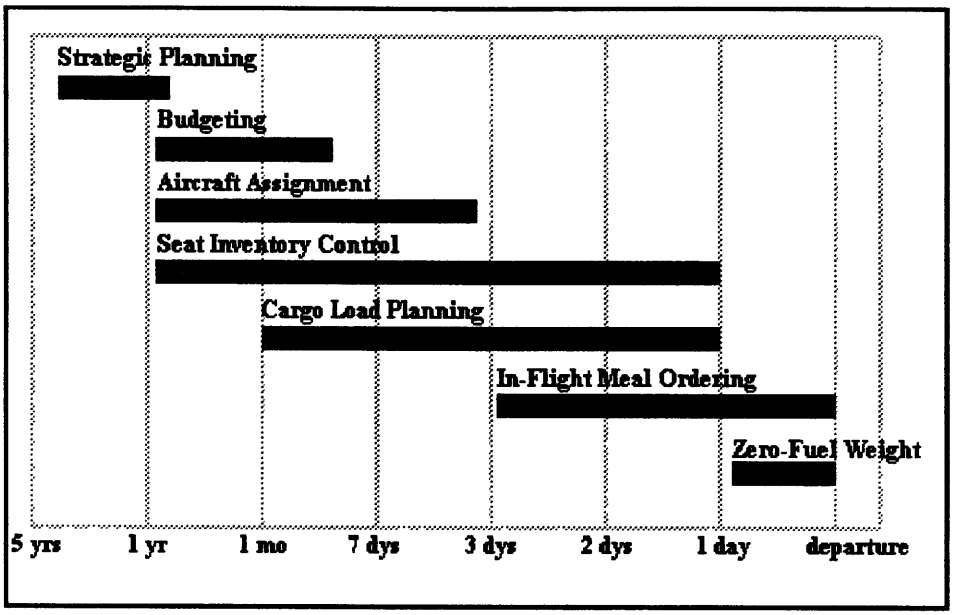


Figure 3.1. Ideal Flight-Specific Timeframes for Airline Forecasting Applications[33]

This stage also includes the initiation of the seat inventory control process for this flight. Booking limits are somewhat uncertain at this stage, since final aircraft assignment may vary with the results of the aircraft assignment exercise. Forecasts for this process are flight-specific estimations of total bookings to come. Other short-term processes requiring forecasts as departure date approaches include *cargo load planning* (which can be done only when aircraft assignment has been fixed due to the variety of cargo pallets used by aircraft, total weight restrictions, etc.) and *inflight meal ordering*, based on immediate-term forecasts of expected show-ups for this particular departure.

The time frames illustrated for this hypothetical flight are generalizations, and are often collapsed if opportunity presents itself. Thus, an airline may decide to immediately serve a market that has been opened via a new bilateral agreement, or that a competitor has suddenly ceased service in due to reallocation of resources to other markets, bankruptcy, or grounding. In this case, most long-term forecasting of market demand (part of the strategic planning exercise) is perfunctory or ignored, while aircraft assignment and seat inventory control are compressed into the available time before the first flight departure.

[33] Figures 3.1 and 3.2 are from Wickham (1995).

3.1.1.2. Forecasting Types and Techniques

Taneja divided forecasting methods used in airline industry applications into three groups by type of analysis: *qualitative* (judgmental), *quantitative* (using mathematical methods), and *decision analysis* (combining the first two methods). Example techniques for each method appear in Figure 3.2. The chosen technique varies by the application, technical expertise, and resource constraints.

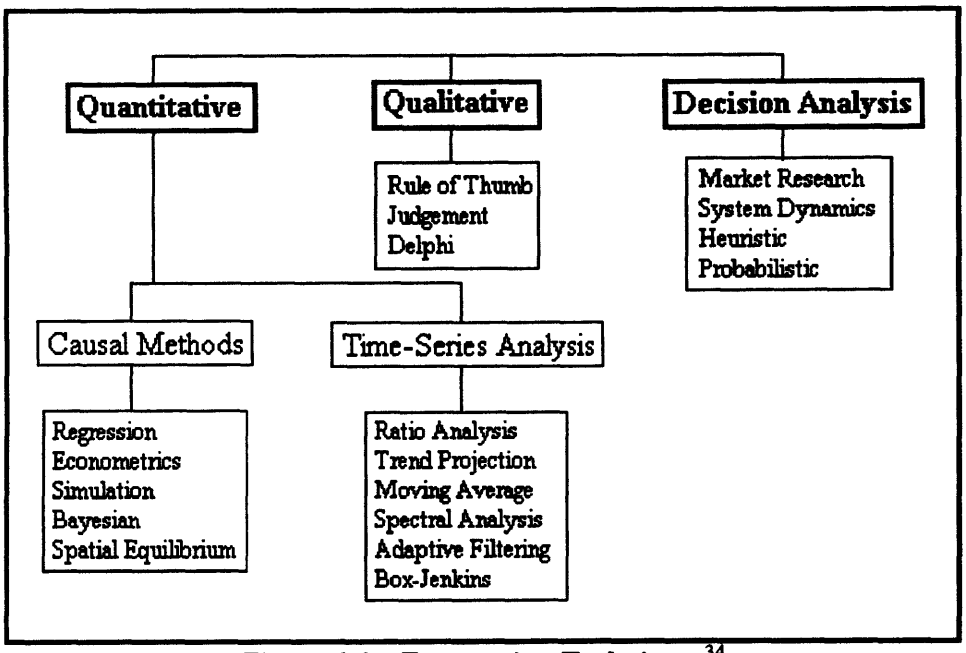


Figure 3.2. Forecasting Techniques[34]

Lee segregated airline forecasting methods according to application[35]. *Macro-level* forecasts detail expectations of large-scale global, national, and regional passenger flows dependent on socioeconomic conditions. Such forecasts are used for strategic planning purposes and do not break out market shares by airline or forecast at the city-pair level. Taneja[36] and Kanafani[37] have written the primary academic treatments of macro-level airline forecasting. The former book discusses regression for national and total airline traffic forecasting, while the latter focuses on forecasting strategy for aggregate airline travel activity, stratified on variables such as

[34] Wickham (1995), pp. 34-35.
[35] Lee (1990), pp. 47-49.
[36] Taneja (1978).
[37] Kanafani (1983).

trip purposes, length of haul, etc. Governmental and quasi-governmental organizations like the Federal Aviation Administration[38] and the International Civil Aviation Organization routinely publish forecasts of airline travel between regions.

Lee's second category, *passenger choice modeling,* involves the allocation of traffic in a given market to the individual alternatives (which may be travel modes or, in the case of the airline industry, individual airlines) offering service in the market. The most common approach to this problem involves logit models which attempt to quantify the (dis)utility a person derives from a trip on the various alternatives, and assigning a probability of travel choice accordingly. The general reference in transportation on this and other "discrete choice modeling" techniques is Ben-Akiva and Lerman[39]; airline-specific examples include Hansen[40] and Alamdari and Black[41]. *Micro-level* forecasting, the final category, involves exactly the concern of this thesis -- forecasting passenger demands by flight, date, and fare class. I cover previously proposed micro-level techniques in Section 3.1.2 below.

Regardless of the technique chosen or the process of interest, every forecast will have a *period* (the unit of time over which the forecast is produced), a *horizon* (the time span covered by the forecast, divided into a number of periods), and an *interval* (the time period over which forecasts are revised). While the forecast interval is typically coterminous with the forecast period (so that forecasts are revised at the end of every period), this is not necessarily so. For the purposes of seat inventory control, the interval and period are equivalent and span the *booking interval* as defined in Section 1.1. The horizon is equivalent to the *booking process* for the flight of interest.

3.1.1.3. Time Series versus Causal Quantitative Forecasting Techniques

This study will exclusively consider quantitative forecasts at the micro-forecasting level. Such techniques make extensive use of data from previously departed flights which are "similar" and predict what total realized bookings should be for the flight of interest. Quantitative methods are divided into two types. *Time-series analysis* examines the movement of the dependent

---

[38] FAA (1995).
[39] Ben-Akiva and Lerman (1985).
[40] Hansen (1990).
[41] Alamdari and Black (1992).

variable of interest over time, and describes the nature of the stochastic variation in the process under study. It assumes that whatever trends or variables influenced the progression of this variable on previous departures will continue into the future. Time series techniques are also referred to as "historical bookings" models, since they use booking information from past departures to infer the booking levels for a flight.

A major difficulty with time-series analysis is that there is no explicit consideration of which factors influence demand levels, so a sudden change in economic or system components influencing the flight will lead to a very inaccurate prediction. *Causal methods* remedy this problem by examining historical data on possible influences of demand to identify relevant factors, which are then weighted according to their importance in affecting historic demand. The most common causal methods are regression and associated econometric techniques, and are also called "advance bookings" models because most use information about $BIH(i)_f$ (bookings already received by interval i for the forecast flight f) to infer eventual total bookings.

All else equal, causal methods should be superior because they consider factors determining demand. However, they have a significant drawback: an extensive database of the many relevant variables must be kept to predict future demands. At any given time, an airline must keep data on all flights which are at any stage of the booking process, so the requirement is for the number of daily flights multiplied by the length of the booking process. With up to 2,000 departures daily and a booking process of up to a year in length, major U.S. airlines require an active historical data base of up to 730,000 flights at any one time. Brevity in the historical data base (HDB) required therefore becomes an important factor in the selection of forecasting mechanisms. A causal model with many independent variables will not be considered if it is impracticable or expensive to obtain values of these required variables on each of an airline's flights.

For the seat inventory control process, airlines typically utilize advance bookings models which use as independent variables only information *internally* generated by the process (i.e., bookings in hand data). Such models are *not* strictly causal, since bookings up to interval i or $BIH(i)$ information does not "*cause*" total bookings received before departure or $BIH(0)$. Rather, both are influenced by the same factors. These "non-causal advance bookings models" are an appropriate compromise, given the short-term nature of seat inventory control process (where

causative factors are less likely to vary significantly) and data storage requirements of true causal models.

3.1.1.4. Forecasting in the Seat Inventory Control Process

The forecasting step of seat inventory control (Step 2 of Figure 1.3) is further detailed in Figure 3.3. Before the beginning of each of the N intervals in the booking process for each flight and fare class, a forecast is generated to determine the total expected bookings $B\hat{I}H(0)_f$ which will materialize before departure on flight f, assuming that demand is not constrained by booking limits. If this is the *first* forecast for this flight (and therefore performed before the *Nth* interval, since we index backwards), there are no bookings yet on this flight. Therefore, the forecaster only has previous total bookings *BIH(0)* from the last departed n flights usable to predict $B\hat{I}H(0)_f$ for this flight. Here n is the choice of the number of past observations to use; some airlines and our PODS experiments assume $n = 52$. Not all n flights need have equal weight: some forecasters exponentially weight most recent departed flights, on the justification that these flights were influenced by conditions most likely to still apply to the forecast flight.

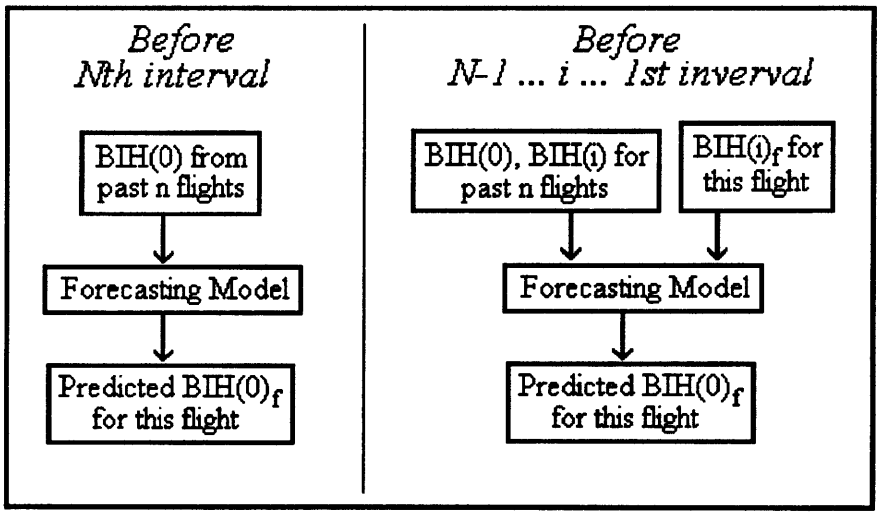


Figure 3.3. Forecasting in the Seat Inventory Control Process (Step 2 of Fig. 1.3)

For all other intervals i where $i \in (N-1,\ldots,1)$, information available to the forecaster includes: bookings $BIH(i)_f$ which have already materialized for this flight up to interval i; incremental bookings *BIH(i)* until i for the n flights in our historical database; and total bookings

BIH(0) for the *n* historical database flights. Some forecasters do not use all this information to predict final bookings $\hat{BIH}(0)_f$ for this forecast flight[42].

3.1.1.5. Assumptions of Forecasting

Forecasting methods in the seat inventory control process assume that the process which is to be predicted -- demand by fare class for a given flight -- has certain properties. These presumed characteristics also apply to the historical data set from which the forecast will be made, and underlay the theoretical treatment of the data. Formally stated, these assumptions are:

- Demands are segregated by fare class and are mutually independent
- Demands by fare class are not constrained by booking limits
- Demands are normally distributed
- Cancellation rates are similar between HDB and forecast flights

I have briefly considered the independence assumption and possible adjustments for observed interdependence of demands in Section 2.5 (analysis of sell-up will be discussed in Section 3.3 below). The second assumption that historical data are unconstrained -- i.e., that bookings were not artificially suppressed because the booking limit for a particular class was reached -- is central to the forecaster's unbiased operation[43]. Predictions based on constrained data will always be too low, since only those flights with high demand reach booking limits (see Section 3.1.3.2 below). Low forecasts lead to inadequate protection levels for higher classes, in turn causing *yield dilution.* Seats which could have been sold to late arriving high-value passengers are taken early by those with lower values (detruncation methods will be discussed in Section 3.2). The unconstrained assumption also applies to the bookings for the present flight -- if a fare class has a constraining limit, materialized demand will be less than predicted $\hat{BIH}(0)_f$. No distortions are introduced by this assumption: the seat optimizer decides how many seats to offer to each fare class, and in constrained situations will limit bookings on low-value fare classes.

---

[42] Note $BIH(0)_f$ and passenger loads are *not* equivalent. *Pax Load = BIH(0) - DB - NS,* where *DB* are denied boardings and *NS* are no-shows. The forecaster does not predict loads net of *DB* and *NS* since these are separately analyzed by the overbooking model (see Step 4 in Figure 1.3).

[43] A few forecasters include internally a detruncation mechanism, in which data from the HDB may be constrained.

Some distributional form for passenger demand must be assumed by detruncation schemes (it is otherwise impossible to extrapolate from data constrained at a booking limit), and by the forecaster (because parameters describing the data must be estimated for input into the seat optimizer). Analysis of reservation patterns indicates that the normal distribution of demand typically holds for flights of low, moderate, and high bookings, with some positive skewness for low-demand flights and a "spike" in the demand distribution at capacity level for high-demand flights[44]. The normal distribution is assumed in virtually all seat inventory control processes used by airlines, and will be followed throughout this study.

A final assumption of most forecasters is that cancellation rates do not vary between the flights in the HDB and the flight *f* being forecast. If a forecaster uses any information from HDB flights or *f* containing reservations which are later canceled (*BIH(i)* is an example) to predict final demand, it is utilizing *gross* measure/s to predict $\hat{BIH}(0)_f$, a *net* measure[45]. Distortions will not occur if cancellations on the predicted flight *f* follow the same proportionate gross/net relationships as occur for HDB flights. Thus, forecasters "predict" a certain proportion of cancellations for the forecast flight on the basis of previous demand. Underprediction of cancellations causes more than expected final bookings, and vice versa.

3.1.2. Literature Review of Available Forecasting Techniques

This section reviews academic and industry literature about forecasting techniques and uses an example HDB booking data matrix to pinpoint the data used by the various models reviewed. I classify the models into three groups, which are defined as follows:

- Historical Bookings: Bookings data from flights in the HDB are input as predictors of the unknown increase in bookings on the forecast flight from the forecast interval until departure.
- Advance Bookings: Bookings-in-hand data from the forecast flight are input as predictors of the unknown increase in bookings on the forecast flight.
- Combined: Both bookings data from the HDB and bookings-in-hand data from the forecast flight are used to predict the unknown increase in bookings on the forecast flight from the forecast interval until departure.

[44] Lee (1990), pp. 122-135. Lee discusses some studies suggesting that airline bookings are log-normal and gamma distributed.

[45] This is true because cancellations by definition cannot occur after the end of the booking period. However, passengers can *no-show*.

Only that data *directly used* to infer the *unknown bookings increase* is relevant for this categorization. For example, a model which takes HDB data to estimate model parameters and then bookings-in-hand data to generate a forecast is an *advance bookings* model, not combined: In this case, HDB data only establishes model parameters.

3.1.2.1. Historical Bookings

The simplest historical bookings approach is the arithmetic mean of historical bookings at the end of the booking process, calculated over selected departures from the HDB. Scandinavian Airlines proposed this basic model in a paper that also addressed related issues like the quantity of historical data necessary for accurate forecasting and outlier removal[46]. This paper is closely related to the work of Duncanson[47], who incorporated improvements to the basic model including exponential smoothing to disproportionally weight the most recent departures and seasonality adjustments. However, his focus was toward stable European markets in the 1970s, and the forecasting horizon was only three months. Neither model broke down forecasting to the fare class level.

Formalized versions of these simple or exponential bookings models by fare class are given by Wickham[48], and are represented by equations (3.1) and (3.2) below, respectively.

$$\hat{BIH}(0)_f = \frac{1}{M-t} \cdot \sum_{i=f-M}^{f-t} BIH(0)_i \tag{3.1}$$

$$\hat{BIH}(0)_f = \sum_{i=f-M}^{f-t} \frac{\alpha_i}{M-t} \cdot BIH(0)_i \quad s.t. \quad \sum_{j=f-M}^{f-t} \alpha_i = 1, \alpha_{f-M} < \ldots < \alpha_{f-t} \tag{3.2}$$

where $\hat{BIH}(0)_f$ is final bookings on day 0 of the flight f being forecast
M is the number of flights considered in the forecast plus the number of flights leaving before f but not yet departed[49].
t is the booking interval from which the prediction is being made.

[46] SAS (1978).
[47] Duncanson (1974).
[48] Wickham (1995), pp. 45-46.
[49] Therefore, M - t is the number of HDB flights used by the forecaster.

The booking information used by these methods is illustrated by the booking data matrix given in Figure 3.4 below. Our objective is to predict the heavy boxed cell, or the total bookings $\hat{BIH}(0)_f$ at the end of the last interval *0* before flight day (*FlD*) on flight *f*. At this particular snapshot in the booking process, flight *f* has two booking intervals before departure. Assuming that booking intervals are equally long, this means flight *f-2* took off today, yielding its final bookings and departing load, or $BIH(0)_{f-2}$ and $Load_{f-2}$. Flights earlier than *f-2* have a complete set of information, while those leaving after have not yet departed and so only have bookings in hand information up to the last booking interval they have traversed. The mean historical bookings models described in equations (3.1) and (3.2) make use only of information contained in the shaded block *A* of Figure 3.4.

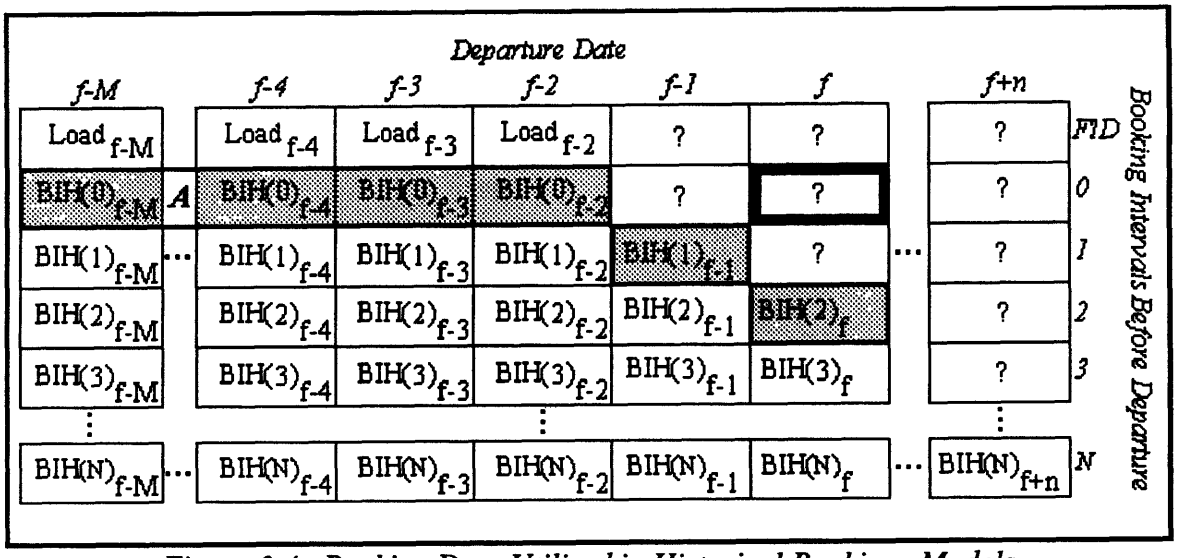


Figure 3.4. Booking Data Utilized in Historical Bookings Models

Sa also proposed a historical bookings model by calibrating two Box-Jenkins' ARIMA (Auto-Regressive, Integrated Moving Average) models for a single fare class on a particular flight[50]. This process makes use of autoregressive and moving average models and the same block *A* of information in Figure 3.4 to estimate parameters that minimize the squared difference between the actual and estimated time series. His results yielded high standard errors, reflecting high inherent variability in the data. The poor results led Sa to abandon the ARIMA approach.

Completing the examination of historical bookings models are *pickup* or "historical moving average" methods, which estimate the average increase in bookings from the interval of

[50] Sa (1987), pp. 75-82.

analysis t to the last booking process for selected HDB flights[51]. Equations (3.3) and (3.4) give the *classical* pickup models for equal and exponential weighting, respectively. Figure 3.5 indicates the bookings data used by the classical pickup model.

$$\text{(3.3)} \qquad \hat{BIH}(0)_f = \frac{1}{M-t} \cdot \sum_{i=f-M}^{f-t} \left(BIH(0)_i - BIH(t)_i\right) + BIH(t)_f$$

$$\text{(3.4)} \qquad \hat{BIH}(0)_f = \sum_{i=f-M}^{f-t} \frac{\alpha_i}{M-t} \cdot \left(BIH(0)_i - BIH(t)_i\right) + BIH(t)_f$$

$$s.t. \qquad \sum_{j=f-M}^{f-t} \alpha_i = 1, \alpha_{f-M} < \ldots < \alpha_{f-t}$$

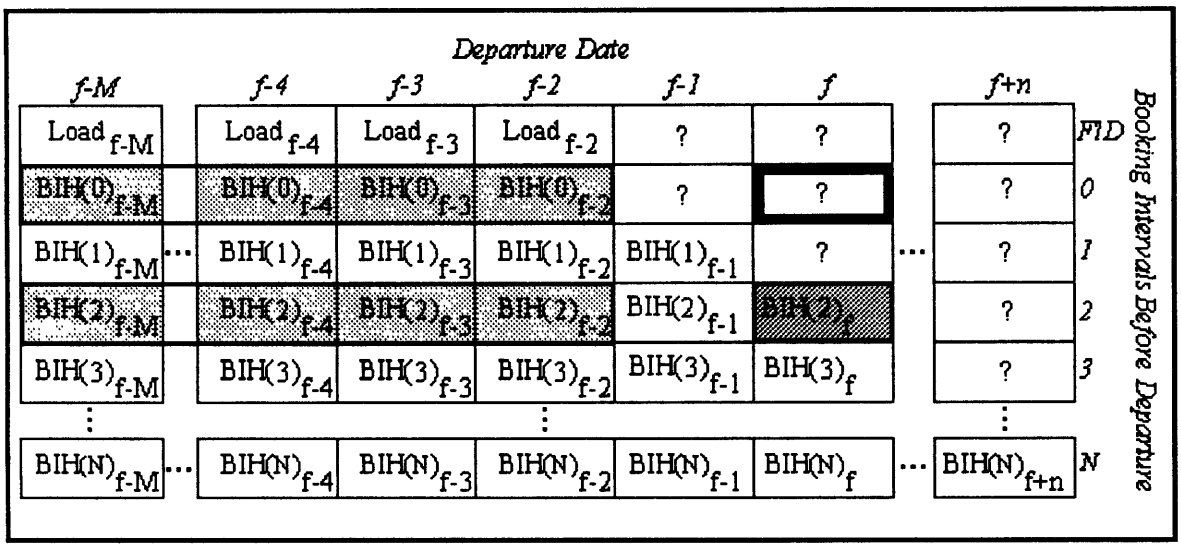


Figure 3.5. Bookings Data Utilized by Classical Pickup

The *advanced* pickup model takes into account information from soon-to-depart flights with incomplete booking histories. This modification, developed at Canadian Pacific by L'Heureux[52], is expected to respond to variations in demand more rapidly. As shown in equation (3.5), the advanced pickup model breaks up the pickup estimation process into t "pickup periods" measuring pickup in each time interval, with subsequent summing of results. This permits incorporation of all data from incomplete flights. The process is illustrated in Figure 3.6, where

[51] Pickup models are *not* combined because bookings-in-hand (advance bookings) is simply *added* to the estimate of pickup (derived from historical bookings) to derive the estimate of $\hat{BIH}(0)_f$. Bookings-in-hand are not used to infer bookings on the intervals t ... 0 which have not yet occurred.

[52] L'Heureux (1986).

advanced pickup uses a series of shifting, overlapping data "blocks" to estimate interval-by-interval pickup.

$$(3.5)\quad \hat{BIH}(0)_f = \frac{1}{M-t}\cdot\left[\sum_{i=f-M}^{f-t}\left(BIH(0)_i - BIH(1)_i\right) + \sum_{i=f-M+1}^{f-t+1}\left(BIH(1)_i - BIH(2)_i\right) + \ldots \right.$$
$$\left. \ldots + \sum_{i=f-M+t}^{f-1}\left(BIH(t-1)_i - BIH(t)_i\right)\right] + BIH(t)_f$$

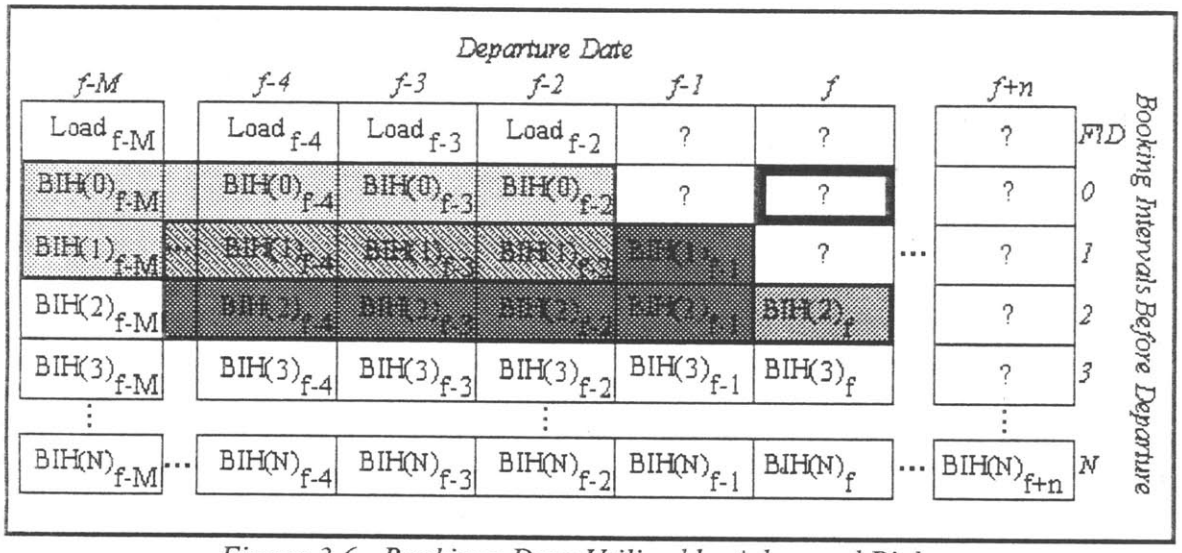


Figure 3.6. Bookings Data Utilized by Advanced Pickup

3.1.2.2. Advance Bookings

Harris and Marucci of Alitalia developed a simple advance bookings model providing forecasts by class of *aggregate* bookings for groups of selected flights on the basis of two data sets: One contained "snapshots" of bookings for the flights in question at five different points in their booking history, and the other described the total booking levels on *all* of Alitalia's flights for a 45-day time period[53]. This aggregation (necessary because of the less-than-daily frequency of most of the company's flights) significantly limits applicability to present forecasting problems, since specificity and sensitivity to variation in particular flights is lost. Additionally, in the U.S. airline industry most routes have at least daily frequencies.

A modified expression for a general regression model proposed by Lee[54] is given in

[53] Harris and Marucci (1983), p. 186.
[54] Lee (1990), pp. 108-110.

equation (3.6). The regression includes three groups of terms: the first for bookings already obtained for the given flight at time *t*, the second for exogenous causal factors, and the last for the random error term in all regression models. This regression model was not calibrated on any data set, nor were possible exogenous variables ***W*** specified. A reduced version of this model utilizing only the *BIH(0)* and *BIH(t)* data from HDB flights was proposed and tested by Wickham[55]. A non-causal regression model is used in Wickham's formulation, since *BIH(t)* is not a variable which "causes" *BIH(0)* data -- both are the results of a common set on exogenous causes.

$$\hat{BIH}(0)_f = \sum_{i=N}^{t} \vartheta_i \cdot BIH(i)_f + \boldsymbol{g} \cdot \boldsymbol{W}(f,i) + v(f,i) \tag{3.6}$$

where ϑ_i are the coefficients on *BIH* in previous time periods
$\boldsymbol{g}$ is a vector of coefficients on exogenous factors
$\boldsymbol{W}$ is a vector of exogenous factors
v is a random error term

The HDB data used for advanced bookings models is detailed in Figure 3.7 below. Blocks ***B1 and B2*** form the endogenous part of the dataset on which the regression is estimated. Only fully departed flights (in our example, up to *f-2*) are included in this estimation. The estimated relationship between the ***B1*** variables and the final *BIH(0)* data in ***B2*** is then used (in concert with exogenous variables) to predict $\hat{BIH}(0)_f$ given the partially complete booking history of flight *f* in block *A*.

An unusual advance bookings model was proposed by Lee, who suggested that the airline booking process was Poisson distributed with a certain probability of a request or cancellation materializing within a specified interval. Lee develops a censored Poisson forecasting model based on this assumption which may be estimated with maximum-likelihood techniques. The development of a censored model thus incorporates the detruncation process into the forecasting mechanism. Formulation of this model is complex and will not be described here[56]. The Poisson approach adopted by Lee assumes that the request probability, cancellation probability, and booking limits for the fare class are all constant within the intervals being forecast[57].

---

[55] Wickham (1995), pp. 47-48.
[56] The necessary equations are 7.6, 7.9, and 7.10 in Lee (1990), pp. 182, 184-185.
[57] Lee (1990), pp 84, 178.

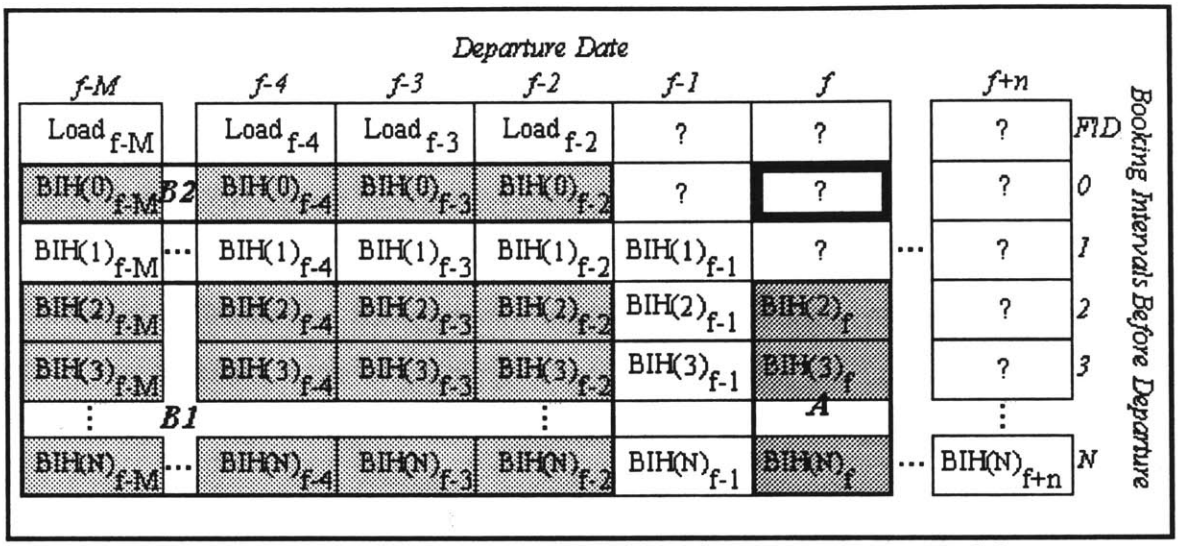


Figure 3.7. Bookings Data Utilized in Advance Bookings Models

Two issues complicate his approach. First, it is computationally intensive. A maximum-likelihood estimation (MLE) procedure in two variables is required for each flight, between each interval. A typical large U.S. airline with 1,500 flights a day, 15 intervals in the booking process, and a booking process of 100 days would have to perform, on average, about 10,500 MLEs a day. Decreasing the number of intervals to reduce computation requirements is infeasible because the three *intra-interval* assumptions underlying the Poisson approach are increasingly less likely to hold -- besides reducing revenues by being slower to reoptimize for developing traffic trends in advance bookings for the forecast flight.

Second, the assumptions about arrival rates, cancellation rates, and booking limits are unrealistic. Lee's adoption of three intervals in his simulations suggests that the booking curve and cumulative cancellation probability is linear but for two "kinks" (see Figure 1.1 for representative booking curves)[58]. While it is true that booking limits BL_i as defined in Section 1.3.1 do not change within the booking interval i, Lee's definition of "booking limit" is actually Mx (maximum bookings in a fare class) which does[59]. The nested nature of most present seat optimization algorithms requires a booking in any fare class to *decrement* the booking limit not only in that class, but some or all *other* classes. In these circumstances, the censored Poisson model's assumption of constant "booking limits" is a requirement that no request occurs in any

[58] Lee (1990), p. 186.

[59] Recall from Section 1.3.1 that maximum allowable bookings Mx within booking interval i is the booking limit BL_i on that fare class less bookings on *other* fare classes which affect the fare class' SA.

other fare class which affects seat availability *SA* on this class. Thus, sufficiently small intervals must be chosen that *Mx* does not vary, i.e., no request occurs either in a lower-valued class or (if all fare classes' *SA* are decremented) any other fare class. Either condition easily increases the interval requirement for a typical flight to far above 30, further taxing computing capacity. Lee acknowledges that assuming one fare class ignores the nested nature of seat optimizers[60], but does not address the implications on intra-interval stability of *Mx*.

3.1.2.3. Combined Historical and Advance Bookings

There are also many models utilizing both historical and advance bookings data. Sa (disappointed with his time series results) calibrated causal regression models for bookings to come on a particular flight given *BIH(t)*, a seasonal index, day of week index, and a historical average of bookings to come or "pickup" between interval *t* and departure. Sa's method is distinguished from advance bookings models because both types of bookings data are used as variables to estimate $\hat{BIH}(0)_f$. In the advance bookings models, one data set (historical bookings) is used to calibrate coefficients which will be applied to the other (advance bookings). This effort was much more successful in predicting demands by fare class, however there were no tests of forecasting ability[61].

Ben-Akiva et. al. proposed a model for microforecasting by flight and class[62], combining a non-causative regression model for advance bookings and a time series model for historical bookings on previous departures. While the calibrated models fit the data set well, the analysis was done on a monthly basis due to data limitations. There were also no validation tests of forecasting ability on future flights.

Lee developed a non-causal "full-information" model combining weighted final bookings information from flights already departed, *BIH(i)* for flights not yet departed, and $BIH(i)_f$ for the forecast flight *f*[63]. A modified version of his formulation is given in equation (3.7) below. The full-information forecasting model takes advantage of the fact that flights with incomplete booking histories (flights which have not yet departed) are likely to reflect recent changes in

[60] Lee (1990), pp. 93-94.
[61] Sa (1987), pp. 83-108.
[62] Ben-Akiva et. al. (1987).
[63] Lee (1990), pp. 115-117.

demand conditions which will have the strongest effect on bookings for the flight being forecast. Thus, the coefficients in equation (3.7) should be ranked as $\varphi_i < \psi_i \leq \lambda$. Lee's model makes use of all the shaded squares in Figure 3.4.

$$(3.7) \qquad \hat{BIH}(0)_f = \sum_{i=f-M+t}^{f-t} \varphi_i \cdot BIH(0)_i + \sum_{i=f-t+1}^{f-1} \psi_i \cdot BIH(t+i-f)_i + \lambda \cdot BIH(t)_f$$

where φ_i are coefficients on total bookings for flights in the HDB
ψ_i are coefficients on most recent *BIH* information for flights which have not yet departed
λ is a coefficient on $BIH(t)_f$ data for the flight in question f at interval t in its booking process

Lee's formulation of the "full-information" model combines forecasting and detruncation in a *recursive substitution* method whereby final demand $\hat{BIH}(0)_f$ is derived by interval-by-interval demand estimation until interval *0* is reached[64]. He uses the same MLE procedure as the Poisson-based model to estimate demand for *each* of the intervals. Unfortunately, this makes it even more computationally intense. An airline must now compute N - t MLE estimates (where N is the total number of intervals and t the interval from which we are forecasting) for each flight, for each interval t, for a total of $N!$ MLE computations per flight[65]. For the typical major U.S. airline case I have described, this implies 42,000 MLEs per day! Additionally, the full-information model retains the assumption of invariant "booking limits" within each interval[66]. I will consider a type of "full-information" model which does not require MLE estimation in Section 5.2.3.

3.1.3. Critical Review of Past Comparative Assessments

Many of the authors I have reviewed above who have formulated forecasting models also evaluated alternative models on some metric, usually forecast accuracy. In this section I review these comparison techniques, and highlight the misleading nature of using measures of forecast error as a comparative metric or to evaluate forecasting performance.

3.1.3.1. Past Comparative Studies of Forecasting Methods

[64] Lee (1990), p. 190.
[65] This is compared to N computations per flight for non-recursive forecast estimation procedures.
[66] Lee (1990), pp. 187-188.

- *Sa: ARIMA Time-series and Regression*

Sa's comparison of ARIMA time-series versus regression models for short-term forecasting rested on goodness-of-fit tests, using bookings data from ten markets. The ARIMA models were dismissed on the basis of poor performance in *one* class on *one* of these markets. Subsequent regression models were estimated for all markets (with the same variable set), and significantly differed in overall model fit and statistical significance of certain coefficients. There were no tests of forecasting ability on a different data set for any of the fitted models, nor was information about time-series fitting on the nine other markets provided. Because airline reservations data are significantly disparate by market and comparative testing was incomplete, the study has limited implications about the relative desirability of the models tested.

- *Ben Akiva: ARIMA Time-series, Regression, and Combined*

The result of Ben Akiva et. al. was a *combined* model with a ARIMA time series component (using historical bookings) and a regression component (using advance bookings). When the two components were run separately, correlation coefficients between the predicted and actual observations declined; however, the regression model was found to fit the data better than the time series[67]. An absence of tests for forecasting ability on a different data set, the monthly nature of data, no consideration of the effect of booking limits on demand levels, and an aggregation of fare class data into two classes limit the applicability of Ben Akiva's conclusions.

- *Wickham: Pickup, Regression, Time-series*

Wickham's study obtained a database of airline booking histories by fare class and day of week, selected several forecasting methods, then evaluated each on measures of forecast accuracy. The forecasting models included classical and advanced pickup (utilizing both simple means and exponential smoothing of the HDB), time series (again with and without weights on the HDB), and regression based on *BIH(t)*. These tests were performed for a variety of HDB sizes and forecast horizons; additionally, a simple unconstraining mechanism was applied,

[67] Ben Akiva (1987), p. 12.

allowing accuracy comparison with forecasts performed using the constrained data[68]. His principal conclusions were as follows[69]:

- Pickup consistently outperformed regression and time-series models.
- Increasing forecast horizons (the distance between the present t and departure date of the flight) favored the advanced pickup model with exponentiation.
- Advanced pickup was more sensitive to dramatic shifts in demand.
- All models had *positive* biases; forecasts overpredicted demand. This was attributed to a positive bias in demand variability.
- Increasing the size of the HDB did not significantly affect model performance.

One primary advantage of this comparative study was its validation of results on an independent data set. Wickham's in-depth consideration of alternative model specifications and various measures of forecast error make his thesis the most complete evaluation of alternative forecasting models on an *accuracy* basis presently available. There were a few minor detractions. First, two fare classes were selected from an airline's database were selected without analysis of whether the relative performance of forecasters might differ according to this selection[70]. The data was aggregated over 24 distinct markets without explicit consideration of characteristics like stage length or dominant passenger type. Finally, positive bias was correctly identified but attributed (probably mistakenly) to outliers. Section 3.1.3.2 below will discuss this issue.

- *Lee: Censored Poisson, Full-Information, Regression, and Pickup*

Lee's thesis on the airline reservations forecasting process compares his censored Poisson and full-information models with a regression and pickup model. An airline dataset for a single market was chosen, with a forecasting period of two months and an HDB of approximately nine months. All forecasters had only three intervals, with forecasts at 60, 28, and 14 days from departure[71]. Class-specific forecasts were generated and the chosen forecasting models compared on the basis of three measures of forecast accuracy. His results indicated that the full-information

[68] See Wickham (1995), p. 59, for a discussion of which specific scenario combinations were tested.
[69] Wickham (1995), pp. 113-117.
[70] Wickham (1995), p. 56.
[71] Lee (1990), p. 192.

model usually was most accurate, followed by the censored Poisson, regression, and pickup models[72]. Expanded tests considering several markets and other fare classes were performed on the pickup and full-information models; the three booking interval restriction was kept. These tests confirmed his initial results: Pickup was decidedly poorer than the full-information model[73]. Such results indicate the power of Lee's models despite the unrealistic assumptions about constancy of "booking limits" and request/cancellation rates.

However, this comparison of error does not address the tradeoffs involved between accuracy and updating. An airline faced with computing constraints might utilize the full-information model but update forecasts less frequently than if a non-recursive, non-MLE forecaster (e.g., pickup) were used. It is not clear that *more* accurate forecasts *less* often is preferable to less accurate forecasters more often (thus taking advantage of developing $BIH(i)_f$ information). A better procedure would compare load and revenue between Lee's advanced models with few booking intervals, and traditional models with frequent updating.

- *Lee: Revenue Cost of Forecast Error*

Lee's thesis comes closest to the present PODS effort by examining the *revenue* effects of forecasting. Here the revenue cost of forecasting error was estimated via simulation. Revenues for a particular flight obtained with the EMSRa seat optimizing algorithm are computed when *forecast demand* by fare class varies by some positive or negative proportion from actual, materialized demand. This is compared with the revenue result when the forecast demand is exactly correct, and differences are attributed to forecasting error. The procedure is repeated for proportionate errors in measuring the *standard deviation* of demand. These tests are performed under a variety of demand levels.

All fare classes are assumed to have equal proportionate variation. Lee constructs a simplified simulation of the seat inventory control process in which all demand arrives at once, in increasing order of value; there is no competition and no interdependence of demand[74]. Results indicated the following:

---

[72] Lee (1990), pp. 199-202.
[73] Lee (1990), pp. 204-215.
[74] Further information on Lee's simulation methodology is found in Lee (1990), pp. 242-245.

- Varying forecast *standard deviation* had little impact on revenues, even at extreme demand levels.
- Varying forecast *mean demand* significantly impacted revenues in high demand conditions. Larger revenue drops were exhibited when the forecast exceeded the actual demand (overprotection) than when it was less (underprotection).

A simplified version of Lee's revenue loss graphs appears in Figure 3.8 for medium and high demand levels[75]. Increased revenue losses with *overprotection* were attributed to the fact that seats protected for high value passengers would go empty, while low-value passengers would be denied. By contrast, with underprotection seats are still filled (since by assumption demand levels are uniformly high for each fare class) but with more low-value passengers, who fill up spaces which are later denied to high-value passengers[76].

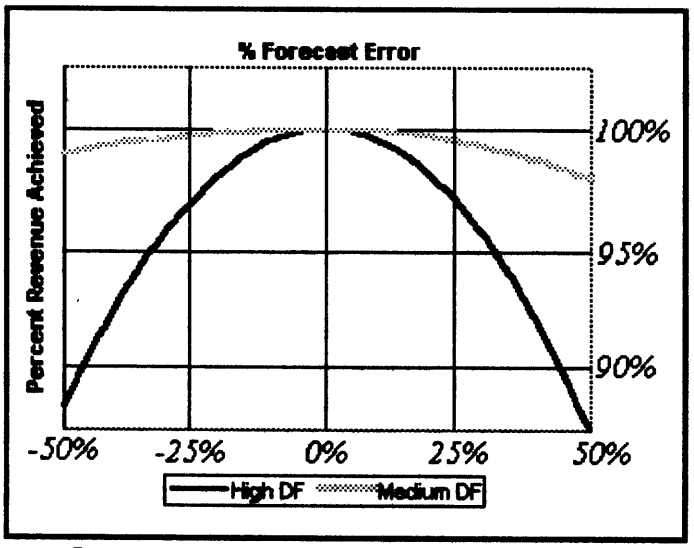


Figure 3.8. Revenue Losses as a function of Forecast Error, High and Low Demands

This conclusion is dependent on the assumption of independent fare class demands. Since demand is interdependent, a passenger denied a seat in a low-value class will often *sell-up* to a higher-value seat (see Section 3.3 below). Interdependence also applies to competitive interaction in the marketplace: passenger loads by fare class result from trade-offs between *all* airlines' fare products in the market; the process is not separable by airline. While these considerations qualify the *asymmetric* nature of revenue decline with forecast error, the general parabolic shape of the

[75] Lee (1990), p. 258.
[76] Lee (1990), pp. 253-254.

curves in Figure 3.8 (indicating greater-than-linear revenue declines with linearly increasing forecast error) is probably correct.

3.1.3.2. Problems Using "Forecast Error" in Comparative Assessments

- *Definition of inherent bias, and conditions under which it occurs*

Most troubling in all these assessments of alternative forecasting methods on error-based metrics is an unclear treatment of what base to use when measuring *error*, and an absent or incorrect analysis of biases created by the comparison methodology. Depending on the base used in error definition, the existence of constrained observations in the dataset, and the forecaster used, *inherent forecast biases* will occur. *Bias* is said to exist when the summed difference between predicted and actual bookings over all flights being forecast is not zero[77]. Failing to eliminate built-in biases confuses the ranking of forecasters: observed differences may be due as much to the construction of the comparative experiment as to inherent performance differences with other models. This analysis, while directly applicable to these comparative efforts, is also relevant to airlines desiring an unbiased measure of forecast error.

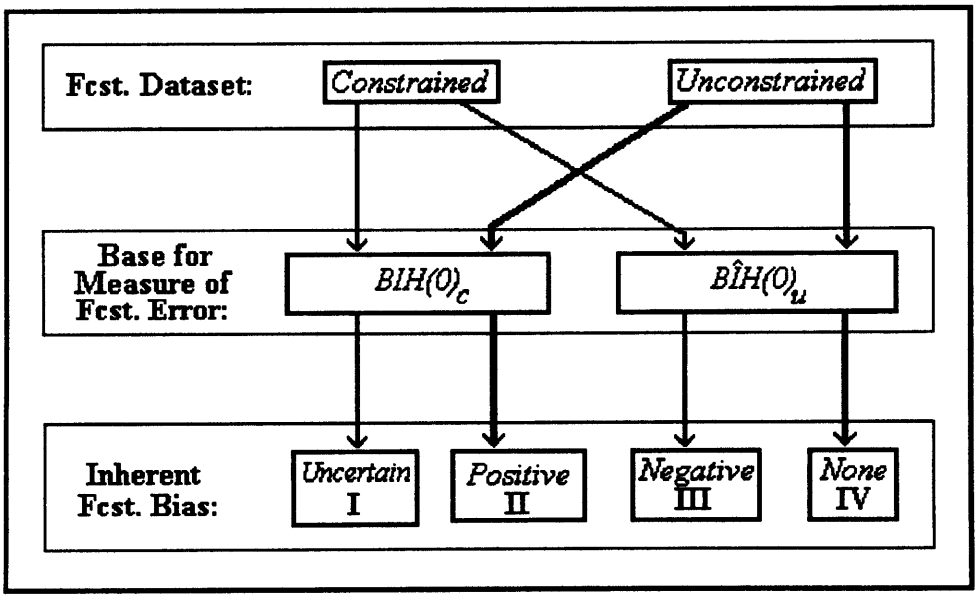


Figure 3.9. Inherent Biases in Measurements of Forecast Error

[77] Wickham (1995), pp. 64-65, 111.

First, most forecasters assume an unconstrained dataset[78], but all comparative analyses of forecasters except Wickham ignore the issue and assume their data are unconstrained. Second, the various error "metrics" comparing the *observed* versus *predicted* bookings[79] operate on one of *two* measures of observed bookings: either $BIH(0)_c$ (observed total bookings before departure *constrained* by closure occurring on some of the forecast flights) or $B\hat{I}H(0)_u$ (expected total unconstrained bookings; this must be estimated). The choices a forecast error analysis makes on these two issues creates four possible combinations which I divide into types, with expected bias as described as described below (see Figure 3.9).

Type I analyses (constrained data and a constrained error measurement) contain *uncertain* forecast bias. This is true because without detruncation, the treatment of censored observations relative to unconstrained observations varies according to the forecaster used. Consider Figure 3.10, which relates the time t before departure and total bookings received $BIH(i)_f$. The average relationship between bookings-in-hand and time is a "booking curve;" for flights which do not close in this example, it is $\boldsymbol{A}$ (the maximum bookings achieved on $\boldsymbol{A}$ flights is U_0, and $U_0 < BL$). A booking curve $\boldsymbol{B'}$ for a flight which exhibits high demand and closes at cl is also given (if there were no booking limit on these flights, the high demand flight's booking curve would be $\boldsymbol{B}$).

The classical pickup forecaster in equation (3.3) estimates the *average* increase in bookings from time period i to 0, which is $U_0 - U_i$ for unconstrained flights. Without detruncation of the historical dataset, the pickup on the constrained flight is $BL - C_i$. Since $BL - C_i < U_0 - U_1$ (that is, the slope of a line a drawn between bookings at i and 0 for unconstrained flights is *steeper* than for constrained flights b'), averaging the pickup on the constrained with unconstrained flights $\boldsymbol{A}$ *decreases* the classical pickup forecasts below U_0[80]. But average realized bookings $BIH(0)_c$ will always be between U_0 and BL, since closed flights have BL bookings. Therefore, under Type I conditions classical pickup has an inherent *negative* forecasting bias.

---

[78] Certain forecasters incorporate both detruncation and forecasting in one methodology. For the purpose of the typology in Figure 3.9, the forecast dataset in these cases is *unconstrained.*

[79] Wickham (1995), pp. 65-67 gives several alternative measures of *forecast error.*

[80] Note also that if $\boldsymbol{B}$ flights are detruncated, the resulting pickup forecast increases as expected, since $C_o - C_i > U_o - U_i$. Chapter 6 discusses this fact in more detail.

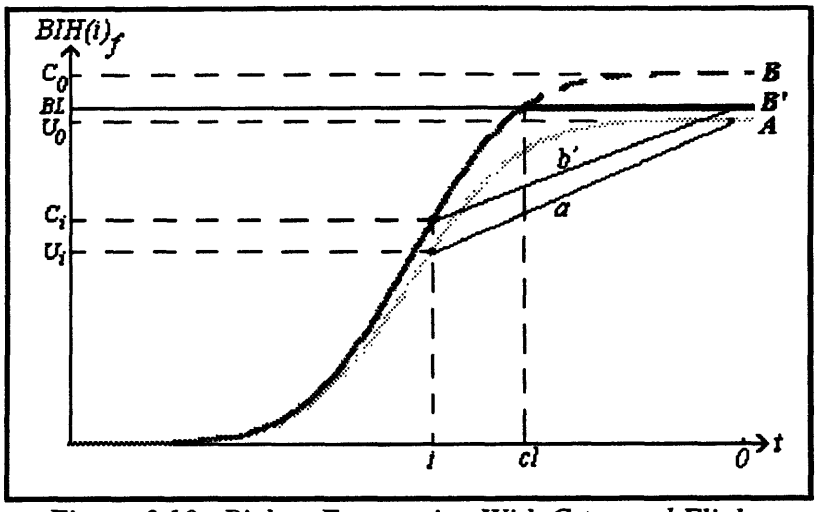


Figure 3.10. Pickup Forecasting With Censored Flights

Now consider the simple time-series of final bookings forecaster in equation (3.1). There is no inherent bias in predicting final constrained bookings $B\hat{I}H(0)_c$, since this forecaster does not treat constrained observations differently than unconstrained observations. Therefore, the inherent bias under Type I conditions is *uncertain*: it depends on the forecaster's treatment of the fact that constrained observations all have final bookings of *BL*. Determining the direction of forecast bias in these circumstances requires specific analysis of each forecaster tested.

Most comparative studies of forecasting (including all the studies examined above) are Type I because they use a constrained measure for error calculations and do not consider detruncation. For example, Lee recognizes the difficulties involved in assuming a detruncated data set. Because the formula for determining the unconstrained bookings in a fare class $B\hat{I}H(0)_u$ given an effective booking limit is "quite complex and ... difficult to apply," especially for his censored Poisson model, he excludes censored data from the data set when calculating forecast performance measures[81]. Wickham finds a positive bias in forecasts based on his *constrained* dataset. He attributes this to outliers, which may have merit (*all* forecasters had the bias) but does not consider whether each forecaster creates a *differential* bias[82].

Type II forecasting error analyses occur when an unconstrained forecast dataset is used in combination with a constrained measure of forecast error. A *positive* inherent forecasting bias

[81] Lee (1990), pp. 156-157. This leaves an "unconstrained" but downward-biased dataset.
[82] Wickham (1995), p. 111.

results because the forecast bookings are not constrained by booking limits, but observed final bookings $BIH(0)_c$ are necessarily limited by BL. The single comparative forecasting study to examine this issue is Wickham, who detruncates his dataset and reruns his model set on the detruncated data. He finds the forecasting models run on the detruncated data have a significantly higher bias than when run on truncated data, which is again attributed to outlier distortions. Here the increased bias is almost certainly due to Type II effects[83]. Some airlines often engage in Type II forecast error analyses in the course of monitoring the effectiveness of seat inventory control. As discussed extensively in this section, airlines forecast total bookings for a flight f on the basis of flights in an unconstrained HDB and/or advance bookings on the flight. This is then compared with actual constrained total bookings $BIH(0)_c$ for the forecast flight to yield *post facto* forecasting ability of utilized forecasters[84].

Under Type III conditions, the dataset includes flights with constrained data, but final booking data $BIH(0)_c$ for each flight is unconstrained by some method for the purpose of calculating error. Such a situation is unlikely, since it seems illogical to develop a detruncation technique to calculate forecast error and not use the same method to detruncate the data used to make forecasts. This circumstance causes a marked *negative bias* in predicted final bookings, because flights in the forecast dataset which close will be have constrained final bookings $BIH(0)_c = BL$ but detruncated "actual" bookings level will be $B\hat{I}H(0)_u > BL$. No comparative studies of forecasts were Type III.

Finally, Type IV error analyses involve a dataset which is unconstrained and a base for the measure/s of forecast error which is similarly unconstrained. This type of analysis is rarely performed; studies of forecast error routinely forget the purpose of the forecaster (to predict *unconstrained* demand by fare class) and suppose *constrained demand* is instead the goal. Forecast errors calculated under Type IV conditions yield the only comparisons free of methodology-induced bias. There is one significant disadvantage: because the base unconstrained bookings $B\hat{I}H(0)_u$ on

---

[83] Wickham (1995), pp. 111-112. Positive bias also results from a "double-counting" problem. Most detruncation procedures assume independence of fare class demands. Thus, a passenger denied space in a fare class because of closure is assumed not to sell-up to a higher-valued fare class. Since this often occurs, the passenger will be counted twice: First, as a denied passenger in the low-valued fare class, and again as an accomodated passenger in the higher-valued fare class.

[84] In this case the "forecast dataset" is the HDB and the comparison is between actual and expected bookings for a specific forecast flight based on information in the HDB -- as opposed to the comparative forecasting methodology case, where the comparison is between actual and forecast bookings on a set of flights within the forecast dataset. The analysis is the same. Note that airline forecast error calculations are not strictly Type II if bookings at time interval i for the flight $BIH(i)_f$ ever reach Mx and an airline does not unconstrain this measure over all i before departure.

which forecast error is calculated must be *estimated* via detruncation for every flight which closes, systematic biases in the detruncation method will similarly bias error calculations. Thus, a Type IV analysis is only truly unbiased if the detruncation method involved is unbiased, or if there are no constrained observations among the flights being compared. Of course, this complaint is also true of the other analysis types (II, III) which require detruncation.

- *Difficulty in achieving zero "forecast error"*

Error measurement and reduction is a desirable goal because (as demonstrated by Lee in Figure 3.8 above) only with zero forecast error can maximal revenue be achieved. However, even if definitional issues about forecast error are resolved, the achievement of zero error in actuality nearly impossible. The problem lies in assuming that an input with zero forecast error *results* in flight loads that exactly adhere to the "zero error" predictions, and thus maximize revenue. Such an event rarely occurs.

To illustrate, suppose a particular constrained flight f departs and the airline is certain that passengers on the next departure $f+1$ of this flight have *exactly similar* demand characteristics. After detruncation, the airline inputs as its forecast for the $f+1$ flight the expected unconstrained bookings based on constrained bookings from f. Does this "zero error" forecast result in a "zero error" result, with passengers materializing as predicted?

No, because the seat optimizer will adjust optimal booking limits to these inputs, causing different class closure properties, sell-up and/or lost passengers, and thus a different (constrained) observed booking pattern than f. The associated unconstrained booking level for $f+1$ will also be different. Hence, "forecast error" for flight $f+1$ will be nonzero despite perfect knowledge about the requests which will materialize on this flight, given present fare structures. There are two conditions in which this rule does not apply: First, if no fare class is constrained at this demand level; and second, if each passenger has been placed in the maximum fare class he or she is willing to pay and the airline effectively prevents dilution. In these two circumstances, bookings will materialize as expected: No closure occurs with low demands (and thereby no sell-up), and with effective segmentation no sell-up occurs in the face of closure because the maximum WTP of passengers has already been identified.

These comments bring us back to the issue addressed in Section 2.5 -- the arbitrary nature of "demand by fare class." It is theoretically problematic to compute forecast errors on fare class categories when the underlying generative process occurs by *passenger type* -- and a precise mapping between passenger *type* and fare *class* happens only in the two limited circumstances above. Otherwise, observed demands in each fare class are the result of the time in the booking class that lower-fare classes close, and the segmentation ability of each fare class. Such difficulties suggest that an emphasis on "forecast error" to compare forecasting methods is misguided. I argue that a more appropriate comparison is a *revenue analysis*, comparing the revenues an airline can expect to earn under a variety of different forecasters, given similar demand and competitive conditions. This avoids the methodological flaws involved in defining and measuring forecast error, and the intermediate step of abstracting from forecast error to revenue performance. I discuss the advantages of PODS for this purpose in Chapter 4.

3.2. Detruncation

3.2.1. Detruncation as a Step in the Flight-Level Forecasting Process

Unconstraining, detruncating, or uncensoring distributions refer generically to the process of estimating parameters of a distribution based on a sample from which certain values have been removed or *censored*[85]. In the airline-specific case, we refer to a situation when bookings-in-hand *BIH(i)* reaches *BL* during some interval *i*. Now the involved fare class is closed, and no more requests will be taken in this fare class unless a cancellation occurs or the seat optimizer determines to make more space available at the next reoptimization (see Section 1.3).

A truncated distribution of demand for a particular flight is shown in Figure 3.11. The random variable $BIH(0)_f$ is, as before, the total bookings received for the flight. Now suppose a maximum of *Mx* bookings will be allowed in this fare class. Then, if bookings are less than *Mx*, the maximum is never reached: the fare class will always be open. However, if *Mx* or more requests are received (which occurs with probability ***A***), only *Mx* will be carried. The probability distribution is therefore censored from above at *Mx*, and will have a "spike" at that point.

[85] In statistics a distinction is drawn between truncation (where offensive values are excluded entirely from the population, and thus not countable) and censoring (where the observations may be identified and counted) [Cohen (1959), p. 217 and Schneider (1986), pp. 1-2]. I use the terms interchangeably.

Without unconstraining the expected value of *BIH(0)* for this flight is $\overline{X_i}$, but actual mean demand for this fare class is $\overline{X}$. The objective of the various detruncation methods is to determine $\overline{X}$ given $\overline{X_i}$, *Mx*, and (often) supplementary information about the booking curve.

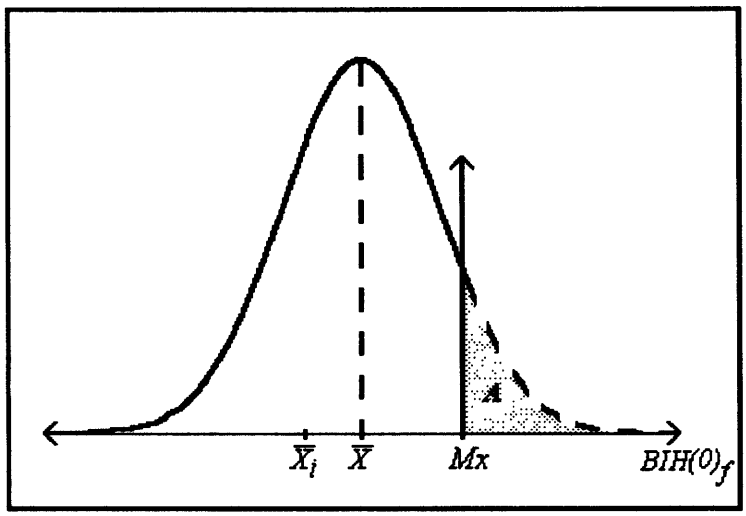


Figure 3.11 A Censored Demand Distribution

Not detruncating could theoretically have severe revenue consequences for an airline. This is due to the self-fulfilling nature of forecasting in airline seat inventory control. If a forecaster underpredicts demands for higher-value fare classes due to not detruncating (which will always occur, since $\overline{X} > \overline{X_i}$), more low-value passengers will be accommodated, leaving no room for high-value passengers who arrive later. This result becomes part of the HDB for future flights, depressing the high-value fare class forecasts further. Dilution becomes extreme as bookings of high-value passengers spiral downward, replaced by low-fare passengers. This worst-case scenario is mitigated to the extent that high-value passengers arrive early, since any indication of higher than expected bookings in a high-value fare class leads to increased protection levels at the next reoptimization. Further, demand levels are typically not so impacted and booking curves for high versus low value passengers are not so disparate (see Figure 1.1) that the plane would *fill* with the latter before the former arrived.

Another issue is the variability of *Mx* over the booking process. It is easily possible that a fare class closed at one point will later open (see Section 3.2.2 below). Theoretically, the HDB could include information on open/close dates for later incorporation into the detruncation

process. However, few methods have been developed to take advantage of this information. Most present detruncation methods and tests in this thesis assume that the flight is closed after the *first* closure until departure. This assumption generally leads to *over*-detruncation and therefore over-forecasting of fare classes with significant closure. Concerns about resultant *spoilage* (seats protected that go empty) are somewhat mitigated because closure occurs most often in lower-value fare classes, with few or no lower fare classes to "overprotect" against.

3.2.2. Literature Review of Available Detruncation Techniques

For flight-level forecasting, we have a situation similar to Figure 3.11: the object is to find the unconstrained mean $\overline{X}$ for final bookings $B\hat{I}H(0)$ among all selected HDB flights, given knowledge of $\overline{X_i}$, its standard deviation, the number of constrained and unconstrained observations, and *Mx*. This circumstance lends itself to maximum likelihood estimation (MLE). In statistical parlance, this distribution is singly Type-I censored from the right at *Mx*. The problem is solved (assuming an underlying normal distribution) without the exhaustive iterative process typically used in MLE problems[86] using a method developed by Cohen[87].

Alas, the detruncation problem in revenue management is complicated by two unfortunate facts. First, on the same flight *Mx* is not fixed over the booking process: The maximum allowable bookings for a fare class changes each time a booking is received or canceled in another fare class which affects this class' seat availability *SA*, and during reoptimizations of the seat optimizing algorithm between booking intervals when bookings limits *BL* are modified (see Section 1.3.1). Second, between flights there is even more *Mx* variability: *Mx* for each flight and fare class depends on *BL*, which is based on the expected forecast of $B\hat{I}H(0)$ for higher-valued fare classes. Obviously, total bookings *BIH*(0) on each flight varies by various trend and cyclical factors. Since the HDB to be detruncated involves multiple flights with *Mx* variation *between* and *within* flights, the distribution of *BIH(0)* is actually *multiply censored.* Solving for the unconstrained mean and standard deviation of $B\hat{I}H(0)$ in this case is *significantly* more difficult than the single censoring case. In fact, no simplified MLE prediction method has been developed for multiply-

[86] Details for the general MLE problem are found in Ben Akiva and Lerman (1985), pp. 20-22, 82-84.
[87] Cohen (1959).

censored normal samples[88]. Preliminary work on the density functions of multiply censored Type-I distributions indicates the difficulty in attempting a solution to this MLE problem[89].

Besides general references to MLE techniques, there is almost no literature on detruncation methods for the airline industry. Brummer et. al. attempted to solve the detruncation problem using an MLE technique, but the random variable was total bookings on the flight *over all fare classes*[90]. Such analysis does not consider *Mx* variability problems, nor the requirement in seat inventory control of forecasts by fare class. The only other reference to detruncation is Wickham. His selected model set (see discussion of Wickham in sections 3.1.3.1 and 3.1.3.2) is run twice on a database of flights, once constrained and once not. The unconstraining process circumvents the difficulties of MLE estimation by assuming that flights which close possess similar *proportionate* relationships between bookings at any interval i and total bookings at 0 as those which do not.

First, a representative booking curve is calculated for each fare class which has closed flights. An average $BIH(i)$ for all intervals i is calculated, selecting the R flights in the entire[91] database which *never* reach booking limits. Then, the proportion of bookings in period i relative to period $i-1$ is calculated by the simple division of averages in equation (3.8)[92]. The resulting proportions $\beta_{i,i-1}$ form an average booking curve derived from those flights which do not close. Applied to closed flights, equation (3.9) is used to multiply together the proportions between closure interval k and departure, creating an "average proportion of *final* bookings received by interval k." Bookings at closure interval k are then divided by this net average to derive unconstrained demand for the closed flight.

$$\beta_{i,i-1} = \frac{\frac{1}{n}\sum_{j=1}^{R} BIH(i)_j}{\frac{1}{n}\sum_{j=1}^{R} BIH(i-1)_j} \tag{3.8}$$

[88] Nelson (1982), p. 327.
[89] Schneider (1986), pp. 59-60.
[90] Lee (1990), pp. 50-51.
[91] "Entire" refers to *all* flights in the HDB, not simply the M most recent flights used by the forecaster.
[92] This proportion will be less than one if positive bookings are received in the ith booking interval.

$$(3.9) \qquad B\hat{I}H(0)_f \mid [SA(k)=0] = \frac{BIH(k)_f}{\prod_{i=k}^{0} \beta_{i,i-1}}$$

This method critically depends on the assumption that flights which *close* have similar booking histories as those which do not. I will discuss in Chapter 5 reasons why this might not be so. I will also examine detruncation models tested in PODS which adjust for the likelihood that flights which close have different booking curves than open flights.

3.2.3. Critical Review of Past Assessments

The typical approach among comparisons of forecasting models is to ignore detruncation step entirely, thus implicitly assuming that their datasets are unconstrained (see Section 3.1.3.2). Wickham discussed extensive comparisons of several forecasting models with and without a detruncated dataset. His detruncation results are summarized as follows[93]:

- Unconstraining had no significant effect on higher booking classes, which close *least* often.
- Unconstraining decreased the spread of performance metrics among the models considered.
- Some performance metrics improved with detruncation.
- Inherent positive forecast bias significantly increased.

These results are generally consistent with our expectations about detruncation[94]. High-value classes will not be affected significantly by detruncation except on flights with *extremely* high demands, since the seat optimizer closes them last. The bias issue has been discussed in Section 3.1.3.2. Unfortunately, there has been no comparative study of different detruncation methods. PODS allows the testing of alternative detruncation methods with revenue performance as the comparative metric. I will test several methods, as detailed in Chapter 5.

3.3. Sell-Up

[93] Wickham (1995), pp. 112-113.

[94] One unresolved issue is the decreasing spread of performance metrics. Some of Wickham's chosen metrics (Wickham 1995, pp. 65-66) include some terms with squared terms, which places more weight on outliers. Clearly, detruncating (i.e., moving from Type I to II error analyses) will *significantly increase* these error metrics.

3.3.1. Sell-Up as a Consequence of Demand Interdependence

Sell-up has already been mentioned in reference to the seat inventory control process (Section 1.3.2) and as a complicating factor in forecasting error definition and cost measurement (Section 3.1.3). It is defined as the situation when a passenger, denied his or her initially desired fare option, instead makes a reservation on the same airline in a higher-valued fare class. This study assumes that sell-up occurs only *within* the flight that the passenger has initially identified, although a passenger denied on an initial flight who accepts a more expensive product on another flight of the same airline also strictly fits the definition.

The phenomena results from the imperfect segmentation of passengers into fare classes. As I have discussed in Chapter 2, airlines use signals or proxies of WTP to create different fare classes, with a gradation of prices according to the airline's estimate of each group's approximate willingness to pay. Some passengers will always be able to evade their designated group because the ability of signals to differentiate is limited, and some will have atypical characteristics (e.g., some businesspeople may *require* a weekend stay as part of their trip, rather than wishing to *avoid* it). These passengers will retain much of their consumer surplus by purchasing a lower-valued fare when available. The airline has therefore *misidentified* these passengers, as its mechanism to segment passengers by WTP fails.

High demand conditions cause closure of low-value fare classes, and the probability increases that these passengers previously able to avoid proper segmentation will be denied their initially selected, low-fare product. In this situation, potential passengers whose signals correctly exhibit their WTP will drop out of the system; they will not pay more. The misidentified passenger, however, will sell-up to a higher-valued fare class. It follows that generally, high demand flights incur the lowest proportion of dilution of passengers to lower-valued fare classes. All sell-up methods aim to produce this effect, and *induce* misidentified passengers to purchase the fare product targeted to their WTP group.

3.3.2. Literature Review of Available Techniques

3.3.2.1. Refine the Fare Structure

There have been several attempts to modify the seat inventory control process to induce sell-up. The first is simply to refine the fare structure, and eliminate the imperfections which cause yield dilution and sell-up possibilities in the first place. Such changes (Step 1b in Figure 1.3) aim to reduce the mismatch between passenger WTP and targeted WTP of the passenger's taken fare class. No formalized approach or adjustment mechanism exists to eliminate fare structure problems; tools available include the number of fare classes, restrictions attached to each category, and absolute fare levels. As discussed in Chapter 2, airlines' ability to distinguish between passengers is limited by potential signals. Lacking a perfect mechanism to discover passengers' WTP and establish the optimal fare structure accordingly, other methods which *assume* an imperfect fare structure are necessary.

3.3.2.2. Modify Booking Limits

In this regard, Belobaba provided a simple modification of his basic EMSRa seat optimization algorithm to account for an estimated probability of sell-up between *adjacent* pairs of fare classes given closure of the lower class[95]. Shifts of more than one fare class were considered unlikely, and so not addressed in his model. Protection levels for higher-value classes increase as the projected sell-up probabilities increase. This modification to the seat optimization algorithm (Step 3 in Figure 1.3) therefore *induces* sell-up by closing off lower-valued fare classes when less demand has materialized.

This basic modification introduced by Belobaba and independently by others[96] was modified for n fare classes and adapted to the EMSRb seat protection algorithm by Belobaba and Weatherford[97]. Like the earlier formation, this heuristic adjusts seat protection levels for higher-valued fare classes by the probability that passengers in lower-level adjacent fare classes will sell-up given closure. It was assumed that fare classes which close do not subsequently reopen (i.e., no cancellations or variable *BL* due to reoptimizations), and that movements over more than one fare class do not occur.

---

[95] Belobaba (1987), pp. 128-131.
[96] Brumelle et al. (1990), Pfeifer (1989).
[97] Belobaba and Weatherford (1996).

Their methodology is explained by comparing the EMSRb seat protection algorithm[98] with and without sell-up. The basic EMSRb algorithm finds the number of seats π_n to jointly protect for the fare classes *1 ... n* (in *declining* order of fare value) by equating the expected revenue of the marginal seat protected for these classes with the fare of the *(n+1)*th fare class, as in equation (3.10a) below. Rearranging (3.10a), we have in equation (3.10b) that the number of protected seats π_n for fare classes *1 ... n* should be set such that the probability of selling that number of seats is equal to the ratio of the fare for the *(n+1)*th fare class and the weighted average fare for the *1 ... n* fare classes. The booking limit *BL* for the *(n+1)*th fare class is then the seats available on the plane[99] less the seats π_n protected for higher fare classes.

(3.10a) $$EMSR_n(\pi_n) = \overline{P_n}(\pi_n) \cdot f_{1,n} = f_{n+1}$$

(3.10b) $$\overline{P_n}(\pi_n) = \frac{f_{n+1}}{f_{1,n}}$$

where $f_{1,n} = \dfrac{\sum_{i=1}^{n} f_i \overline{X_i}}{\sum_{i=1}^{n} \overline{X_i}}$ is the weighted average fare between fare classes *1* and *n*.

Belobaba and Weatherford modified equation (3.10b) for nonzero sell-up probabilities $SU_{n+1,n}$ between adjacent fare classes *n+1* and *n* in equation (3.11)[100]. Increasing the sell-up probability $SU_{n+1,n}$ decreases the R.H.S. of (3.11), which requires commensurate increases in protections π_n for high-valued fare classes.

(3.11) $$\overline{P_n}(\pi_n) = \frac{f_{n+1} - f_{1,n} \cdot SU_{n+1,n}}{f_{1,n}\left(1 - SU_{n+1,n}\right)}$$

---

[98] Further details are found in Belobaba (1992).

[99] In most implementations, this will be the aircraft capacity *cap* before the beginning of the booking process and $cap - \sum_{c=1}^{n} BIH(i)_c$ for intermediate booking intervals *i* where *c* indexes all fare classes, assuming no overbooking.

[100] Belobaba and Weatherford (1996) discusses the derivation of this modification.

One important assumption of this technique is that misidentified passengers in each fare class targeted for sell-up do not arrive *earlier* than other passengers in their respective fare classes. Otherwise, those willing to pay more take up seats in low-value fare classes early, denying later-arriving passengers who are not. A policy of adjusting protection levels is most successful if misidentified passengers in low-value classes arrive *last*. If misidentified passengers have even a modest tendency to plan trips later than *bona fide* discretionary passengers in a given fare class (which is likely), they will arrive toward the end of the low-value fare classes' availability periods.

3.3.2.3. Prematurely Close Fare Classes

A third published attempt to induce and measure sell-up is discussed by Bohutinsky[101], who documented the common *ad hoc* approach to inducing sell-up by airline revenue management analysts. This involves the *premature* closure of low-value fare classes before their normal expiration as specified by advance purchase (AP) restrictions. It is essentially a specific case of fare structure modification (Step 1b of Figure 1.3), as it tightens the advance purchase restriction for identified fare classes. A fare class could be identified as having a high proportion of misidentified passengers if, for example, a consistent spike in bookings occurs in higher-value fare classes immediately upon closure of the class. Targeted premature closure dates are structured to avoid closure of identified fare class/es *before* the normal AP cutoff of *lower*-valued fare classes. This prevents incongruities like a higher-value fare class being closed while a low-value fare class is still open (which could precipitate further *dilution* if targeted passengers otherwise going to the prematurely closed class are able to satisfy the restrictions of the low-value class). Like other methods, this approach assumes that misidentified passengers do not arrive early.

3.3.3. Difficulties in Estimation of Sell-Up

The methods I have examined in 3.3.2 above all rely on information about how imperfect the present fare structure is in segmenting passengers. Adjacent fare class sell-up probabilities, for example, depend on the degree to which misidentified passengers dilute revenues and "get off

[101] Bohutinsky (1990), pp. 67-68, 77-80.

cheap." Such information is by definition difficult to determine (since these passengers have atypical signals), and misidentified passengers obviously have no incentive to identify themselves to the airline.

Further, sell-up is difficult to identify even when it occurs. About 70 percent of passengers purchase tickets from a travel agent, in which case airlines only have information about the final reservation accepted. The initial fare class request and subsequent examination/exclusion of available fare classes by the potential passenger is completely unknown. Nothing is known about would-be passengers who make requests to a travel agent for a low-fare product, but upon denial seek accommodation on another carrier (or decide not to make the trip). Thus, airlines cannot determine any information about sell-up from bookings occurring through travel agencies.

Even if the potential traveler makes requests directly with the airline, only by direct monitoring of phone calls can airlines estimate the proportion of passengers who are initially denied but then sell-up to another fare class. These procedures are expensive, time-consuming (since opportunities for sell-up occur only on the relatively few flights which close), and possibly misleading. Passengers often make a provisional reservation on a higher-valued fare class when denied a lower fare, only to cancel after they search other airlines for cheaper fares[102].

Sell-up is also a function of a myriad of variables besides the particular fare structure adopted by an airline. Competitive environment exerts a primary influence: markets with many alternative flights (on either the same or other carriers) will have limited sell-up potential, since the probability of finding the initially-requested fare on an alternative flight is higher. Fluctuations in demand composition (which occur around holidays, or on Monday and Friday with many business passengers) influence sell-up probabilities because nondiscretionary travelers are more likely to sell-up. Sell-up between adjacent fare classes also becomes more probable among classes of higher value, since misidentified passengers are progressively less likely to meet the increasingly stringent requirements of successively lower-value fare classes[103]. The extreme

[102] Bohutinsky (1990), p. 58. Bohutinsky discusses limited surveys performed by American Airlines and the Canadian Transport Commission (CTC) about passenger willingness to sell-up. As she correctly notes, these surveys are of very limited use, since the studies were of very small size (the American study considered only 30 passengers) or had specificity problems (the CTC survey aggregated over all airlines and markets out of Toronto and Vancouver, and relied on stated preferences of surveyed passengers. The survey also suffers from the well-known problem of relying on stated as opposed to revealed preferences).

[103] This occurs unless misidentified passengers are completely insensitive to restrictions imposed on the lowest fare classes.

specificity of sell-up and the difficulty in procuring information to estimate its probabilities indicate the caution with which airlines should approach this issue.

3.3.4. Critical Review of Past Assessments

There have been few published assessments of fare structure modification to maximize revenues. This is due in large part to an inability to systematize a methodology for identifying the maximizing fare structure, as discussed above.

3.3.4.1. Closure Experiments at Delta Air Lines

Bohutinsky documented an example implementation of the *premature closure* strategy at Delta Air Lines. This study was intended to illustrate the *ad hoc* closure methods often practiced by analysts. A small number of flights with historically high demand levels were selected, and "paired" two departures of the same flight on the same day of week (for example, the July 21 and July 28 departure of flight 1131 might be paired). One of the flights served as a *control* or base case flight which provided expected loads and revenue data without intervening for premature closure. On the other *adjusted* flight, two lower-value fare classes were closed earlier than as specified by AP restrictions[104]. Revenue differences between the control and adjusted flights were attributed to the premature closure. In total, 108 flight pairs in 21 markets were tested. Her results are summarized below:

- The premature closure policy was generally revenue negative; i.e., more revenues were lost by premature closure than gained by induced sell-up to higher-value fare classes.
- Comparisons of flights within the same week and across two weeks typically did not affect the negative revenue result.
- Sell-up was more prevalent in higher-value fare classes, and almost nonexistent at the lowest fare classes.

These results have discouraging implications for the sell-up methodologies, especially since they occurred in an actual airline context. However, some issues prevent the conclusion that sell-up generally is an unprofitable pursuit. First, the selected flights all had historically high demand levels, but there was no analysis on any of the many factors which Bohutinsky identified

[104] Bohutinsky (1990), pp. 77-79.

as being important influences on sell-up (e.g., competition, day-of-week, passenger composition)[105]. Second, revenue comparisons were drawn on the basis of the paired control and adjusted flights in the same market, but no information was provided on inherent passenger demand variability on the markets studied. This issue could have been partially addressed by using average revenue information from all other departures of this flight on the same day-of-week, within similar seasons, etc. Without a sense of the normal expected variation between flights in loads, observed revenue differences could be entirely due to chance demand variation in *both* the control and adjusted flights. However, this concern is significantly mitigated because of the large number of paired flights Bohutinsky tested.

Third, it was unclear how analysts selected the particular interval in the booking process in which to prematurely close a flight. Selection of closure interval has significant revenue effects: if closure occurs too early, many passengers unwilling to sell-up will be denied space relative to the few passengers willing to buy a more expensive fare. If closure occurs too late, most passengers willing to sell-up have already arrived -- leaving little room for improving revenues. Optimal closure interval also depends on the arrival within the booking process of passengers willing to sell-up relative to those whose maximum WTP is being extracted.

Similarly, analysts' evaluation of demand on prospective flights *before* closure was not clear. Poor analysis causes revenue deterioration: A flight which has *already been closed* by the seat inventory optimizer will not be affected at all by the "premature" closure, while a flight with low demand will incur significant revenue losses if closure is attempted. Bohutinsky's study provided many useful insights into sell-up, but numerous issues indicate that the *ad hoc* premature closure strategy should be considered only with careful analysis.

3.3.4.2. Belobaba and Weatherford's Sell-Up Simulations

The other study to assess a sell-up methodology is by Belobaba and Weatherford. As part of their work on modifying the EMSRb seat optimization algorithm to account for sell-up in *n* fare classes, a simulation was constructed to quantify the revenue benefits of this heuristic relative the basic EMSRb framework, and the previous decision rule formulated by Belobaba and others. The simulation assumes an isolated market with one airline and Poisson generated demands per

[105] Bohutinsky (1990), pp. 62-65.

class over each of the 18 intervals in their booking process. Price and demand data were taken from actual airline cases. Revenues with and without incorporation of the EMSRb sell-up modification were calculated under several demand factors, *assumed* versus *actual* sell-up probabilities (which declined with declining fare class value, and ranged from 40 to 15 percent), and sizes of the fare class structure. Their results can be summarized as follows:

- Belobaba and Weatherford's method outperformed both EMSRb and the previous formulation by increasing amounts as DF increases.
- As the number of fare classes increases, this superiority increases over the previous formulation but decreases over EMSRb. With more fare classes basic EMSRb performed *better* than the previous formulation.
- Revenue superiority over EMSRb and the previous formulation increases with the sell-up propensity of demand.
- The new heuristic outperforms basic EMSRb by up to two percent at reasonably high demand factors with few fare classes and up to one half percent with many fare classes.
- Increasing the difference between *assumed* and *actual* sell-up probability decreased the revenue improvement with sell-up adoption.

This simulation confirms expectations about sell-up adjustment mechanisms regarding superiority over EMSRb and improving performance with impacted demand. Further, the decreased superiority over EMSRb with more fare classes suggests that additional segmentation decreases misidentification of passengers -- thus limiting the benefit of sell-up adjustments. While the relative revenue performances established in the Belobaba/Weatherford study are reasonable, one assumption made by the model may qualify the estimated absolute percentage revenue improvements.

The study assumed an unbiased forecaster -- it had perfect knowledge of the mean demand and standard deviation for analyzed flights (*realized* demands were unknown but based on the distribution created by the known mean and standard deviation of demand). In reality, forecasters lack this information, and must infer it from observed bookings on previous flights. The result of this imperfect information is that forecasts are rarely unbiased. In actual application, this forecaster property could cause overprotection problems for high-valued classes if a large *overforecasting* error is combined with the heuristic sell-up adjustment. Of course, large negative

forecasting errors may also occur, in which case the Belobaba/Weatherford heuristic *moderates* the dilution associated with underprotection.

These two studies are the only airline-specific tests of proposed sell-up methods. Belobaba and Weatherford's study compared two *alternative* sell-up methodologies among the few available, but little work has been done on the relationship between estimated sell-up rates and revenue performance given competition *and* unknown actual sell-up propensity. PODS will be used to address this issue in Chapter 6.

IV. The Passenger Origin / Destination Simulator (PODS)

This chapter provides an overview of the Boeing PODS simulator[106], which is used in this thesis for all simulations. A detailed investigation into the structure and theory behind it has already been written by Wilson (1995), thus my focus is only on areas relevant to this thesis. I will also discuss the differences between the PODS simulator and other comparative assessments of forecasters, detruncators, and sell-up mechanisms. This thesis is the second in a series of reports resulting from the research collaboration between Boeing and the MIT Flight Transportation Laboratory. Our joint objective is to use PODS to examine how different tools and methods used in seat inventory control affect revenues, in as realistic a simulation environment as possible.

4.1. PODS System Flow

This section provides a *macro* perspective on PODS; Sections 4.2 and 4.3 will provide limited *micro* details on some components. First, some definitions: A *case* is defined to be a complete set of input values for all the variables used in the PODS system. An *observation* is a *complete* run through the seat inventory control process for the flights in the input case. To determine the revenues, loads, and other statistics which would result from this case, a number of these observations are performed and averaged together. There are assumed to be no *trends* in any of the input values of variables over the observations; However, many variables are stochastic. Further, a completed observation affects the performance of *subsequent* observations because PODS correctly simulates the forecasting process of predicting bookings on a forecast flight by extrapolating (in part) from completed flights in the HDB. Similar to the steps described in Figure 1.3, after each observation has run entirely through the seat inventory control process, it becomes part of the HDB which is then used to predict *future* observations.

At the beginning of the case, PODS has no HDB of completed observations from which to predict subsequent observations. It therefore *initializes* an HDB by running some (currently 200) observations under "cold-start" conditions, with user-input initial forecasts[107]. Gradually, these

[106] PODS Version 5C or 5D was used for all simulations.
[107] All experiments used the same input values which were arbitrarily derived. These numbers have no affect on final results because of the burn process, but are required for the simulation to begin.

forecast values are replaced with completed observations (since the HDB always includes the most recent departures), until the HDB contains no flights either with the initial forecasts or substantially affected by them. These affected observations are "burned" from the simulation, and are *not* included in final case statistics.

Because of the absence of trends and the reliance of the forecast for the present observation on HDB data, it was noted that the revenue and load results of *successive* observations would eventually stabilize. However, it was also discovered that the stabilized values occasionally varied significantly depending on random events in the "burn phase" of the first 200 observations, or shortly thereafter. To combat this problem, the total observations which are averaged to derive average results for the case are split into a number of *trials* (currently 20). At the start of each trial, the HDB initialization process is run and observations after the first 200 are recorded into case results. Each trial consists of 1,000 observations total, the last 800 of which are included in results for the case. Thus, each case is the *averaged* result of 16,000 observations, taken over 20 trials. The results reported in this thesis should be seen as the expected results from a flight on a particular day with all trend effects removed.

PODS system flow is illustrated in Figure 4.1. After the input of variables in Step I (see Section 4.2) and the HDB generation process for this first trial, demands for each market and passenger type for the first observation are stochastically generated (Step II; see Section 4.3). No-shows (if included) are estimated using HDB data, allowing the computation of overbooking "rates" (the percentage by which total aircraft capacity *cap* is multiplied to derive total "authorized units" which may be sold for this flight)[108].

Then, for each booking interval, the forecaster estimates demand for all flights based on data from *historical bookings* data (from the most recent 52 observations in the HDB) and/or *advance bookings* information (bookings-in-hand $BIH(i)_f$ on the forecast flight f). This is Step III in Figure 4.1, and is discussed in Section 4.4. On this basis the seat optimizer sets booking limits. Separately, the total proportion of simulated actual requests which are to arrive in this booking interval for each flight is generated (Section 4.3), with some variation around a booking curve *for each passenger type*. Cancellations which will occur in this period are calculated, according to an

[108] In PODS this *overbooking model* occurs before the setting of booking limits unlike my description of seat inventory in Figure 1.3, where it immediately followed the seat optimizer.

assumed constant input probability that a request made in a *previous* booking interval will be canceled in a following interval.

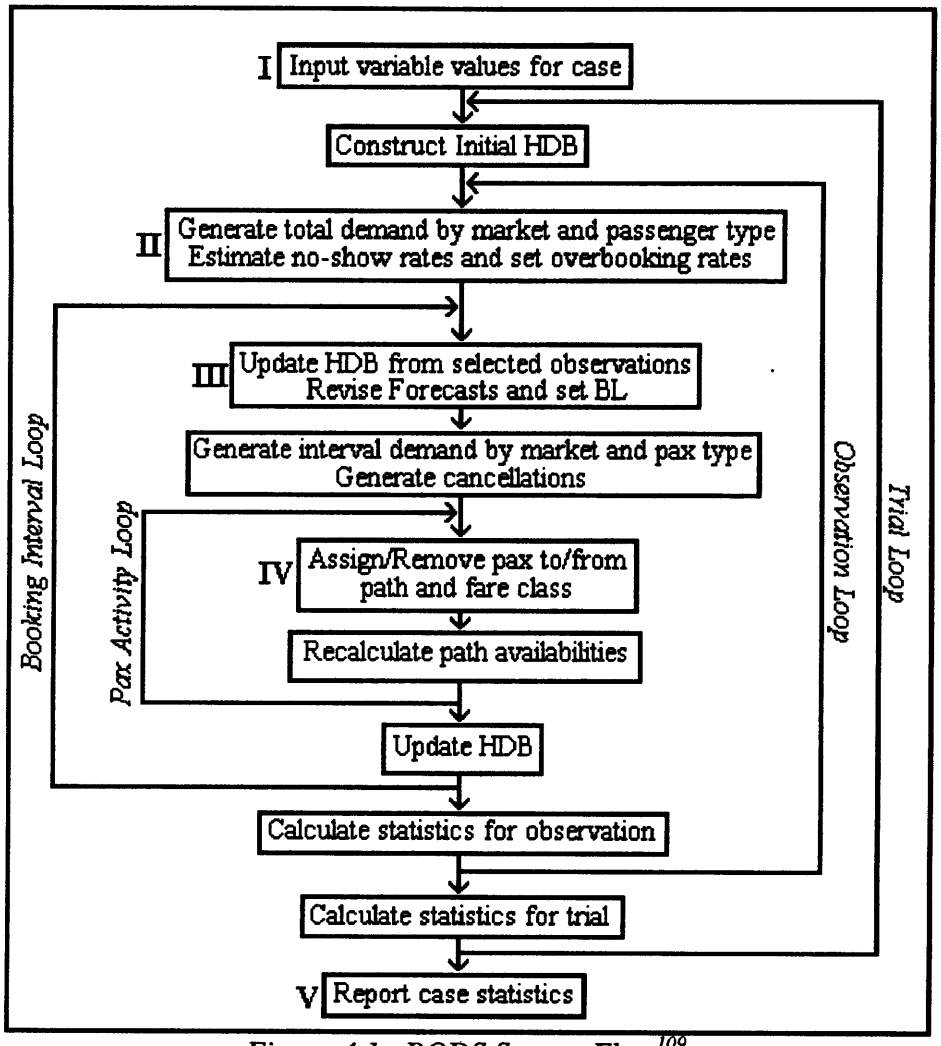


Figure 4.1. PODS System Flow[109]

For each passenger (pax) request or cancellation in this period, the "pax activity loop" is followed. If the activity is a request, choice of flight and fare class occurs by evaluating available alternatives against a passenger's willingness to pay and monetary valuation of service attributes (Section 4.5). This process is also done for each possible itinerary involving more than one flight, and also includes disutilities associated with connections, longer flight duration, etc. inherent in services which are not non-stop. For cancellations, an existing booking made in any previous

[109] Adapted from Hopperstad (1996), p. 5.

booking interval is randomly selected. Increment or decrement of seat availability is performed as appropriate for each passenger activity.

Once all passenger activity in the booking interval has been accomplished for all flights, the "booking interval loop" is followed. New forecasts are performed on the basis of HDB flights and/or bookings-in-hand *BIH(t)* information up to the *t*th interval. The simulation proceeds to the *(t-1)*th interval and repeats the passenger activity loop according to the stochastic number of passengers arriving and canceling in the period. Once all bookings intervals have occurred, relevant statistics for each flight in this observation are kept, and the process is repeated for the next observation. After all observations have been recorded, average statistics are recorded and this trial is complete. Upon the repetition of this process for all trials and the generation of final statistics, this case is complete.

4.2. PODS Inputs (Step I)

The simulator takes three levels of inputs for each trial. *System-level* inputs operate over all flights and markets in the network; *airline-level* inputs are specific to the airline operating a given set of flights; finally, *market-level* inputs concern passenger and airline characteristics by airline. I list the present input variables in PODS in Tables 4.1-4.3 below. No further explanation will be provided on these inputs except in the following sections, or where revenue effects of their variability are being tested. More explanation on these inputs is provided in Wilson (1995).

Table 4.1. PODS v5D System-Level Inputs

Sizing Variables:	*Booking Process Variables:*	*Simulation Variables:*
No.of Airlines	No. of Booking Intervals	No. of Observations
No. of Legs	No. of HDB Obs. Used by Fcster	No. of Obs. Burned
No. of Markets	Length of Each Booking Interval	No. of Trials
No. of Pax Types		

Measures of Stochasticity:	*Variables by Passenger Type:*
System K-factor (*skf*)	Schedule Tolerance
Market K-Factor (*mkf*)	Booking Curve
Pax Type K-Factor (*tkt*)	Acceptable Cost Ratio (ACR)
Attributed Cost K-Factor (*ckf*)	*Attributed Costs:*
Z-Factor 1, by Pax Type	Cost for Replanning
Z-Factor 2, by Pax Type	Cost by Restriction Category, Airline
	Cost for Degraded Paths
	Cost for Disfavored Airline

Table 4.2. PODS v5D Airline-Level Inputs

Seat Inventory Control Methods:	*Variables by Leg:*
Seat Optimizer Method	Capacity on Leg
Forecaster Method	Distance of Leg
Detruncation Method	Initial Forecast Demand, by Fare Class
Scaling for Closed Obs. (Bk Cv Detrunc)	Initial Forecast Error, by Fare Class
τ (Projection Detrunc)	
Adjustment for Overbooking	*Variables by Fare Class:*
Prob. of SU, Adj. Fare Classes	Last Available Booking Interval
	Cancellation Penalty
	Applicability by Restriction
	Probability of Cancellation
	Probability of Show-Up

Table 4.3. PODS v5D Market-Level Inputs

Demand Characteristics:	*Miscellaneous Variables:*
Mean Demand by Pax Type	Fare by Airline
Time-Of-Day Demand Curve	Denied Boarding Penalty
	Distance in Market
Variables by Path:	Preference Coefficient by Airline
Scheduled Arrival/Departure Times	Delta-T
Quality Index	

4.3. Demand Generation by Market and Booking Interval (Step II)

To generate simulated demand, PODS first takes for each observation measures of stochasticity (Table 4.1) and the input mean demand d_{mp}^{μ} for a market m by passenger type p (Table 4.3). This generation process is completely hidden from the forecaster, which must divine loads for an observation from historical flights and/or the developing booking history of the observation. The generated "demand" is *unconstrained* -- it represents total requests which are, on average, received for the flight -- *not* average passenger *loads*. A single multiplier $kmult_{mp}$ accounting for stochasticity on this observation is derived by equation (4.1a) below, which multiplies each component of demand which is k-variable[110] by independently drawn standard normal variates *NV*. This measure is then combined with d_{mp}^{μ}, another independent standard

[110] A k-factor specifies a constant relationship between the standard deviation and mean of a random variable, i.e., $\sigma = k \cdot \mu$. System, market, and passenger type are k-variable because averaged loads over large numbers of flight data, grouped by each of these categories, exhibit this property. This variability is also called "cyclical." See Swan (1978), pp. 84-86.

normal variate NV, and a z-factor[111] zf_1 in equation (4.1b). The equation yields total demand d^o_{mp} to be realized during this observation in market m by passenger type p.

(4.1a) $$kmult_{mp} = 1 + NV \cdot skf + NV \cdot mkf + NV \cdot tkf$$

(4.1b) $$d^o_{mp} = d^\mu_{mp} \cdot kmult_{mp} + NV \cdot \sqrt{d^\mu_{mp} \cdot kmult_{mp} \cdot zf_1}$$

Now that total demand for this observation has been determined, allocation of the requests to each of the booking intervals in the booking process proceeds according to an input *booking curve* (see Table 4.1 and Figure 1.1). Actual demand d^N_{mp} for time period N, the initial booking interval in the booking process, is given by equation (4.2), where $pbook_{pN}$ is the average cumulative booking probability for passenger type p at booking period N.

(4.2) $$d^N_{mp} = d^o_{mp} \cdot pbook_{pN}$$

From this the conditional probability cpb_{pt} that a type p passenger who will appear in observation o will book in booking interval t, given that the passenger has not booked previously (i.e., from intervals $N \ldots t+1$), is given in equation (4.3).

(4.3) $$cpb_{pt} = \frac{pbook_{pt} - pbook_{p(t+1)}}{1 - pbook_{p(t+1)}}$$

This conditional probability is applied to the total demand which has not yet arrived by interval t in the first term of equation (4.4). If there were no variability in passenger arrival interval (i.e., passenger arrivals strictly followed the booking curve), this term would be the requests realized in each period N - 1 ... 1 for passenger type p. However, PODS also allows variability around the booking curve by the second term of equation (4.4). The standard normal variable NV is independently drawn for each interval t, and zf_2 (the "secondary z-factor") is constant over all time periods N.

---

[111] A z-factor specifies a constant relationship between the variance and mean of a random variable, i.e., $\sigma^2 = z \cdot \mu$. According to Swan's taxonomy, this is the remaining variability after cyclical factors have been removed (see preceding footnote). The inclusion of the z-factor here results from Boeing studies indicating that such a relationship obtains for residual variability in air travel demand.

$$(4.4) \qquad d^{i}_{mp} = \left(d^{o}_{mp} - \sum_{j=N}^{t+1} d^{j}_{mp} \right) \cdot cpb_{pt} + NV \sqrt{\left[\left(d^{o}_{mp} - \sum_{j=N}^{t+1} d^{j}_{mp} \right) \cdot zf_2 \cdot cpb_{pt} \cdot \left(1 - cpb_{pt}\right) \right]}$$

The secondary z-factor zf_2 is a measure of *intratemporal* demand variation which occurs within a given booking interval. As zf_2 approaches zero, passenger demands by interval are less likely to vary from the relationships provided by the booking curve. Correlation in bookings *between* intervals is clearly higher as the booking curve is more closely followed. At the limit with $zf_2 = 0$ and $j > t$, $BIH(t)$ is simply a multiple of $BIH(j)$ given by $pbook_t / pbook_j$ [112]. Further, it is apparent that exactly d^{o}_{mp} requests will eventually arrive over all booking intervals (since, in the final interval before departure, $cpb_{p0} = 1$ and the second term in equation [4.4] is zero).

PODS' division of the booking process into a number of intervals (with interspersed arrivals of each passenger type and fare class according to specified booking curves) is more realistic than simulations which collapse the booking process into one instantaneous period, or impose an "arrival in increasing order of value" condition. Under these simulation conditions, we account for differences in input methodologies' ability to adjust to booking patterns revealed gradually, as a flight progresses through each booking interval. Clearly, if PODS did not include interspersed arrivals by fare class, there could be no meaningful comparison between advance and historical bookings forecasting models. The principal advantage of advance bookings models (i.e., adjusting protection for high-value classes according to developing trends on the flight) is defeated if no information is received until *after* all lower-value passengers have arrived.

4.4. Forecasting and Inventory Control (Step III)

A basic construction in PODS is the separation of *generative* demand processes in Section 4.3 from *forecasting* processes. Although the simulation has stochastically determined total demands, cancellations, and no-shows which will occur on the flight, the yield management

---

[112] This assumes there is no closure. Total bookings $BIH(t)$ at interval t is related to total demand d^{t}_{mp} received in interval t by the simple relation $BIH(t) = \sum_{j=N}^{t} d^{j}_{mp}$.

system can only estimate these activities as they are revealed throughout the booking process. This realistic separation draws out differences in *inherent bias* among between alternative forecasting and detruncation methods, and allows comparison of performance degradation as stochastic variation is increased.

Additionally, the zero-error problem (see Section 3.1.3.2) is avoided by using revenue performance as the comparative metric, and especially by having completed observations become a part of the HDB for future observations in the case. As shown in Figure 4.2, this mimics the recursive nature of forecasts affecting booking limits, which then affect realized loads, which affect future forecasts.

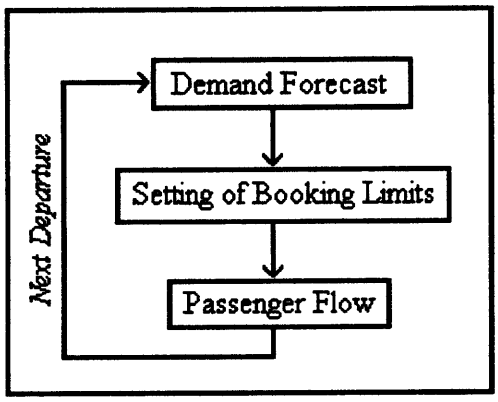


Figure 4.2. Recursive Nature of Seat Inventory Control

Comparison on a revenue basis (combined with passenger generation by *passenger type* rather than *fare class*) avoids the pitfalls of assuming independent demands. The resulting choice process by passengers among various fare classes (see Section 4.5) allows for *misidentification* of passengers and resulting sell-up opportunities. This realistic simulation process casts uncertainty on Lee's assertion that both positive and negative forecast "error" or bias causes revenue declines (Figure 3.8). It is theoretically possible that a systematic forecast bias towards overprediction of higher-valued fare classes may actually induce sell-up and therefore be *revenue positive*. There are two necessary (but not sufficient) conditions for this event to occur. First, the fare structure assumed by the forecaster must be *imperfect,* so some passengers are misidentified. Second, the flight being forecast must be constrained in some fare classes as a result of the overprotection (otherwise no misidentified passenger is forced into a higher-value class).

The treatment in PODS of forecasting, detruncation, and seat optimizing is exactly analogous to Steps 1a, 2, and 3 of the general seat inventory control process in Figure 1.3. Methods and algorithms for each of these processes may be tried in various combinations as desired for comparative purposes, as I will detail in Chapter 5.

4.5. Passenger Assignment/Cancellation (Step IV)

Passenger assignment is the process of selecting for each request an available path, or "spilling" the passenger from PODS if no paths are available which meet the passenger's WTP. Cancellations occur when a passenger decides *before* the end of the booking process not to travel according to originally set plans, thus freeing a space previously unavailable. I will briefly discuss the theoretical assumptions made by PODS with respect to passenger assignment (Section 4.5.1), then cover the implementation of this theory in PODS (Section 4.5.2).

4.5.1. Boeing Decision Window Concept

PODS' passenger assignment process assumes that the governing factor in flight selection is *schedule* relative to the times individuals wish to travel. Generally, travel demand is almost always defined within a concrete band of time. A nondiscretionary traveler, for example, may only travel within a tightly defined interval to accommodate a time-inflexible schedule. Even the most discretionary passenger has a limited duration (e.g., the time a person has set aside for vacation) within which travel must occur. The matching of the *schedule* of flights or groups of flights which may be taken in a given airline O/D market with the limited times within which passengers' travel must occur forms the core of the *decision window* concept. Also developed by Boeing, this concept is used to assign each passenger generated in Section 4.3 to the various flights in the market.

Formally, any O/D airline market has a number of *paths* by which a passenger travels from the origin to destination. A *path* is defined as one non-stop flight or a set of connected flights by which a passenger may travel over the O/D market[113]. The available paths in the O/D market create a number of *schedule states*, each of which is a unique (Boolean) combination of paths

[113] In PODS, we presently restrict *paths* to contain flights on only one airline, and assume that passengers always take the *first available* flight to their destination from the connecting airport/s.

which the passenger in that state is equally willing to travel. A "1" indicates inclusion of a path and "0" exclusion. Thus, for the simple O/D market described in Figure 4.4 with two paths (**A** leaving at 0900, arriving at 1130; **B** leaving at 1700, arriving at 1930), there are three possible states. Passengers in state "10" could possibly take flight **A** but not **B**, and "01" state passengers are reversed; Passengers in the final state "11" could take either flight. Every passenger is assumed to have an *earliest* possible departure from the origin and a *latest* possible arrival at the destination. This creates a "decision window" within which the passenger can choose paths. The duration and placement over the day of a passenger's decision window depends on three input variables:

- Distribution of passenger demand over the day ("time of day" curves)
- Delta-T, the "sunk time" of the shortest flight over the O/D[114]
- "Schedule tolerance" or average window length, by passenger type

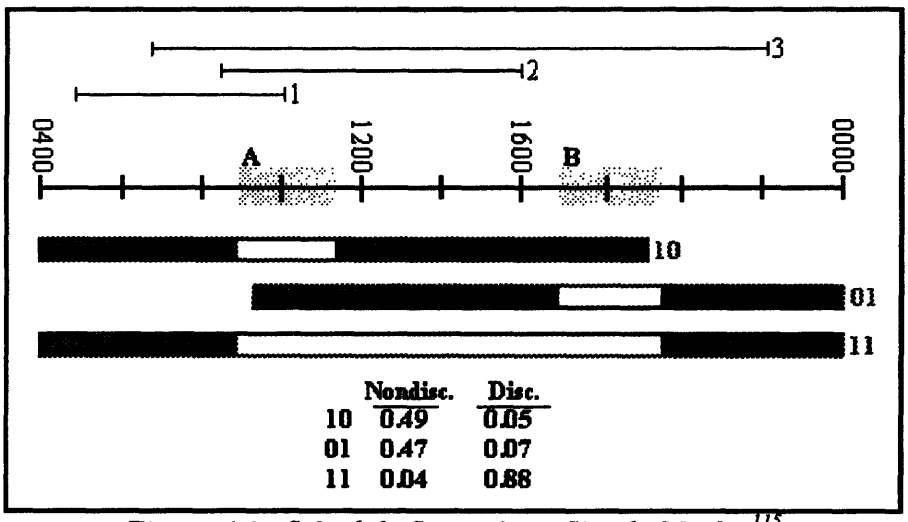


Figure 4.3. Schedule States for a Simple Market[115]

In Figure 4.3, passengers 1 through 3 have decision windows of varying lengths. Passenger 1 will not fit in any state, since the latest arrival time of his/her flight occurs before the arrival of **A**, the earliest flight. Passenger 2 could possibly take flight **A** but not **B** (and therefore falls in state 10), while passenger 3 could take either flight (i.e., state 11). The *clear* portion of

[114] This includes *nominal* differences induced by time zones. All passengers are assumed to have decision windows at least as large as the minimum "sunk time."
[115] Adapted from Wilson (1995), p. 51.

each state in Figure 4.3 indicates the minimum length of a passenger's decision window to belong to the state; the *shaded* portion indicates the maximum length without falling into another state.

PODS builds the probabilities of being in a particular state for each market *by passenger type,* with consideration of the three variables listed above. The hypothetical probabilities in Figure 4.3 for nondiscretionary and discretionary passengers are the state probabilities for this market. Clearly, nondiscretionary passengers are much less likely (and discretionary travelers much more likely) to fit in state 11 because they are by definition time sensitive. Because states 10 and 01 have approximately equivalent probabilities for both passenger types, the time of day distribution in this case has two approximately equal peak demand periods (morning and evening). PODS decision window technology presently assumes that all trips are one-way and a passenger has uniform preference for the *location* of a flight wholly *within* his/her decision window.

4.5.2. The Passenger Activity Loop

Figure 4.4 expands the passenger activity loop of Figure 4.1. If the activity is a *cancellation*, PODS simply increments seat availability *SA* over all flights in the involved path and proceeds to the next passenger activity. For a *request,* the process is more involved. First, the potential passenger's *type* is chosen according to input probabilities with some variability, and paths which are unavailable (e.g., because $SA = 0$) are eliminated from consideration. Given this type, the potential passenger's maximum WTP for travel is set as some multiple of the lowest fare in the market (the acceptable cost ratio or ACR), which is again stochastic. Next, "attributed costs" or monetary valuations of the (dis)utility incurred with various travel attributes are calculated, including: path quality (the number of stops and/or connections), use of a disfavored airline, restrictions on certain fare products, and state replanning (which will be explained below). All available paths and fare classes for which the *nominal* fare -- not including attributed costs -- is greater than the passenger's maximum WTP are summarily excluded from the potential passenger's choice process. If no paths remain, this individual is *spilled* from the system, and we proceed to the next passenger activity.

If some paths are still possible, the probabilities attached to states with at least one retained path are updated, proportional to each included state's previous probability[116]. Next total costs by path and fare class are calculated by adding attributed costs and the nominal fare, and the potential passenger's state is selected given the updated probabilities. A "replanning cost" is added to all retained paths not included in the selected state.

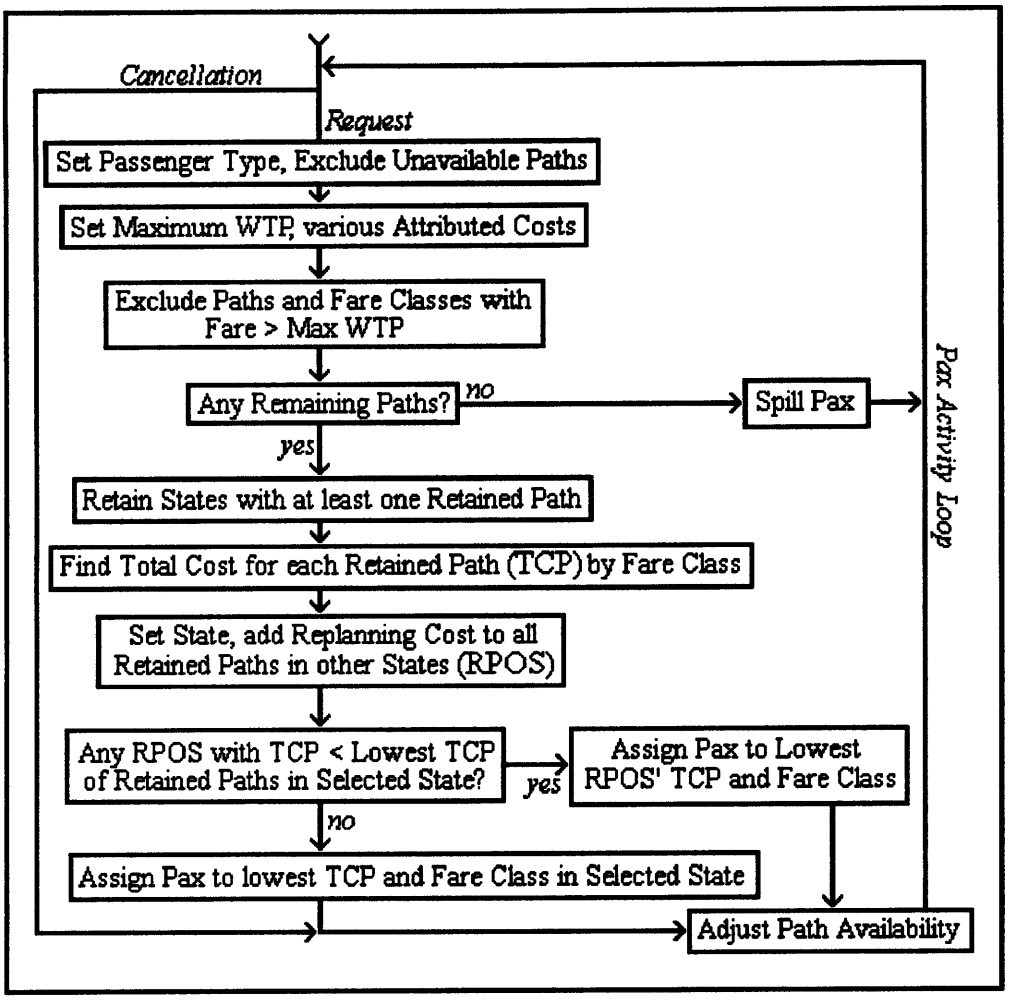


Figure 4.4. Passenger Activity Loop

Now the passenger compares the path with the lowest total cost in his/her selected state with all retained paths in other states (RPOS). If there are RPOS with total costs (*including* the replanning cost) less than the lowest cost path in the passenger's selected state, he/she will replan

[116] The process is really a simple case of Bayesian updating where the conditioning event ("removing the path does not make a state's minimum decision window equivalent to another's, or eliminates the minimum decision window completely") is a binary random variable.

-- i.e., accept the cost of changing one's initial schedule to take advantage of better options outside the initial decision window. The RPOS with the *least* total cost is chosen, seat availabilities *SA* are decremented on all involved paths and the appropriate fare class, and we proceed to the next passenger activity. Essentially, the replanning step is a simplified accounting for passenger willingness to travel *beyond* the bounds of the initially generated decision window, though at higher cost than *within* the window. Obviously, the replanning cost is significantly higher for time-sensitive passenger types. If a passenger's replanning cost is high and/or RPOS are not attractive, the passenger will select the lowest-cost path in the initially-identified state, *SA* for the appropriate paths and fare class/es will be decremented, and we proceed to the next passenger activity.

This chapter has discussed the general PODS system flow, input values, the demand generation process, and passenger assignment using decision window methodology. While some of the input values (mentioned in Section 4.2) utilized in this thesis have been discussed in passing, no formal statement of the environmental conditions assumed by this case study has been presented. Chapter 5 begins with such a presentation, and continues with a discussion of the forecasting, detruncation, and sell-up methods to be compared using the PODS simulator.

V. Description of Evaluated Methods and Simulation Environment

In this chapter, I first provide a review of the standard market and passenger characteristics (Section 5.1) used in our simulations of alternative input methodologies for forecasting, detruncation, and sell-up. These "base-case" values for the variables listed in Tables 4.1 through 4.3 provide a simplified, standardized context from which to discuss revenue differences in these methodologies. Results in Chapter 6 will establish revenues under this "base" environment, and also examine revenue effects when many of these environment variables are *altered*. Further information on these base conditions is provided by Wilson (1995), who uses a substantially equivalent base environment. Next, this chapter reviews the models for each of the three input methodologies that will be analyzed using PODS: forecasting (Section 5.2), detruncation (Section 5.3), and sell-up (Section 5.4). A brief list of alterations under which each methodology is tested is also included.

5.1. Base Case Simulative Environment

5.1.1. System-Level Inputs

The simple base case has two airlines, each of which operates one non-stop flight in a single market. Hence, there is no consideration of *network effects* -- all passengers on each airline's flight belong to one O/D market only[117]. There are only two passenger types, *nondiscretionary* and *discretionary*, which correspond to Types I and III (respectively) of the typology in Section 2.3.1. Booking processes for both airlines consist of eight booking intervals. The length of each booking interval and booking curves for the two passenger types is depicted in Figure 1.1. The total number of observations per trial, number of trials, and "burn" information is described in Section 4.1.

Significant stochastic variation is included in the base case. System and attributed cost K-factors are 0.3, i.e., the standard deviation for systemwide demand and each attributed cost listed in Table 4.1 is 30 percent of their respective means. This level of system demand variability is

[117] Extension to a full-network base simulative environment is expected within the year.

consistent with past studies of airline booking variability[118]. However, the market K-factor is zero in this single market case (this measure is only useful in multiple-market scenarios), as is passenger-type K-factor. The primary and secondary z-factors are 4.0 and 2.0, respectively, implying *moderate* correlation in bookings between intervals and variation in bookings received for each observation d^{o}_{mp} (see equation [4.1b]).

The monetary valuations of attributed costs and the acceptable cost ratio are system-level inputs, but describing their values here is not informative without consideration of the underlying fare structure. I will therefore consider these issues under Section 5.1.3, where fare levels are defined.

5.1.2. Airline-Level Inputs

Each airline is assumed to use the EMSRb seat optimizer originally proposed by Belobaba[119]. Use of EMSRb alone is appropriate because our purpose is to test revenue differences induced by input methods to seat optimizers, not the optimizers themselves. Further, EMSRb technology is in use at a variety of airlines worldwide. In the base case, there are no cancellations nor no-shows (all initial requests will result in a show-up if an available path is successfully chosen). Thus, the two airlines do not adjust for no-shows (so the *overbooking model* step of Figure 1.3 is not performed).

Other airline inputs include the choice of forecaster and detruncator. Unless variation in the input methodology is being specifically tested, the default forecaster is a simple classical pickup model as described in equation (3.3). The default detruncation method uses booking curves derived from unclosed observations, as in equations (3.8) and (3.9). There is no scaling for the possibility that closed observations have different proportional booking relationships than unclosed observations. No adjustment of booking limits for the assumed probability of sell-up is performed.

5.1.3. Market-Level Inputs and all Fare-Related Inputs

---

[118] E.g., Sa (1987), pp. 35-70 shows results for 28 flights between January and June, 1986. Derived K-factors are generally well above 0.3. However, this study mixed seasons and did not control for changes in the operational environment (e.g., frequencies offered by the airline).

[119] Belobaba (1992).

At the market level, the airlines' non-stop flights are offered at the same time each day. Thus, they have equivalent arrival and departure times. A demand factor (DF) which relates mean total unconstrained demand to total capacity available in the market is a convenient way to describe demand conditions when no leg carries passengers from more than one market. The base DF is 0.9, which is sufficiently high that yield management is of some benefit (with stochastic variation, many observations will have closure in some fare classes) -- but not so high that the optimal airline response, given its particular fare structure, is simply to offer only the highest-value fare.

Both carriers are assumed to adopt the same fare structure, which is given in Table 5.1[120]. There are four fare classes, each of which has a variety of restrictions and associated attributed costs by passenger type. As the monetary fare decreases, more restrictions are introduced which limit access to the fare: Advance purchase, Saturday night stay, round-trip, and other restrictions are added. This is in addition to increasing restrictions on *seat availability* for lower-fare classes, which is accomplished automatically by the seat optimizer. The base fare class structure shown in Table 5.1 draws heavily on investigations into common industry practice, as discussed by Wilson[121]. Monetary fare levels (e.g., $100 for unrestricted Y class and $40 for the most restricted, cheapest Q class) are approximately representative of present *relative* differences in fare products (compare Table 4.1 with Table 2.2). Note that nondiscretionary (ND) travelers' *total* attributed costs for restrictions are significantly higher than for discretionary (D) passengers, and that individual restrictions have *differential* impacts on the passenger types. For example, the attributed cost of a Saturday-night stay is five times greater to a nondiscretionary passenger, while refundability matters more to discretionary travelers (who must pay for their trips, as opposed to the nondiscretionary group primarily composed of business travelers not directly paying for the trip).

Several cost-related passenger type variables are described in Table 5.2. These variables

[120] This approximates airline practice, where very low information costs enabled by the CRS prevent significant differences in fare structures.
[121] Wilson (1995), pp. 57-61.

Fare Code	Price	Adv Purchase	Restriction Categories and Individual Costs								Total Average Attributed Cost	
			Sat Night Stay		Round-Trip		0% Refundability		50% Refundability			
			ND	D	ND	D	ND	D	ND	D	ND	D
Y	$100	-	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	$0	$0
B	80	3	N/A	N/A	$15	$0	N/A	N/A	$5	$10	20	10
M	50	7	$25	$5	15	0	$10	$15	N/A	N/A	50	20
Q	40	14	25	5	15	0	10	15	N/A	N/A	50	20

Table 5.1. PODS Base-Case Fare Structure[122].
Note: N/A = restriction not applicable to fare product

Variable	Passenger Type	
	Non-Discretionary (ND)	Discretionary (D)
Acceptable Cost Ratio (ACR)	5.0	2.0
Cost for Replanning	$50	$20
Cost for Degraded Path	$0	$0
Cost for Disfavored Airline	$0	$0

Table 5.2. PODS Base-Case Passenger Type Details.

[122] Adapted from Wilson (1995), p. 60.

are all subject to stochastic variation according to the attributed cost K-factor (Section 5.1). As described in Section 4.5.2, the acceptable cost ratio (ACR) governs what multiple of the lowest nominal fare ($40) each passenger type is willing to pay, on average, in the initial path screening process. Thus, a nondiscretionary passenger will never exclude any fare product, while discretionary passengers will on average include all fare classes but Y under the base conditions. Table 4.2 indicates that the *replanning* cost is significantly higher for nondiscretionary passengers, as is expected given their time sensitivity. Costs for degraded paths and disfavored airline use are not relevant in the present PODS scenarios: With only non-stop paths in the simple one-airline, two-airline base environment, no paths are "degraded." Finally, disfavored airline use is not considered in the base case, where there are equal preferences among airlines.

5.2. Forecasting Models

The variety of forecasting models and the myriad of simulation variables under which revenue performance may be tested presented a significant problem: Limited resources required either testing many forecasting models under a limited set of conditions, or testing a small or selected set of models under a variety of environmental conditions. This thesis adopts the latter approach. Testing a select number of representative models under many conditions yields insight into how well fundamental differences in forecasting approaches adapt to changing environments, and is more rewarding than examining methodological nuances which yield substantially similar results. To this end, the three forecasting models I will consider are described in Sections 5.2.1 through 5.2.3 below. I analyze one advance bookings, one historical bookings, and one combined model. PODS revenue comparisons of these alternative forecasters avoids the pitfalls associated with "error" comparisons (see Section 3.1.3.2), and incorporates a passenger choice framework that avoids the difficulty with simulations assuming independence of demand between fare classes.

5.2.1. Non-Causal Regression (Advance Bookings)

The first selected forecasting method is a non-causal regression model relating bookings in hand $BIH(t)_f$ at booking interval t to either bookings to come from intervals t to 0 (equation [5.1]) or total bookings $BIH(0)_f$ at completion of the bookings process (equation [5.2]). If *bookings-to-*

come were estimated via regression, final bookings $\hat{BIH}(0)_f$ would be derived by adding the result of equation (5.1) and $BIH(t)_f$.

(5.1) $$BIH(0)_f - BIH(t)_f = \alpha_t + \eta_t \cdot BIH(t)_f + \xi_t$$

(5.2) $$\hat{BIH}(0)_f = \gamma_t + \vartheta_t \cdot BIH(t)_f + v_t$$

where ϑ_t and η_t are coefficients on $BIH(t)$ found via OLS estimation
v_t and ξ_t are random error terms
α_t and γ_t are intercept terms
t is the booking interval from which estimation occurs

It has been speculated that predicting total bookings yields different results than prediction using bookings-to-come data (which are then added to bookings-in-hand). The latter method was expected to be superior, since predicting $\hat{BIH}(0)_f$ intuitively introduces prediction error for that part of final bookings already in hand by interval t. This information $BIH(t)_f$ is *already known* by interval t, it seems obvious that only the unknown part of $\hat{BIH}(0)_f$ should be forecast, i.e., bookings-to-come. I will show, however, that equations (5.1) and (5.2) are equivalent (assuming no cancellations).

Following the standard technique of bivariate ordinary least squares estimation (OLS)[123], the coefficient ϑ_t in equation (5.2) above is given by the solution to equation (5.3a). For the discussion which follows I have introduced the following simplified notation: $BIH_t = BIH(t)_f$, and $BTC_t = BIH(0)_f - BIH(t)_f$.

(5.3a) $$\vartheta_t = \frac{\sum_{j=1}^{N}\left(BIH_j - \overline{BIH}\right)\cdot\left[\left(BTC_j + BIH_j\right) - \left(\overline{BTC + BIH}\right)\right]}{\sum_{k=1}^{N}\left(BIH_k - \overline{BIH}\right)^2}$$

where $\overline{BIH}$ is *average* $BIH(t)$ over the N flights
$\overline{BTC + BIH}$ is the average $BIH(0)$ over the N flights
N is the number of flights in the HDB used for estimation of the regression

[123] Further details on the OLS regression process can be found in any standard econometrics or statistical analysis textbook, e.g., Cohen and Cohen (1975).

Distributing the terms of equation (5.3a), we have

$$\vartheta_t = \frac{\sum_{j=1}^{N}\left(BIH_j - \overline{BIH}\right)\cdot\left(BTC_j + BIH_j\right) - \sum_{j=1}^{N}\left(BIH_j - \overline{BIH}\right)\cdot\left(\overline{BTC + BIH}\right)}{\sum_{k=1}^{N}\left(BIH_k - \overline{BIH}\right)^2} \tag{5.3b}$$

The second term of equation (5.3b) self-evidently cancels out. The first term, after appropriate multiplication and distribution, can be transformed to equation (5.3c).

$$\vartheta_t = \frac{\sum_{j=1}^{N}\left(BIH_j \cdot BTC_j - \overline{BIH}\cdot BTC_j\right) + \sum_{j=1}^{N} BIH_j \cdot \left(BIH_j - \overline{BIH}\right)}{\sum_{k=1}^{N}\left(BIH_k - \overline{BIH}\right)^2} \tag{5.3c}$$

If the numerator and denominator of the second term of equation (5.3c) are distributed, the second term reduces to one, and the derivation of ϑ_t has been sufficiently simplified. Next, OLS estimates the coefficient η_t for equation (5.1) by equation (5.4a) below.

$$\eta_t = \frac{\sum_{j=1}^{N}\left(BIH_j - \overline{BIH}\right)\cdot\left(BTC_j - \overline{BTC}\right)}{\sum_{k=1}^{N}\left(BIH_k - \overline{BIH}\right)^2} \tag{5.4a}$$

Distribution of terms and canceling yields equation (5.4b) below, which is equivalent to (5.3c) minus one. That is, $\eta_t = \vartheta_t - 1$.

$$\eta_t = \frac{\sum_{j=1}^{N}\left(BIH_j \cdot BTC_j - \overline{BIH}\cdot BTC_j\right)}{\sum_{k=1}^{N}\left(BIH_k - \overline{BIH}\right)^2} \tag{5.4b}$$

Now the intercept terms can be estimated. The standard OLS formula for the intercept α_t of equation (5.1) is given by equation (5.5) below. The intercept γ_t for equation (5.2) is given by

equation (5.6). Simple algebra establishes that equations (5.5) and (5.6) are equivalent, i.e., $\alpha_t = \gamma_t$.

$$\alpha_t = \overline{BTC} - \overline{BIH} \cdot \left(\frac{\sum_{j=1}^{N} \left(BIH_j \cdot BTC_j - \overline{BIH} \cdot BTC_j \right)}{\sum_{k=1}^{N} \left(BIH_k - \overline{BIH} \right)^2} \right) \tag{5.5}$$

$$\gamma_t = \overline{BTC + BIH} - \overline{BIH} \cdot \left(\frac{\sum_{j=1}^{N} \left(BIH_j \cdot BTC_j - \overline{BIH} \cdot BTC_j \right)}{\sum_{k=1}^{N} \left(BIH_k - \overline{BIH} \right)^2} + 1 \right) \tag{5.6}$$

Substituting $\eta_t = \vartheta_t - 1$ and $\alpha_t = \gamma_t$ in equation (5.1) reduces it to (5.2), and the equality of the two is proved. Not all descriptive statistics for the two models will be equal: For example, R^2 will in general be different. Predicting total bookings given bookings in hand will be equivalent to predicting bookings-to-come, so long as results are not adjusted by descriptive statistics (e.g., if a weighting scheme were employed which placed more confidence in the regression model as R^2 improved).

5.2.2. Classical Pickup (Historical Bookings)

The second model I will analyze is simple classical pickup, introduced in Section 3.1.2.3 and repeated in equation (5.7) below. Booking matrix data utilized by classical pickup is indicated in Figure 3.5. This forecaster and the regression model in Section 5.2.1 are representative of the two primary forecasting models presently utilized by many airlines. Unlike non-causal regression models (which assume total bookings are a function of the present bookings in the forecasting interval t), pickup disregards the booking history of the forecast flight f until interval t and assumes that absolute increases in bookings until departure will mimic patterns on previously-departed flights. Only historical bookings are used as signals to estimate $\hat{BIH}(0)_f$; advance bookings are simply *added* to the result of the estimation process.

$$\hat{BIH}(0)_f = \frac{1}{M-t} \cdot \sum_{i=f-M}^{f-t} \left(BIH(0)_i - BIH(t)_i \right) + BIH(t)_f \tag{5.7}$$

5.2.3. *Boeing Efficient Forecaster (Combined)*

A "full-information" model potentially yields significant benefits, since it estimates final bookings $\hat{BIH}(0)_f$ with consideration of all possible signals of demand on the forecast flight *f*. This includes historical bookings from flights already departed and flights with incomplete booking histories, and advance bookings on the forecast flight. However, Section 3.1.2.3 pointed out the computational difficulties with "full-information" models which rely on maximum likelihood estimation (MLE). Some estimation procedure is necessary because truncated flights (i.e., completed flights which close and incomplete flights) do not have measures of unconstrained $BIH(0)_f$. MLE has desirable statistical properties under a wide range of conditions[124], but it is not presently practicable for very-large scale airline forecasting problems.

An alternative "efficient" forecaster has been proposed by Hopperstad[125] which uses all possible bookings information[126], but does not rely on cumbersome MLE estimation. This procedure is similar to Lee's approach in that the forecaster and detruncator are combined into one model. Suppose we desire to forecast the final bookings on a flight *f*, and we are at booking interval *t* in this flight's booking process. The efficient forecasting method involves three steps; first is detruncation of the last $M - t$ observations used as the HDB[127]. Instead of relying on a mathematically involved estimation procedure, Hopperstad takes advantage of the observed stability in *proportional* relationships between $BIH(i)$ and $BIH(j)$, where *i* and *j* are any two intervals in the booking process. This is true regardless of demand level[128]. A limited set of these proportional relationships form the *booking curve* relating bookings received in each booking interval i $(i > 0)$ as a percentage of *final* bookings $BIH(0)$.

---

[124] Ben-Akiva and Lerman (1985), pp. 20-22. For example, with MLE the underlying relationships between variables need not be linear or transformable-to-linear, unlike OLS estimation.
[125] Hopperstad (1991).
[126] Besides the information previously mentioned, a version of the efficient forecaster has been developed which includes advance bookings data on flights which depart on dates *after* the forecast flight. See Hopperstad (1991). This version is not incorporated into PODS.
[127] Recall from Section 3.1.2 that *M* is the number of observations used by the forecaster *plus* an "offset" due to the *t* most recent, incomplete flights. Since the full-information and efficient forecasters also use data from incomplete flights, it follows that the HDB for these forecasters includes the most recent $M - t$ flights. The HDB for forecasters using complete flights only includes *M* flights, of which the $M - t$ *earliest* (and therefore complete) flights are used.
[128] In fact, flights which *close* are more likely to have atypical booking curves. The adjustment for this fact (see Section 5.3.1) is applied to the efficient forecaster's detruncation step.

Of the last M - t previous complete booking histories, some will be truncated: Some completed flights will have closed fare classes (so *BIH(0)* does not represent final *unconstrained* bookings), and all uncompleted flights will not yet have *BIH(0)*. The efficient forecaster estimates the booking curve for each market based on all complete, unclosed observations since the beginning of the trial. This is exactly equivalent to Wickham's generic detruncation method described in equations (3.8) and (3.9). Using the information provided by equation (3.9), equation (5.8) gives $upbook_{ch}^{m}$, the *estimated* cumulative booking probability for leg m in fare class c for each booking interval h[129]. Unconstrained final bookings $B\hat{I}H(0)_g$ for each truncated flight g is estimated by equation (5.9), where h is the interval of closure in the case of closed flights, and the latest booking interval available in the case of uncompleted flights.

$$upbook_{ch}^{m} = \prod_{i=h}^{0} \beta_{i,i-1} \tag{5.8}$$

$$B\hat{I}H(0)_g = \frac{BIH(h)_g}{upbook_{ch}^{m}} \tag{5.9}$$

Now that unconstrained final bookings on all flights have been estimated, the second step of efficient forecaster generates a weighted mean of unconstrained final bookings $\overline{WBIH(0)_c^m}$ for leg m and fare class c. As shown in equation (5.10), this mean is calculated over the M - t flights in the HDB. Each flight i is weighted by the correlation coefficient $CC_i^2(j)$, calculated between *BIH(j)* at flight i's booking interval of truncation j and total bookings *BIH(0)*. A completed flight k which does not close has the highest weight, where $CC_k^2(0) = 1$. The correlation coefficient is also calculated over all the M - t flights which do not close.

[129] The $\beta_{i,i-1}$ must be calculated over the specific fare class c and leg m. We forecast at the leg level because the seat optimizer used in these simulations (EMSRb) operates at this level. Other optimizers operate at the O/D market level.

$$(5.10) \qquad \overline{WBIH(0)_c^m} = \frac{\sum_{i=1}^{M} CC_i^2(j) \cdot \hat{BIH}(0)_i}{\sum_{i=1}^{M} CC_i^2(j)}$$

where $CC_i^2(j)$ is the estimated correlation coefficient at period of truncation between *BIH(j)* and *BIH(0)* for the *i*th observation, calculated over all $M - t$ flights[130]
j is the interval of truncation for the *i*th observation

Essentially, the efficient forecaster assumes that the most confidence can be placed in observations not requiring detruncation, and that correlation coefficients are a good measure of how much confidence should be placed on $\hat{BIH}(0)$ of detruncated flights. There is no differentiation among detruncated flights according to the *cause* of their incomplete booking history. The correlation between *BIH(j)* and *BIH(0)* increases as departure date nears, since the number of booking intervals remaining (over which variability in pickup may occur) declines. Hence, the earlier in the booking process a flight is truncated, the less is the weight on that observation.

The final step is to combine this weighted average of total bookings for HDB flights with estimated final bookings information $\tilde{BIH}(0)_f$ for the forecast departure *f*. The forecast flight *f* is detruncated using the same procedure (described in equation [5.9]) as is followed for other flights with incomplete booking histories, yielding $\tilde{BIH}(0)_f$. The weight *WP* attached to $\tilde{BIH}(0)_f$ is given by equation (5.11) below, where *t* is the present booking interval for flight *f*. Final estimated forecast bookings $\hat{BIH}(0)_f$ for flight *f* immediately follows in equation (5.12).

[130] Assuming that all demand stochasticity is included in the z-factors (see Section 4.3) and perfectly adjusted pbook (see below), an alternative definition of the correlation coefficient is given by
$CC^2(j) = apbook_j \cdot zf_1 \Big/ \left(apbook_j \cdot zf_1 + \left(1 - apbook_j\right) \cdot zf_2\right)$.

$$(5.11) \qquad WP = \frac{1 + 1\Big/\sum_{i=1}^{M} CC_i^2(j)}{1\Big/CC_f^2(t) + 1\Big/\sum_{i=1}^{M} CC_i^2(j)}$$

$$(5.12) \qquad B\hat{I}H(0)_f = WP \cdot B\tilde{I}H(0)_f + (1 - WP) \cdot \overline{WBIH(0)_m^c}$$

5.2.4. Conditions to be Tested

The three forecasting methods discussed in this chapter are compared under a variety of conditions which vary from the base scenario in Section 5.1. Each of these conditions is tested with each detruncation model, which both airlines share. Following is a list of the conditions, and a brief rationale for changing each:

- Base simulation context (Section 5.1)
- Low and high demand factor (DF = 0.7, 1.2). Forecasters may not adjust equally well to different underlying demand levels.
- Low and high secondary z-factor (zf_2). Forecasters may not adjust equally well to booking curve variability in HDB flights.
- Low and high systemwide *k*-factor (*skf*). Forecasters may not adjust equally well to total demand variability *between* HDB flights.

Two other conditions not listed above were also tested. The first was nonzero cancellation rates. It was supposed that not all forecasters adjust equally to the presence of bookings which later vanish. However, no substantial differences between forecasters were revealed under a variety of cancellation rates. Second was multiple frequencies. As will be discussed extensively in Chapter 6, one primary difference between forecasters is *differential protection* for high-value fare classes, which induces sell-up. Multiple frequencies was hypothesized to limit passenger captivity and thereby sell-up opportunities. Again, no significant differences by forecasting method were revealed. The presentation of results for nonzero cancellation rates and multiple frequencies has been suppressed.

5.3. Detruncation Models

To date, there have been no published comparisons of detruncation methods. Comparing unconstraining alternatives in the PODS context provides a valuable contribution to the airline

forecasting literature. The detruncation methods I will study include two of the most common methods (pickup detruncation and booking curve detruncation), and a third developed by Boeing.

5.3.1. No Detruncation

If an airline does not detruncate observations in the HDB which close, they are simply ignored by the forecaster. This is distinguished from the method used by some airlines, where "no detruncation" means that observations which close are input to the forecaster without either adjustment or exclusion. Such a practice would *magnify* any negative effects of not detruncating reported here, since an extreme *downward bias* in pickup (i.e., zero pickup) from closure interval until the end of the booking process is implicit with unadjusted inclusion.

5.3.2. Booking Curve Detruncation

Simple booking curve detruncation has been described in Sections 5.2.3 and 3.2.2. As implemented in PODS, booking curves have no monotonicity restriction or "shape" requirement. Therefore, in the presence of significant cancellations booking curves may not be strictly increasing, and may not exhibit the typical "sideways-S" shape depicted in Figure 5.1. Presently in PODS, booking curves are estimated from unclosed flights only by fare class and leg, and are exponentially weighted (the most recent flight departures have the highest weight). The number of observations used to estimate the booking curve includes the entire observation set for each trial, up to the flight being forecast.

Booking curve detruncation assumes that the proportional relationships between *BIH(i) and BIH(j)* (where *i* and *j* are two booking intervals) are constant over all flights. Therefore, positive correlations between *BIH(i)* and *BIH(j)* are assumed. The posited proportionality is probably true over all *unclosed* flights -- so unadjusted booking curve detruncation of open but *incomplete* flights (which have not yet departed) is appropriate. However, flights which close typically *do not* exhibit equivalent proportionate booking relationships as those which remain open throughout the booking process. This section describes a modification to account for this possibility.

Suppose a hypothetical market and fare class has typical *unclosed* booking curves given by ***B*** in Figure 5.1. Booking curves to the *left* of ***B*** (e.g., ***A)*** fall in **R1**, and receive more of their

total bookings early in the booking process relative to ***B***. Booking curves to the right (e.g., ***C***) fall in **R2**, where an opposite relationship occurs: Most of their bookings are received later than the typical unclosed flight. Thus, for a given booking interval before departure *i*, flight ***A*** will have received A_i percent of its bookings, while flight ***C*** has received a much smaller C_i percent.

Under revenue management, flights in **R2** are much more likely to close than those falling either in **R1** or following the typical path ***B***. Suppose we have data on three flights in a given fare class. For simplicity, final unconstrained bookings on these flights are equal, but with three distinct booking curves according to the three paths in Figure 5.1. If the fare class involved is *high-value*, the high early bookings flight (path ***A***) induces additional seat optimizer protection for the fare class, *reducing* closure probabilities in that class. This effect is magnified if the forecaster includes advance bookings information to predict final bookings. However, the high late bookings flight (path ***C***) has an elevated closure probability, since less-than-anticipated bookings earlier in the booking process causes the seat optimizer to reduce seat protections for the high-value fare class. Hence, low-fare passengers are allowed to fill seats which would otherwise be occupied by the (unanticipated) late spate of high-value passengers. Therefore, closed flights in high-value fare classes are more likely to have path ***C*** booking curves.

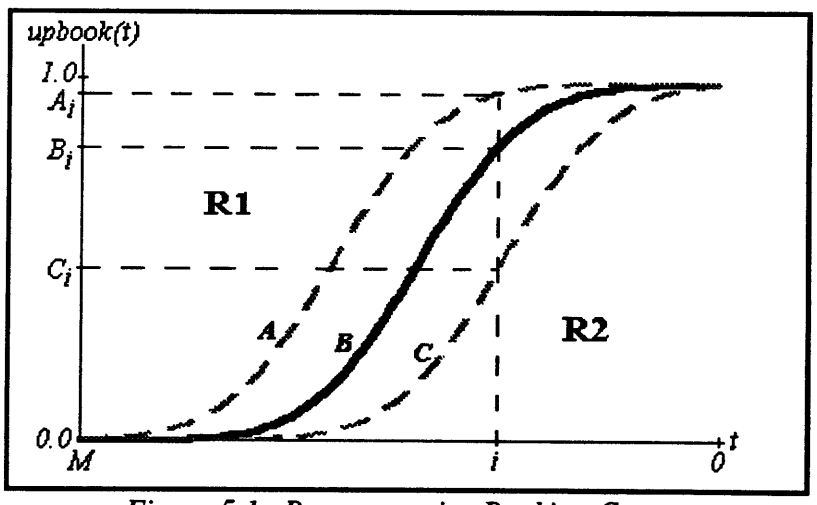


Figure 5.1. Representative Booking Curves

Now suppose the three flights involve a *low-value* class. If this low-value class has high early bookings (path ***A***), its booking limits are unaffected: The seat optimizer is sensitive only to changes from expected bookings levels on *higher-valued* fare classes. Similar reasoning holds for

the "typical" flight with booking path ***B*** in the low-value class. Since high-value bookings generally arrive *later* than low-value bookings, neither paths ***A*** nor ***B*** will be affected by atypical booking behavior in higher-valued fare classes: The bulk of the low-value bookings will already have been received by the time the variation in high-value bookings occurs. It follows that neither the high early bookings flight nor the "typical" flight with booking path ***B*** will be especially likely to close. In contrast, the flight with high *late* bookings in the low-value class (path ***C***) is more likely to be affected by a surge in high-value fare class demand. Low-value passengers under path ***C*** conditions will not have claimed seats *before* the high-value arrivals, and the seat optimizer quickly reduces booking limits for the former to accommodate the latter. Thus, regardless of class value, closed flights are more likely to involve atypically high *late* arrivals (path ***C***) than open flights (which will involve paths ***A*** and ***B***).

If booking curves estimated on the basis of unclosed path ***A*** and ***B*** flights are indiscriminately applied to closed path ***C*** flights, unconstrained demand on the latter flights will be *underpredicted* -- causing a negative bias in the unconstrained data set and underprotection of the involved fare class(es). A simple heuristic adjustment for this possibility is described in equation (5.13) below, where the unadjusted estimated booking curve $upbook_{ch}^{m}$ is modified by a constant term *pbscl* ($0 < pbscl < 1.0$), yielding the *adjusted* cumulative booking probability $apbook_{ch}^{m}$ for each market m, fare class c, and booking interval h. Our comparative detruncation tests will include unadjusted and various levels of *pbscl*-adjusted booking curves. Note the chosen *pbscl* is based on reasonable hypotheses and not systematic comparison of booking curves for flights which do close and those which do not, since the latter is by definition unavailable.

$$apbook_{ch}^{m} = pbscl \cdot upbook_{ch}^{m} \tag{5.13}$$

5.3.3. Projection Detruncation

The second detruncation method this thesis will consider is projection detruncation, developed by Hopperstad[131]. For any flight which closes, the conditional probability that unconstrained bookings on a flight g are greater than its unconstrained forecast (given closure)

[131] Hopperstad (1995).

can be specified, if some assumptions are made about the distribution of closed flights. This detruncation method proceeds from the idea that this probability should be a specified constant. Suppose we are at forecasting a flight f from booking interval t, and must estimate $B\hat{I}H(0)_g$ for some HDB flights g which close at various intervals x (all $x < t$)[132]. The projection detruncator first calculates the mean μ_t and standard deviation σ_t of "pickup" -- the increase in bookings between intervals t and 0 -- based on those flights among the $M - t$ in the HDB which do not close at any time during the booking process. Next, projection takes the input conditional probability, the (nondetruncated) pickup statistics, and (for each truncated flight) incremental pickup cl between the interval of prediction t and the interval of closure x. It then projects estimated pickup $\underline{cl}$ for the closed observations g, assuming that closed observations come from the same demand distribution as the unclosed observations[133]. This situation is illustrated for one flight g in Figure 5.2, where area $\boldsymbol{A}$ is the probability of receiving between cl and $\underline{cl}$ pickup, area $\boldsymbol{B}$ the probability that more than $\underline{cl}$ pickup occurs, and area $\boldsymbol{C}$ the probability of less than cl pickup.

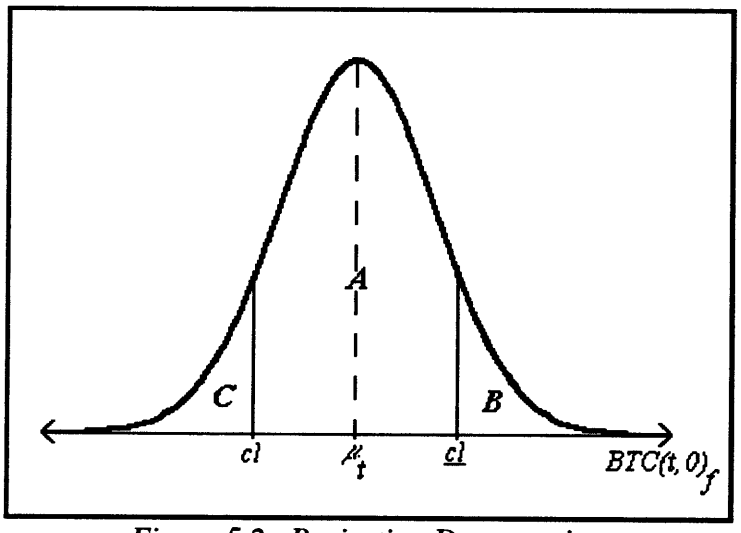


Figure 5.2. Projection Detruncation

[132] Any HDB flight which closes in booking interval t or before *cannot* be detruncated by projection, and is ignored in detruncation calculations.

[133] This seems unlikely -- clearly closed observations will have *greater* pickup, on average, than open flights. As I will show below, projection detruncation always estimates $\underline{cl} > \mu_t$. Recalculating average pickup with inclusion of detruncated observations would then increase μ_t, as expected.

Define τ to be the conditional probability of receiving more than $\underline{cl}$ bookings given closure at cl. Then $\tau = \frac{B}{A+B}$, and the projection $\underline{cl}$ is found via iteration on equation (5.14), or using normal distribution functions in equation (5.15) where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are the cumulative and inverse cumulative normal distribution functions, respectively. From equation (5.15), our estimate of final bookings for the detruncated flight g under projection detruncation is simply given by equation (5.16).

$$\frac{1}{\tau}\int_{\underline{cl}}^{\infty}\left[\frac{1}{(2\pi)^2}\cdot\mathbf{exp}\left(-\frac{1}{2}\left(\frac{x-\mu_t}{\sigma_t}\right)^2\right)\right]dx = 1-\int_{-\infty}^{cl}\left[\frac{1}{(2\pi)^2}\cdot\mathbf{exp}\left(-\frac{1}{2}\left(\frac{x-\mu_t}{\sigma_t}\right)^2\right)\right]dx \tag{5.14}$$

$$\underline{cl} = \Phi^{-1}\left[\tau\cdot\left(\Phi\left(\frac{cl-\mu_t}{\sigma_t}\right)-1\right)+1\right] \tag{5.15}$$

$$B\hat{I}H(0)_g = BIH(t)_g + \underline{cl} \tag{5.16}$$

Like booking curve detruncation, projection also creates a positive correlation between "incremental pickup" from prediction interval t to closure interval x and final projected bookings, since increasing cl in equations (5.14) or (5.15) increases $\underline{cl}$. Decreasing the input τ has a similar effect; as $\tau \to 0$ the probability of *underpredicting* demand decreases (protection levels are increased), with a concomitant increase in the probability of overprediction (given by area A[134]). Input τ should satisfy the inequality $\tau < 0.5$, since otherwise the resulting estimated pickup $\underline{cl}$ for the closed observation is *less* than the average pickup μ_t for unclosed observations. Base case PODS tests let $\tau = 0.15$. Tests on the sensitivity of projection detruncation results to various τ have been performed, and indicate generally that $\tau > 0.15$ makes projection forecasting substantially inferior to other detruncation methods[135].

5.3.4. Pickup Detruncation

[134] This assumes that closed observations have the same unconstrained distribution as unclosed flights' $N(\mu_t, \sigma_t)$, which is not likely. Overprediction probability should therefore be less than A.
[135] Skwarek (1996a), p. 13.

In contrast to projection methods, pickup detruncation assumes that no proportional relationship exists between bookings in hand *BIH(x)* at the closure interval x and final bookings *BIH(0)*. Instead, it asserts that *absolute* increases in bookings between x and the last interval *0* on unclosed flights are the best indicator of what pickup would have been on the closed flight g, had space been available. As shown in equation (5.17), the estimate of final unconstrained bookings $\hat{BIH}(0)_g$ is derived by adding the simple average of pickup *after* the closure interval on unclosed HDB flights to bookings received until closure $BIH(x)_g$.

$$\hat{BIH}(0)_g = BIH(x)_g + \frac{1}{C} \cdot \sum_{j=1}^{C} \left(BIH(0)_j - BIH(x)_j \right) \tag{5.17}$$

where j indexes the C (of M - t[136]) flights in the HDB which do not close at any time during the booking process
x is the interval of closure for the HDB flight g being detruncated

5.3.5. Conditions to be Tested

The detruncation methods are compared under a variety of conditions. The condition set is exactly similar to those discussed in Section 5.2.4 for forecasting methods. Base-case pickup forecaster is used by both airlines in all comparisons of detruncation methods. All detruncation methods are compared utilizing the base-case booking curve detruncation without scaling as a benchmark. Following is a list of the selected conditions, which repeats Section 5.2.4:

- Base simulation context (Section 5.1)
- Low and high demand factor (DF = 0.7, 1.2).
- Low and high secondary z-factor (zf_2).
- Low and high systemwide k-factor (*skf*).

Nonzero cancellation rates and multiple frequencies were also tested for each detruncation method comparison. Because no significant differences among methods was revealed, these results will not be shown.

5.4. Sell-Up Models

[136] In pickup detruncation the M - t earliest (complete) HDB flights are used. See Footnote 127 in this chapter.

5.4.1. Modify Booking Limits

In this thesis, I will examine the Belobaba/Weatherford modification of booking limits strategy only, as has been extensively described in Section 3.3.2.2. For simplicity, airlines in PODS experiments will assume a constant sell-up probability $SU_{n+1,n}$ for all adjacent fare class combinations $n+1$, n, given closure of the nth fare class. This is somewhat unrealistic and ignores Bohutinsky's conclusion (see Section 3.3.4.1) that sell-up is more likely to occur between adjacent fare classes of higher value. However, this avoids the complex issue of estimating sell-up rates by fare class combination, and yields a sufficient "first order" approximation to expected revenue differences.

This PODS simulation improves on the more limited simulation environment used by Belobaba and Weatherford. As with any simulation, PODS cannot claim the direct *actual* experience of testing performed by Bohutinsky, but as described in Section 3.3.4.1, "real world" experiments are subject to many underlying variables and random events which cannot always be controlled for. Because PODS abstracts away from the trends and random events influencing demand for flights, I argue that a simulation is appropriate for assessing the *systematic* revenue impact of incorporating sell-up estimates in setting booking levels.

5.4.2. Conditions to be Tested

Most base market conditions (Section 5.1) are retained. Estimates of sell-up probability $SU_{n+1,n}$ rates range 0.0 to 0.8. Three competitive sell-up environments are considered: *no adoption* (neither of the two carriers adopts an estimate of sell-up), *single adoption* (one airline adopts, the other does not), and *joint adoption* (both airlines adopt). These three environments and sell-up probabilities are tested against the following conditions:

- Base simulation context (Section 5.1)
- High demand factor (DF = 1.2). Sell-up may be more effective at high DF. Low DF is not considered because there are few sell-up opportunities if demand is not high.
- High price sensitivity, at normal and high DF (DF = 0.9, 1.2). Sell-up is less effective as passengers become more price sensitive.
- Scaling of booking curve (*pbscl* = 0.2, 0.4, 0.6, 0.8, 1.0), at normal and high price sensitivity. Previous studies indicate that excessive scaling induces

sell-up[137]. Combining scaling with estimations of sell-up may separate out these factors.

- Additional frequencies in market. More alternatives reduce passenger captivity on a flight, thus limiting sell-up possibilities. This effect may vary depending on whether the airline with additional frequencies also (in the *single* adoption case) estimates sell-up.

Now that model description and the outline of testing conditions is complete, I proceed to the PODS simulation results in Chapter 6. These results will emphasize, where possible, consistent theoretical explanations for observed revenue differences.

[137] Skwarek (1996a).

VI. PODS Revenue Results

This chapter details the revenue comparisons for the three input methodologies under the conditions described in Chapter 5. While I provide a theoretical justification for presented revenue differences, these results should be treated with caution: the multiple levels of stochasticity and interaction between input methodology choices, passenger flows, and subsequent model behavior allow many interpretations of observed differences. Sections 6.1 (comparing forecasting methods) and 6.2 (comparing detruncation methods) are closely interconnected because forecasting results depend heavily on the detruncation methodology chosen. Thus, most of the theoretical distinctions I draw for detruncation methods will be introduced in Section 6.1 and simply elaborated on in Section 6.2. Bullet summaries of principal conclusions are included for each major section in this chapter.

6.1. Forecasting Model Comparisons

This thesis compares alternative forecasters in a "pair-wise" fashion against the base-case pickup forecaster, under a variety of scenarios as described in Chapter 5. This creates two pairs: pickup versus regression, and pickup versus the efficient forecaster. In each sub-section I give results for these two comparisons. Recall from Section 3.1.1.4 that the purpose of forecasting in seat inventory control is to predict final bookings $B\hat{I}H(0)_f$ on a forecast flight f before each of the N booking intervals in the booking process. Depending on the forecaster chosen, this prediction is made on the basis of historical bookings data (final bookings $BIH[0]$ on previous departures of the same flight), advance bookings data (bookings-in-hand $BIH[i]_f$ for the forecast flight f until the forecast interval i), or both. All historical data is taken from the Historical Data Base (HDB), which contains bookings-in-hand $BIH(i)$ data over each booking interval i for recent departures of the flight being forecast.

6.1.1. Pickup/Regression Comparison

6.1.1.1. Base Case and High/Low Demand Factor Scenarios

Table 6.1 lists the percentage revenue difference of pickup over regression forecasting under three demand factors (DF). These differences are calculated for each detruncation method,

where both airlines use the same method. Methods include no detruncation; booking curve detruncation with extreme scaling[138] (pbscl = 0.6), moderate scaling (pbscl = 0.8), and no scaling (pbscl = 1.0); projection detruncation; and pickup detruncation.

Neither regression nor pickup dominates these results. It is immediately apparent that the detruncation pair used in the market significantly influences the relative performance of the forecasters. With limited exceptions, there are no substantial revenue differences at low or moderate demand levels. Pickup forecasting performs markedly better than regression forecasting at high demand factors under booking curve detruncation, but worse under projection detruncation.

	% Rev Difference, Airline with Pickup over Airline with Regression[139]					
DF	None	Bk Crv (0.6)	Bk Crv (0.8)	Bk Crv (1.0)	Projctn	Pickup
0.7	-0.26%	-0.79%*	-0.38%*	0.06%	-0.17%	-0.38%*
0.9	-0.05%	0.19%	-0.13%	-0.74%*	-0.57%*	-0.12%
1.2	-0.50%*	-0.45%*	4.89%*	7.95%*	-1.54%*	-0.49%*

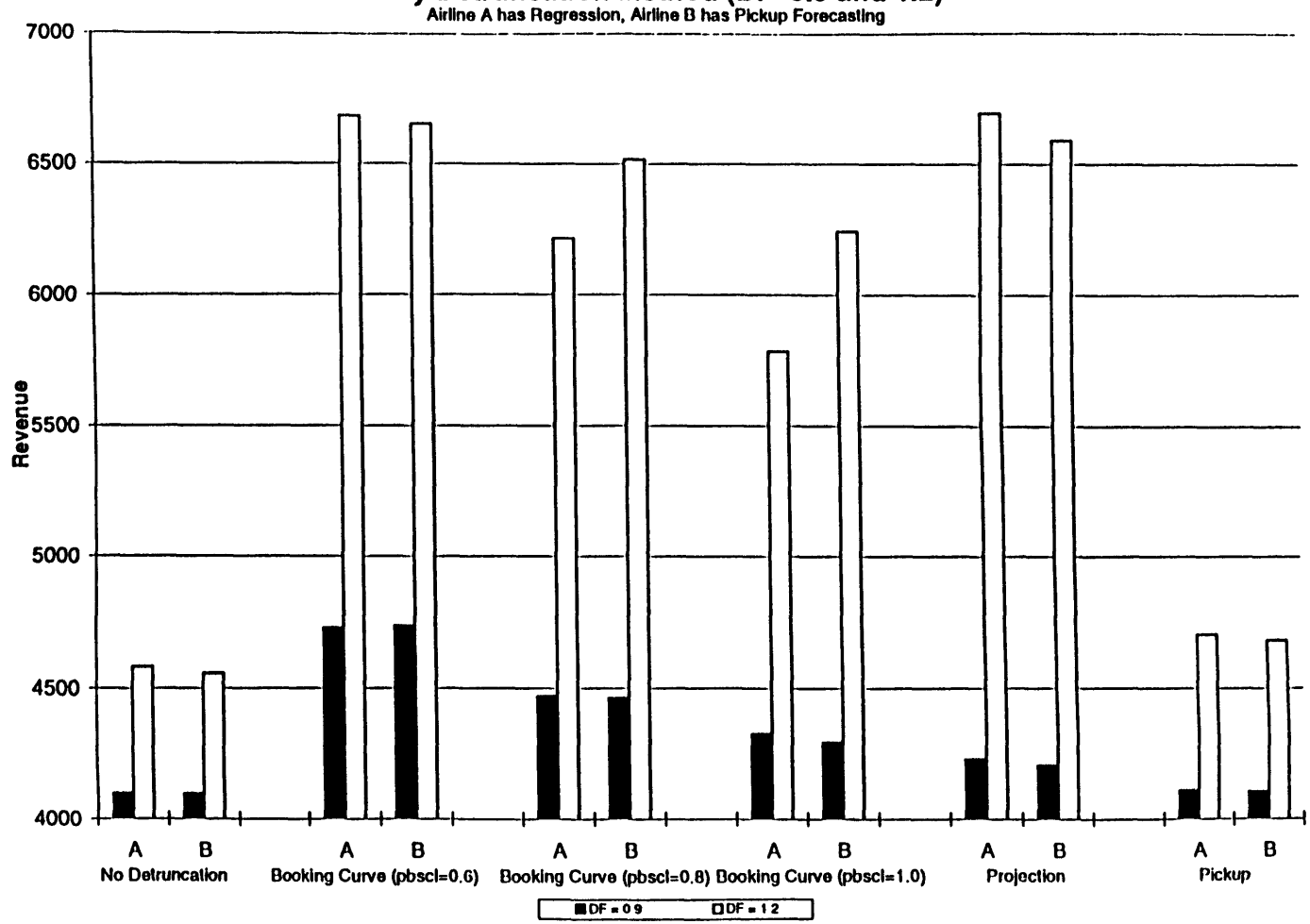
Table 6.1. Pickup versus Regression Forecasting Relative Revenue Performance

Figure 6.1 graphs *absolute* revenues for each airline for each detruncation pair; low demand factor revenues have been omitted because no differences are substantial. In these comparisons, Airline A has regression and Airline B has pickup forecasting. At moderate demand factors (DF = 0.9), revenues are more influenced by detruncation method choice than forecasting choice (i.e., Airlines A and B achieve approximately equal results for each detruncation pair choice). This changes under high demand conditions, when detruncation choice is still controlling but significant differences emerge between the airlines according to *forecaster* choice. In particular, regression significantly underperforms pickup forecasting with booking curve detruncation and no scaling, but is superior with projection detruncation. Not detruncating yields the lowest revenues as expected, but adopting pickup detruncation yields marginal improvements over this worst-case scenario. I will now discuss theoretical explanations for observed revenues *by detruncation method pair.*

[138] Recall from Section 5.3.2 that scaling of booking curves is necessary because closed flights exhibit different proportional booking relationships than unclosed flights.

[139] Results marked with * have statistically significant revenue differences using a paired t-test, at a 95% level of confidence. This notation is repeated throughout Chapter 6.

Figure 6.1. Regression versus Pickup Forecasting Revenue Results, by Detruncation Method (DF=0.9 and 1.2)



- *No Detruncation*

When neither airline detruncates, there are no substantial performance differences between regression and pickup forecasting, used by Airline A and Airline B respectively -- even under high demand conditions. As shown in Table 6.2, passenger loads under both forecasting methods are equivalent. From this we can infer that their forecasts were largely equal.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
DF=0.7	9.1	4.3	2.6	50.9	66.9	9.0	4.3	2.7	51.0	67.0
DF=0.9	10.0	5.6	3.2	62.3	81.1	9.9	5.6	3.2	62.5	81.2
DF=1.2	9.8	7.2	3.7	71.1	91.8	9.3	7.2	3.7	71.2	91.8

Table 6.2. Loads for No Detruncation Scenarios

These results may be explained with the aid of Figure 6.2, which has a representative booking history for a fare class on an unclosed flight. Suppose that the booking curve in Figure 6.2 represents the "path of average bookings by interval t" for unclosed flights in the HDB. Then at booking interval h, B_h bookings on average have been received. Average total bookings at the end of the booking process are L_d. A line drawn from average final bookings L_d to any previous interval (e.g., ray be between intervals h and 0) represents the truncated pickup forecasting predicted increase in bookings between the specified intervals (see Section 5.2.2). Assuming approximately zero variation around the booking curve (i.e., $zf_2 \approx 0$), the ray be also represents *truncated* regression forecasting predicted final bookings L_d given B_h bookings at h. Thus, without detruncation and assuming stationary demand conditions, regression and pickup forecasting *on average* always yield equivalent forecasts.

To see this, consider the unadjusted booking curve information estimated on the basis of unclosed observations in the HDB, and given by $upbook_h$ for each interval h (see Section 5.2.3). With no variance from the proportional relationships specified by $upbook_h$, the regression slope relating bookings to come $BTC(0)$ by departure (L_d - B_h in Figure 6.1) to bookings-in-hand $BIH(h)$ (B_h in Figure 6.2) will be equal across all the unclosed HDB observations[140]. Therefore,

[140] As noted in Section 5.2.1, we could also estimate a regression curve between $BIH(x)$ and *total bookings before departure* $BIH(0)$ with equivalent results.

the R^2 for the regression is 1, and its slope is equal to $\frac{1}{upbook_h}$ for the regression performed at booking interval h[141]. If a forecast flight f has B_h bookings at h (i.e., it has the *mean* bookings by interval h on HDB unclosed flights), regression and pickup forecasting yield equivalent estimates of final demand at L_d.

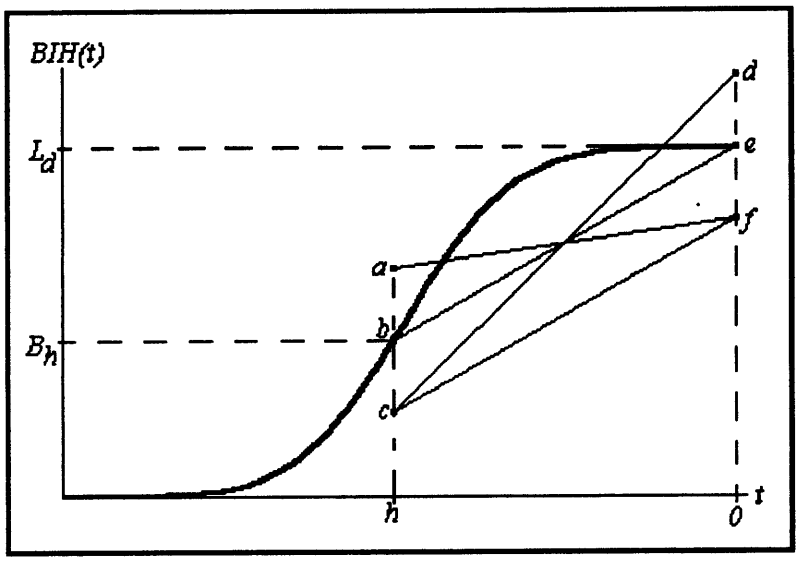


Figure 6.2. Sample Booking History

What if bookings on the forecast flight f at interval h are not equal to the average for unclosed observations, e.g., $BIH(h)_f = B_h + \varepsilon$? In this case the forecasts are no longer equivalent: The pickup forecast of final demand is unchanged at L_d, but regression differs by $\frac{1}{upbook_h} \cdot \varepsilon$. Thus, there is no guarantee that the forecast for f based on an undetruncated HDB will be equal if f does not have the average bookings-in-hand on unclosed flights at forecast interval h. However, for comparative purposes we are interested in *average* forecast and revenue differences. Since these comparisons are drawn over 16,000 observations, the mean of ε is zero. Over this large sample size, forecasts and revenues are equivalent.

[141] This assumes the intercept term in OLS has been suppressed. Also, the unclosed flights in the HDB must form a similar booking curve as is constructed for booking curve estimation (recall that booking curves are constructed using *all* observations since the beginning of the trial). Two facts suggest that this is true. First, both the forecast based on unclosed HDB data *and* booking curves are estimated on the basis of unclosed observations. Second, the minimum sample size for 95% confidence that a forecast will be within 10% of its "true" value under forecast errors of about 35% is 47.06 observations -- less than 52, the size of the HDB. The 35% is above the demand variability evinced by present system k-factors of 0.3. See Wickham (1995), p. 58.

One consideration has until this point been ignored. What if there is significant variation in *proportional* booking relationships among observations? Suppose that poor correlation exists in bookings between intervals over HDB flights, so a given bookings-in-hand at the forecast interval is not a good signal of final bookings on the forecast flight *f*. Will regression and pickup still yield, on average, equivalent forecasts?

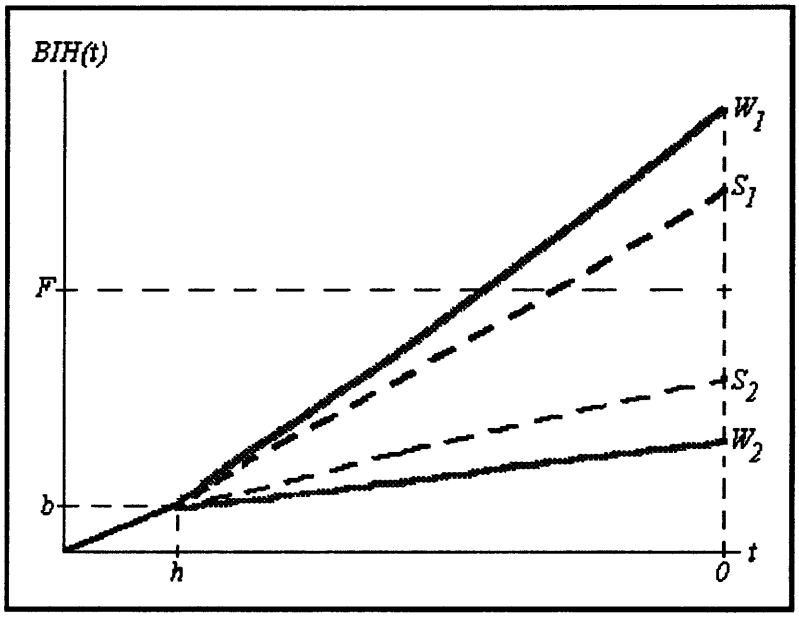


Figure 6.3. Booking Correlation Between Intervals and Range of Final Demand

In Figure 6.3 a hypothetical flight is at booking interval *h* in the booking process, with *b* bookings-in-hand. If the correlation in demand between booking intervals is high ($zf_2 \approx 0$), demand in the each remaining interval is closely related to the bookings in the interval before. A relatively tight distribution in final bookings between S_1 and S_2 results. But if the correlation between intervals is poor (zf_2 is high), a much wider possible distribution of final bookings between W_1 and W_2 occurs. The implication of this phenomenon is an increased presence of *outliers* as correlation between intervals declines. Suppose knowledge of *b* bookings received by *h* yields an expectation of receiving *F* bookings by departure. As the correlation between booking intervals declines and the range of final demand increases, the likelihood of receiving *F* declines significantly.

Regression and pickup forecasting *do not* treat outliers equally: Pickup weights an outlier observation the same as all others at $\frac{1}{M-t}$, where M - t is all observations used by the forecaster. Regression forecasting is more influenced by outliers because the OLS regression technique minimizes mean square error (MSE) -- which grows with the *square* of the distance between the regression forecast and actual demand for each observation. Thus, outlier observations' weight on the slope will be *greater* than $\frac{1}{M-t}$.

This implies that the addition of an outlier observation to the HDB induces a larger absolute change in the regression than pickup forecast. Consider the linear regression curve $F(BIH[t])$ between bookings-to-come BTC and bookings-in-hand at interval h in Figure 6.4. The curve is upward sloping (indicating that more bookings-in-hand at interval h imply greater increases in BTC), and always passes through the means of the two variables -- in this case, through the point (B_h, P_d). Pickup forecasting assumes that bookings to come will be simply the mean of past BTC for HDB observations, or P_d. This yields the zero-slope line Pk representing the pickup forecast of BTC for any bookings in hand $BIH(h)$ level at interval h.

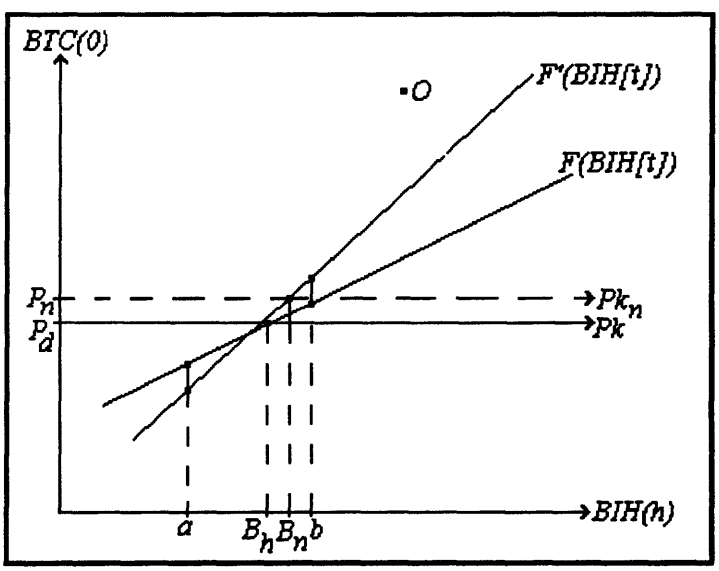


Figure 6.4. Pickup versus Regression Forecasting -- Treatment of Outliers

Now suppose an outlier O is added to the HDB. The minimum MSE property of regression significantly affects the regression forecast, yielding a new forecast curve $F'(BIH[t])$ with a significantly higher slope. Means for the two variables are now (B_n, P_n). Pickup forecast

increases by the differences of means P_n - P_d, yielding a new pickup forecast curve Pk_n. Under regression forecasting, however, the forecast has changed by *more* than this amount for all $BIH(h) \in (0,a) \cup (b,\infty)$. Thus, outliers have potentially significant impacts on regression forecasts. We therefore expect regression and pickup to perform differently without detruncation as variability in proportional booking relationships increases. This topic will be discussed in Section 6.1.1.2. As we will see, the *differential treatment of outliers* property has many implications for observed differences between forecasters..

- *Booking Curve Detruncation*

Next I turn to a comparison between pickup and regression forecasting when both airlines have booking curve detruncation. Referring again to Table 6.1, the challenge is to explain first why pickup and regression forecasters are almost exactly equal at all but the highest demand factors (where the latter's revenue performance declines significantly); and second, why scaling of the booking curve used for detruncation eliminates the revenue gap. Without detruncation, the "path of average bookings by interval *t*" for unclosed HDB observations is shown as ***U*** in Figure 6.5. A flight ***C*** which has departed and closed in interval *x* is to be added to the HDB[142].

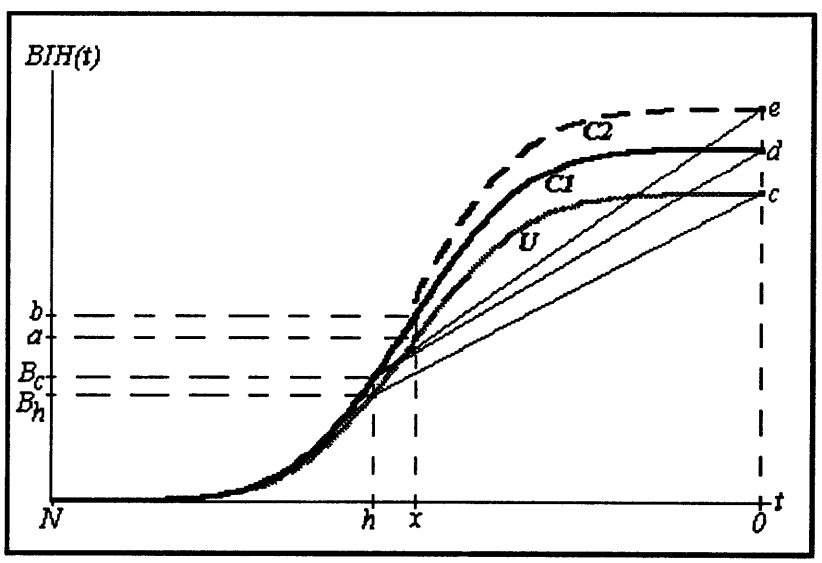


Figure 6.5. Booking Curve Detruncation

[142] Since the HDB size is constant at 52 observations, it is assumed throughout that *dropping* the last observation to add flight *C* has no effect on forecasts.

If booking curve detruncation without scaling is applied to this HDB observation between its closure interval x and the end of the booking process at 0, curve ***C1*** results. Final bookings d for the now-unconstrained observation ***C1*** are a multiple of average bookings on the unclosed flights ***U*** given by $d = \frac{1}{upbook_x} \cdot b = \frac{c}{a} \cdot b$. Given this situation, the curves ***U*** and ***C1*** cannot be parallel at any point in the booking process: Both have the same origin (0 bookings at interval N, the first booking interval) yet ***C1*** has more bookings by closure interval x. After interval x, the constant term $\frac{1}{upbook_x}$ is multiplied to b and a, again ensuring nonparallel curves: Two curves are parallel if their difference is an *additive* (not multiplicative) term. It follows that ***C1*** has a *steeper* curve than ***U***.

Now suppose we wish to forecast a future flight f which is presently at interval h. Pickup forecasting supposes that the increase in bookings for the forecast flight f is best approximated by the *average* bookings increase on HDB observations from the forecast interval h until departure. Among unclosed observations, the average increase between intervals h and 0 is $c - B_h$. With booking curve detruncation, closed observations like ***C1*** are now included. Since the slope of ***C1*** is always steeper than ***U***, we may conclude that booking curve detruncation always *increases* pickup forecasts, since the ray $B_c d$ is steeper than $B_h c$.

The change in regression forecast, however, is uncertain. Suppose that the HDB detruncated flight ***C1*** perfectly follows the booking relationships given by $upbook_t$ for all intervals t before closure at x (it will necessarily follow $upbook_t$ for periods after closure since it is detruncated using this information). In this case, detruncation *does not* change the slope of the BTC/BIH regression curve over not detruncating: The newly detruncated observation fits exactly on the regression line $F(BIH[t])$ (see Figure 6.4) calculated singly on unclosed observations! This analysis critically depends on whether "no detruncation" means *exclusion* of closed observations (as is assumed here and in PODS) or *inclusion* with no adjustment. In the latter case, booking curve detruncation without scaling will always change the regression curve, since the inclusion of closed observations with zero pickup from closure until departure obviously creates booking relationships that do not follow the $upbook_t$ relationships (specifically, $upbook_x = 1.0$ for the truncated flight, which is clearly implausible for $x > 0$).

Now suppose the HDB detruncated flight ***C1*** does not perfectly follow the $upbook_t$ relationships. This may occur for every booking interval before x, since bookings on the closed HDB flight *before closure* need not follow $upbook_t$[143]. Figure 6.6 repeats the $F(BIH[h])$ curve of Figure 6.4, and has a slope of $\frac{1}{upbook_h}$. If total bookings in hand $BIH(h)_w$ for the detruncated HDB observation w by interval h are *less* than $upbook_h$ percent of total estimated detruncated bookings $d = B\hat{I}H(0)_w$ but more than the average received for unclosed HDB observations by interval h (i.e., $BIH(h)_f > B_h$), the regression forecast between intervals h and 0 will increase. This occurs because the detruncated HDB observation falls in region *Ia* of Figure 6.6, thus shifting up the regression curve *and* increasing its slope[144]. A dark shading is applied to indicate this effect.

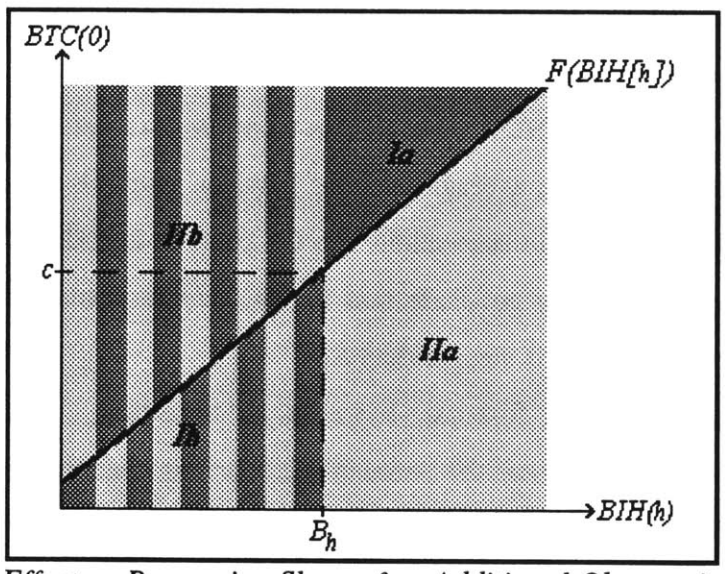


Figure 6.6. Effect on Regression Slope of an Additional Observation, by Region

But if more than $upbook_h$ percent of total bookings on the detruncated observation w are received by interval h yet less than the average bookings at h have been received (i.e., $BIH(h)_w < B_h$), the observation falls in region *Ib* and the slope of $F(BIH[h])$ increases -- but the regression curve also shifts down. The addition of an HDB observation in region *Ib* therefore has an

[143] It is a requirement that the forecast interval h for a forecast flight f be *before* the closure interval x on the closed HDB observation w for the booking curve detruncation of w to have this effect on the regression forecast. If $h = x$, the closed observation must follow $upbook_t$.

[144] This is true because if $BIH(h)_w > B_h$ and $BIH(h)_w / B\hat{I}H(0)_w < upbook_h$, if follows that $B\hat{I}H(0)_w > F(BIH[h])$, since $F(BIH[h]) = BIH(h) / upbook_h$.

uncertain effect on the regression forecast between intervals h and 0, and both dark and light shadings are used. Detruncated observations in this region are unlikely, since closure usually does not occur on a flight w with *fewer* than anticipated bookings by interval h. Ray cf in Figure 6.2 illustrates this type of relationship on a representative booking curve at interval h[145].

Now if the detruncated flight w being added to the HDB has higher bookings than average before closure without commensurably higher bookings after closure, i.e., $BIH(h)_w > B_h$ and $BIH(h)_w / B\hat{I}H(0)_w > upbook_h$, the detruncated observation falls in region *IIa*. This situation decreases the slope and shifts the regression curve downward, thereby decreasing regression forecasts. See ray af in Figure 6.2. Finally, if low bookings on the detruncated observation *before* interval h are combined with more than the expected $BIH(h)_w / upbook_t$ bookings afterward, the effect on regression forecasting in region *IIb* is again uncertain. The slope of the regression curve decreases, but the curve is shifted up (see ray cd in Figure 6.2).

Thus, the effect on regression forecasts of booking curve detruncation is almost completely *ambiguous*, depending entirely on the booking properties of detruncated relative to undetruncated flights. I have previously established that detruncated observations often come from flights with low bookings in early intervals and higher bookings in later intervals (Section 5.3.2). This suggests that many detruncated observations fall in the uncertain region *IIb* -- providing no further clues. Detruncated observations are also sure to include region *Ia* examples, i.e., flights which have higher than expected bookings throughout their open booking periods. These two facts suggest weakly that booking curve detruncation (like pickup) *increases* regression forecasts. More precise details must be inferred from the observed revenue and load differences.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
DF=0.7	9.3	4.3	2.8	50.9	67.2	9.1	4.4	2.6	51.6	67.6
DF=0.9	12.0	7.9	6.9	53.8	80.6	10.3	8.6	7.9	54.6	81.4
DF=1.2	17.0	19.5	22.4	35.0	94.0	18.1	38.7	28.1	8.2	88.2

Table 6.3. Loads for Booking Curve (No Scaling) Detruncation Scenarios

[145] To see this, note first that $c < B_h$ (i.e., less than average bookings by interval h). Second, ray cf has f total bookings at the end of the booking period, which is below the average e.

To this end, Table 6.3 shows load results for both airlines under the three demand factors (DF) when both carriers have booking curve detruncation without scaling. Because pickup (Airline B) unambiguously increases with booking curve detruncation, it protects significantly more than regression at high demand factors. This effect is clear from the dramatically reduced Q loads for Airline B at DF = 1.2 relative to the Airline A (see shaded boxes). While regression does protect additional seats for higher-valued fare classes (vertically-lined boxes), this increase is substantially less than for pickup. It appears that the weight of outliers under regression (with presumably more outliers at high demand factors) does not successfully counteract the inconsistent effect of booking curve detruncation on regression forecasts. Significantly more revenues for the airline with pickup results.

How does this analysis change if the booking curve for detruncation is also scaled? According to Table 6.1 and Figure 6.1, the pronounced revenue difference at DF = 1.2 between pickup and regression forecasting vanishes. As shown in Figure 6.5, scaling beyond closure interval x of the closed HDB observation pushes its detruncated curve out further to ***C2***. Final bookings for the detruncated HDB observation with scaling are $e = \frac{c}{a} \cdot \frac{1}{pbscl} \cdot b$. The pickup forecast increases (since ray $B_c e$ is steeper than $B_c d$), as does the tendency of the detruncated observation w to fall in regions *Ia* and *IIb* of Figure 6.6 (since $B\hat{I}H(0)_w$ increases). The effect on loads as *pbscl* decreases from 1.0 to 0.6 at DF = 1.2 is shown in Table 6.4. With detruncated HDB observations showing high bookings increases, pickup forecasting almost completely shuts down Q class and largely closes M class (shaded boxes). This stimulates further increases in B class bookings (lined boxes). Tightening low-value class availability by lower *pbscl* also causes a net decrease in total loads.

Airline A also experiences dramatic declines in Q class loads -- which is consistent with the hypothesis that detruncated observations in regions *Ia* and *IIb increase* regression forecasts. Curiously, M class loads first increase as *pbscl* decreases and then decline as B class loads increase. The phenomenon results from the *gradual closeout* of lower value fare classes as lower *pbscl* is applied: First Q class is closed when *pbscl* = *0.8* (shaded box in Figure 6.4), yielding increased bookings in M class (lined box). A lower *pbscl* = *0.6* repeats this same pattern between M and B classes.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
pbscl=1.0	17.0	19.5	22.4	35.0	94.0	18.1	33.7	28.1	8.2	88.2
pbscl=0.8	19.4	26.3	26.8	20.9	93.3	19.8	42.3	20.9	2.9	85.8
pbscl=0.6	22.6	39.0	16.1	12.5	90.2	20.2	50.0	11.7	1.0	83.0

Table 6.4. Loads for Booking Curve (DF=1.2) Detruncation Scenarios, Variable Scaling

The gradual tightening of low-value class availability explains why pickup forecasting's revenue advantage over regression declines to approximately zero with lower *pbscl*. Without scaling, regression forecasting is a poor performer because it does not protect enough seats for high-valued passengers, allowing low-value passengers to flock to it. As *pbscl* decreases, the increased observations in *Ia* and *IIb* create more *outliers* to which the regression curve aggressively responds by increasing protection levels for high-value classes. Passengers are left without an inexpensive alternative, and must sell-up. Further tightening by pickup forecasting fails to maintain its revenue lead because successively more protection begins to have *diminishing returns*: Instead of reducing booking limits for low-value fare classes, they are simply made *unavailable* -- yielding little *additional* sell-up. Regression forecasting results in a passenger distribution with more bookings in lower-value fare classes, but this is offset by the fewer total bookings pickup forecasting receives with almost complete closure of low-value classes.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
pbscl=1.0	12.0	7.9	6.9	53.8	80.6	10.3	8.6	7.9	54.6	81.4
pbscl=0.8	12.9	9.1	9.9	49.0	80.9	10.9	10.7	13.8	45.6	81.0
pbscl=0.6	14.1	11.0	19.5	36.6	81.2	12.1	14.0	27.5	25.8	79.4

Table 6.5. Loads for Booking Curve (DF=0.9) Detruncation Scenarios, Variable Scaling

Curiously, this phenomenon occurs only at high demand levels. Under more reasonable demand conditions (DF = 0.9), pickup and regression forecasting have approximately equal revenue results regardless of scaling. Table 6.5 details load results for DF = 0.9, and indicates that regression and pickup forecasting have substantially the same forecasts. Lower *pbscl* again induces the differential protections (pickup directs significantly more Q class passengers to M than regression forecasting), but this is offset as Airline A retains equality with its competitor in

high-class bookings, and offsets revenue losses induced with lower M class loads by more Q class passengers. Thus, Airline A's loads *increase* with scaling, while Airline B's loads decrease.

Moderate demand levels prevent the dramatic *underprotection* of regression relative to pickup forecasting because closed observations are much less likely. With fewer observations to operate on, the load differences induced by detruncation do not occur (recall Table 6.2, which suggests that *without* detruncation, pickup and regression forecasts are equal). This revenue equality is maintained if scaling is applied to booking curve detruncation): Scaling does induce increased closure of low-value fare classes under pickup forecasting, and makes *outliers* of the detruncated observations for regression forecasting. But the "super-weighted" treatment of outliers under regression and the few observations to which this technique is applied ensures the retention of revenue equality.

- *Pickup Detruncation*

Booking curve detruncation adopts the approach that *proportional* booking relationships are constant between closed and unclosed HDB flights. Pickup detruncation instead assumes that the *absolute* increase in bookings from closure interval x to departure on the closed HDB observation is best approximated by the *average* increase in bookings between these time periods on HDB flights which do not close. This situation is illustrated by Figure 6.7, where the increase in bookings *after closure* interval x on a closed HDB flight $\boldsymbol{C}$ is the mean pickup $c - a$ in bookings between x and 0 for unclosed HDB flights $\boldsymbol{U}$. The difference between the curves $\boldsymbol{C}$ and $\boldsymbol{U}$ is therefore constant at $b - a = d - c$ between intervals x and 0.

As is apparent by inspection of the curve, this results in pickup detruncation yielding low estimates of final bookings d on closed HDB flights -- certainly lower than booking curve detruncation (compare Figure 6.7 and 6.5). Suppose a flight f is being forecast from booking interval h. In Figure 6.7, if the forecast interval h is before the closure interval x for the detruncated HDB flight $\boldsymbol{C}$, pickup detruncation results in a *slight* increase in the pickup forecasting estimate of final bookings $B\hat{I}H(0)_f$ for the forecast flight f. This is true because the slope of the ray $B_c d$ representing pickup from interval h for the detruncated HDB flight $\boldsymbol{C}$ is marginally *greater* than the slope of $B_h c$, representing average pickup for all unclosed HDB

flights[146]. The effect on regression forecasting is uncertain for reasons already described. Relative to booking curve detruncation, the regression forecasts under pickup detruncation will certainly be *lower*.

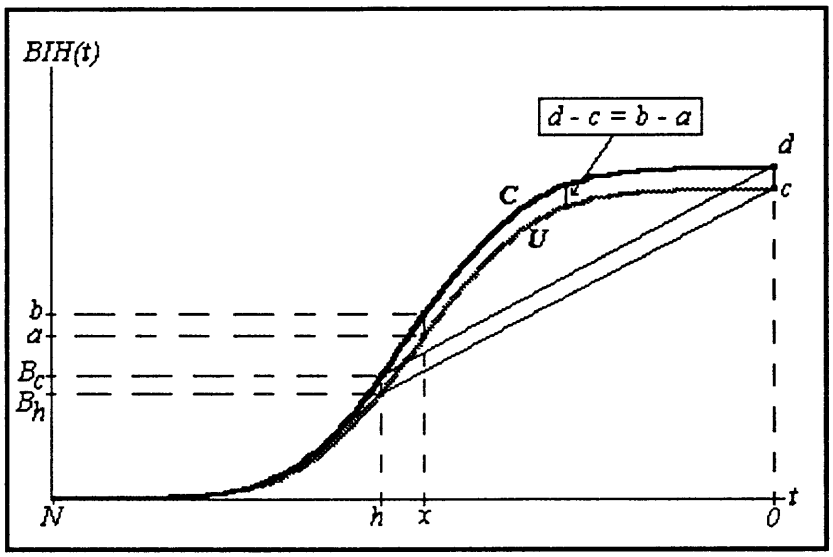


Figure 6.7. Pickup Detruncation

In Table 6.6, loads are approximately equivalent for both forecasters. This is explained as another result of Figure 6.7: A pickup detruncated HDB observation w will not have substantially different estimated final bookings $B\hat{I}H(0)_w$ than the mean final bookings $BIH(0)$ for all unclosed observations. No *outliers* are created, which are treated differently according to forecasting method.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
DF=0.7	9.2	4.3	2.7	51.1	67.3	9.1	4.4	2.7	51.0	67.1
DF=0.9	10.2	5.6	3.3	62.0	81.0	10.1	5.7	3.3	62.0	81.0
DF=1.2	11.3	7.7	3.9	69.2	92.0	11.0	7.6	3.9	69.4	91.9

Table 6.6. Loads for Pickup Detruncation Scenarios

This explains why regression does not underperform pickup forecasting at high demand factors, as it does when booking curve detruncation is used. There is simply little impact of the "larger" forecast slope with inclusion of the detruncated HDB observation. Comparing Table 6.6

[146] However, if the forecast interval h for the flight being forecast is equal to the closure interval x of the detruncated HDB flight C, pickup detruncation *has no effect* on the pickup forecast.

and Table 6.2 (loads with *no* detruncation), we see almost no difference -- except for a modest increase in Y and decrease in Q bookings (as shown by the lined and shaded boxes, respectively). The implication is that pickup detruncation is marginally better than no detruncation at all. I will return to this topic in Section 6.2 below.

- *Projection Detruncation*

Finally, we have projection detruncation. In Figure 6.8, this detruncation method predicts the increase in bookings for a closed HDB observation between booking intervals h to 0, given closure at interval x ($x < h$). Following the terminology of variables already introduced, interval h is the forecast interval for a flight f that is being forecast. Interval x is the closure interval on an HDB observation w which closes[147]. Between intervals h and x, $b - B_c$ bookings are received on w. The average increase in bookings for HDB flights which remain open over these intervals is $a - B_h$. Then the average increase on unclosed HDB observations from intervals x to 0 is $c - a$, yielding average total "pickup" from intervals h and 0 for unclosed HDB flights of $\mu_t = c - B_h$. Projection detruncation estimates final estimated unconstrained pickup $\underline{cl}$ between the forecast interval h and 0 on the closed HDB flight w given this information and τ, an input conditional probability (see Section 5.3.3).

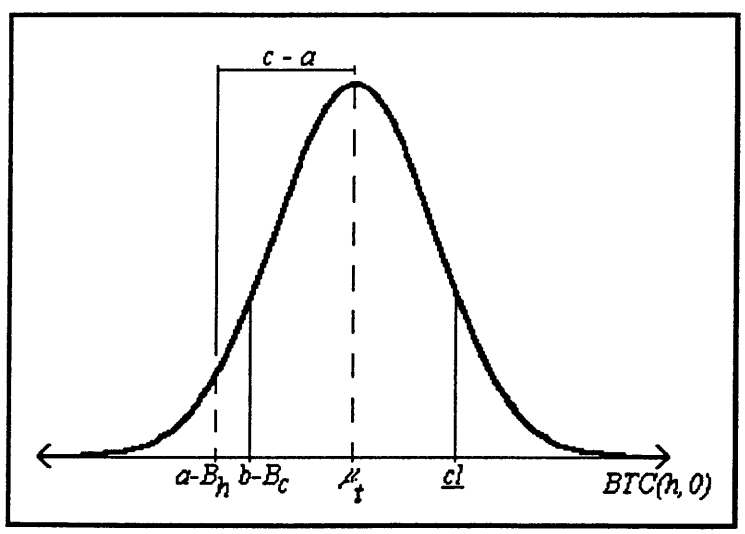


Figure 6.8. Projection Detruncation for a Closed Flight

[147] Projection detruncation may also be used when more than one observation in the HDB is closed.

In this situation the several input variables complicate general statements about regression or pickup forecasts under projection detruncation. Effects on these forecasts may be inferred by examination of the results of the projection detruncator relative to pickup and booking curve detruncation already discussed. First I compare projection and pickup detruncation. Projection estimates of final $B\hat{I}H(0)_w$ for the closed HDB flight w will be *greater* than the pickup detruncation estimate (i.e., $b - B_c + c - a < \underline{cl}$) if $a - B_h \approx b - B_c$ and $\tau < 0.5$. The latter condition will always be true (see Section 5.3.3); the former is not likely to be true (bookings received up until forecast interval h are usually higher on HDB observations which later close than those which do not). As $a - B_h < b - B_c$, the projection detruncation estimate $B\hat{I}H(0)_w$ on the closed HDB flight w tends to *decrease* relative to pickup detruncation. Tests of sample values indicate that the projection detruncation estimate of unconstrained demand will be larger than pickup detruncation if $b - B_c$ is not larger than a relatively generous 150-175% of $a - B_h$. This condition is robust under a wide range of values for the other variables. As $\tau \rightarrow 0$, the projection detruncated estimate increases relative to pickup detruncation; our base case $\tau = 0.15$ is relatively low. Based on this information, pickup detruncation should consistently yield *lower* detruncated estimates than projection.

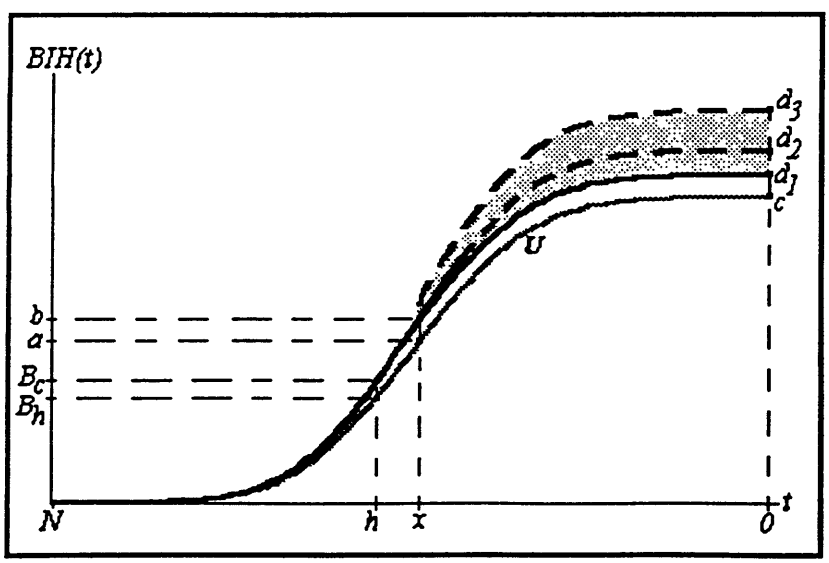


Figure 6.9. Projection Detruncation

Projection detruncation may be compared to booking curve detruncation using Figure 6.9. Unclosed observations U in the HDB have, on average, c total bookings by the end of the

booking process. Repeating the detail of Figures 6.7 and 6.5, pickup detruncation will estimate d_1 total bookings on the closed HDB flight w. Booking curve detruncation without scaling estimates d_2 final bookings, and d_3 with scaling. I have established that projection detruncation generally estimates $B\hat{I}H(0)_w$ higher than pickup detruncation at d_1. Since increasing τ causes projection estimates of $B\hat{I}H(0)_w$ to increase, projection detruncation may be anywhere within the shaded area of Figure 6.9, and even higher than d_3 for sufficiently small τ.

Table 6.7 (loads for both forecasters under projection detruncation) indicates no especially consistent difference in bookings between forecasters. At DF = 1.2, regression forecasting outperforms pickup forecasting for Y class bookings -- but it underperforms on B and M classes, and its excess of Q bookings suggests slight underprotection relative to Airline B's pickup. On a revenue basis, regression outperforms pickup forecasting, which increases with demand factor (Figure 6.1). More surprising is the *extreme* shift in bookings from Q-class to M and especially B class with high demand and projection detruncation. This far surpasses other detruncation methods in the degree of fare class distribution shift as demand factor increases (compare Table 6.7 with Tables 6.6, 6.5, and 6.4). Projection is apparently *very* sensitive to high demand situations.

This suggests that projection detruncation responds to high demand conditions by causing extreme outliers in the HDB dataset (i.e., projected demands on closed HDB flights are greater than d_3 in Figure 6.9). Such reasoning is consistent with the observed superiority of regression over pickup forecasting under projection detruncation: Estimating $B\hat{I}H(0)_w$ on a closed flight w beyond the level d_3 (which occurs under booking curve detruncation with extreme scaling at *pbscl* = *0.6*) further improves the performance of regression forecasting due to the disproportionate weight of outliers.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
Scenario	Y	B	M	Q	Total	Y	B	M	Q	Total
DF=0.7	9.2	4.3	2.7	50.8	67.1	9.1	4.3	2.7	51.0	67.1
DF=0.9	11.1	6.8	4.6	58.8	81.3	10.5	6.9	4.6	59.3	81.4
DF=1.2	22.0	42.8	14.2	9.0	88.0	19.4	43.9	18.9	5.0	87.1

Table 6.7. Loads for Projection Detruncation Scenarios

- *Summary*

The following may be concluded from this discussion of base case and high/low demand scenarios:

- Neither regression nor pickup dominates revenue results. The difference between the two is dependent both on underlying demand and detruncation method choice.
- Detruncation tends to introduce outliers, though the extent to which these observations deviate varies significantly by detruncation method.
- Because regression gives disproportionate weight to outliers, it tends to perform better when large outliers are associated with high demand situations.
- Applying detruncation always increases pickup forecasts, but not necessarily regression forecasts.
- Substantial differences between the forecasters emerge only under moderate and high demand conditions.

6.1.1.2. High/Low Booking Curve Variability Scenarios

The discussion surrounding Figure 6.3 indicated two facts: First, *outliers* in the HDB are more likely as the correlation in bookings between intervals decreases. Second, regression and pickup forecasting do not treat outliers equally. This section tests the revenue effect of increasing "variability around the booking curve" to determine how this phenomenon changes the revenue comparisons reported in Table 6.1. Table 6.8 gives revenue results as the z-factor zf_2 varies from 1 to 4, where the base zf_2 is 2 (see Section 4.3 for a discussion of the influence of zf_2 on variability around the booking curve). The low-demand scenario has been suppressed for lack of significant revenue differences.

		% Rev Difference between Airline with Pickup over Airline with Regression, by Detruncation Method Pair					
DF	zf_2	None	BkCrv(0.6)	BkCrv(0.8)	BkCrv(1.0)	Projctn	Pickup
0.9	1.0	-0.22%*	-0.11%	-0.36%*	-2.90%*	-0.60%*	-0.41%*
	2.0	**-0.05%**	**0.19%**	**-0.13%**	**-0.74%***	**-0.57%***	**-0.12%**
	4.0	-0.10%	1.92%*	0.98%*	0.31%	-0.47%*	-0.32%
1.2	1.0	-0.76%*	-0.33%*	2.26%*	4.18%*	-1.34%*	-0.49%*
	2.0	**-0.50%***	**-0.45%***	**4.89%***	**7.95%***	**-1.54%***	**-0.49%***
	4.0	-0.37%*	0.43%*	11.89%*	16.31%*	-0.76%*	-0.02%

Table 6.8. Percentage Revenue Differences between Pickup and Regression under Variable zf_2

These results consistently show that increasing variability from booking curve relationships specified by $upbook_t$ causes the performance of regression forecasting to *deteriorate*. This is true regardless of the detruncation method considered, and is magnified at high demand factors. The revenue underperformance of regression is exacerbated when the selected detruncation method is booking curve detruncation *without scaling*. Curiously, the revenue difference almost completely vanishes as more scaling is applied (see two shaded regions in Table 6.8). To explain these results, Table 6.9 details loads for booking curve detruncation with *pbscl* = *0.6* and *1.0* (extreme and no scaling, respectively). I examine the high demand scenario, where the revenue effect of not scaling is most pronounced.

		Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
pbscl	zf_2	Y	B	M	Q	Total	Y	B	M	Q	Total
0.6	*1.0*	23.4	30.3	21.3	14.9	89.9	20.5	41.9	19.0	1.2	82.7
	2.0	**22.6**	**39.0**	**16.1**	**12.5**	**90.2**	**20.2**	**50.0**	**11.7**	**1.0**	**83.0**
	4.0	21.1	46.8	11.9	11.4	91.1	19.4	58.0	6.6	0.9	84.8
1.0	*1.0*	20.1	17.9	23.7	31.7	93.3	18.2	30.1	31.2	8.6	88.1
	2.0	**17.0**	**19.5**	**22.4**	**35.0**	**94.0**	**18.1**	**33.7**	**28.1**	**8.2**	**88.2**
	4.0	12.3	17.8	17.2	47.4	94.6	17.2	37.2	21.8	12.3	88.6

Table 6.9. Loads with Variable zf_2, Booking Curve Detruncation, pbscl = 0.6 or 1.0, DF = 1.2

Without scaling, regression performs worse than pickup forecasting as variability over the booking curve increases because their bookings by fare class (and thereby revenues) move in *opposite* directions. Airline A (with regression forecasting) loads in Y and M class decline precipitously, while Q class loads increase. In complete contrast, Airline B's Y and Q class loads are not affected by increased zf_2, while M class loads decline significantly and B class loads increase. When scaling is applied, regression assumes the same fare class distribution as pickup forecasting: M loads decline while B class loads increase, with relatively little movement in other classes.

The absolute revenue effect of these dramatic differences is illustrated in Figure 6.10, which shows the revenues achieved by Airline A (with regression) and Airline B (with pickup) for each of the three zf_2 levels by detruncation method. From this graph we may determine how increasing variability around the booking curve affects the *absolute* revenues achieved by the

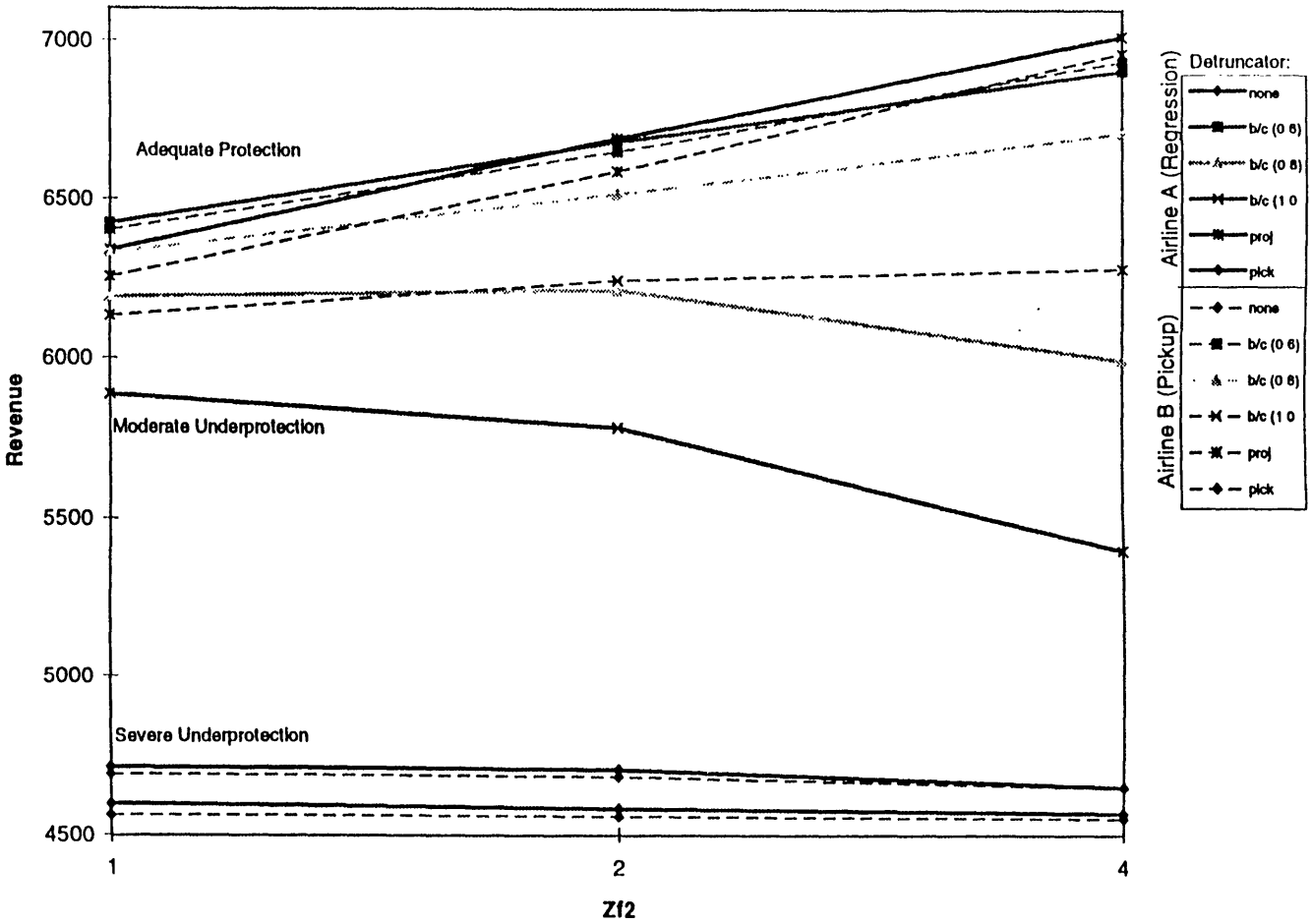
forecasters, not the proportionate difference between them at each zf_2. There are many revenue lines which complicate reading of the graph, but they fall into three easily identifiable groups. First are no detruncation and pickup detruncation, which *severely underprotect* for higher-valued fare classes. Regression and pickup forecasting yield equivalent revenue results, and revenues *decline slightly* as zf_2 increases.

Second, projection and booking curve detruncation with *pbscl = 0.6 adequately protect*: Revenues are again approximately equal between regression and pickup forecasting. Moreover, they *increase* with increasing booking curve variability! Finally, intermediate cases (booking curve detruncation with little or no scaling) offer *moderate underprotection*. The only significant differences between forecasting methods are exhibited here: Regression performs *worse* than pickup forecasting. The difference increases with larger zf_2. It is apparent from Figure 6.10 that the revenues resulting from the *initial* fare class distribution due to forecaster and detruncator choices is most relevant in predicting the effects of variation in proportional booking relationships.

The causes of these revenue differences are not clear. If revenues increase simply because of the presence of additional HDB outliers of final bookings as zf_2 increases, causing forecasters to adjust by increasing protections (e.g., booking curve with scaling in Table 6.9), why does it apply only to B and M classes? The *positive* outliers (where present bookings-in-hand for an HDB observation are less than $upbook_t$ percent of final bookings, represented as ray *cd* in Figure 6.2) implicit in this explanation are no more likely to occur than *negative* outliers, which have the opposite effect on forecasting.

In fact, in a detruncated HDB *negative* outliers are more likely than positive outliers. The latter, with unexpectedly high later bookings, are much more likely to close (see Section 5.3.2).Increasing zf_2 magnifies the difference between expected booking relationships $upbook_t$ for each booking interval t and actual relationships for the *positive* outlier HDB observation w. Thus, $upbook_t >> upbook_t^w$. Therefore, relative to a moderate zf_2, all detruncation methods -- even with scaling -- subsequently applied to a closed outlier HDB flight w tend to *underestimate* the actual $upbook_t^w$. Negative outliers, however, fail to close -- which suggests that protections for higher-valued fare classes *decline* as zf_2 increases. This is implausible, given the observed load

Figure 6.10. Revenue Performance for Variable Zf2 by Airline and Detruncation Choice, DF = 1.2



shift toward higher-valued fare classes in Table 6.9. At present, no consistent explanation may be offered for observed revenue shifts as zf_2 is varied.

The following conclusions may be drawn from this discussion of pickup versus regression forecasting as variability around the booking curve changes:

- Regression deteriorates relative to pickup forecasting as booking curve variability increases.
- This difference is magnified at high demand factors.
- The difference is almost completely suppressed if booking curve detruncation with scaling is the chosen detruncation method.

6.1.1.3. High/Low System K-factor Variability Scenarios

Section 4.3 described the effect of system k-factor *skf* on total demand for each observation. A larger *k*-factor increases the variability between observations of demand by passenger type. Testing several *skf* is of interest because regression and pickup forecasting may not equally adjust to this variation. To this end, alternative *skf* values of *0.1* and *0.5* were tested in addition to base case *skf* = *0.3* at DF = 0.9. The high demand case will not be discussed, since it largely repeats the results of DF = 0.9. Revenue results are listed in Table 6.10.

	% Rev Difference between Airline with Pickup over Airline with Regression, by Detruncation Method Pair					
skf	None	BkCrv(0.6)	BkCrv(0.8)	BkCrv(1.0)	Projctn	Pickup
0.1	-0.12%	0.06%	0.02%	-0.05%	-0.02%	0.00%
0.3	**-0.05%**	**0.19%**	**-0.13%**	**-0.74%***	**-0.57%***	**-0.12%**
0.5	-0.71%*	-0.49%*	-2.85%*	-6.73%*	-1.76%*	-0.58%*

Table 6.10. Percentage Revenue Difference under Variable System K-factor; DF = 0.9

It is clear by inspection that increasing systemwide demand variability benefits regression forecasting at the expense of pickup forecasting, regardless of detruncation method chosen. The relative loss due to pickup forecasting is significantly influenced by the detruncation method. As indicated by the shaded regions of Table 6.10, the difference is reduced under booking curve detruncation when scaling is applied. When there is little variation in demand between observations (*skf* = *0.1*), there is no statistically significant difference between regression and

pickup forecasting under any detruncation method choice. These results strongly suggest that *differential treatment of outliers* is responsible for the observed revenue differences.

This effect is demonstrated in Table 6.11, which compares loads under each *skf* when booking curve detruncation without scaling is the selected detruncator. Here, both airlines record moderate declines in bookings as systemwide demand variability increases. Airlines A and B begin with equivalent bookings in all classes. However, Airline A's declines come completely from the lowest-value Q class, while higher-valued class bookings *increase*. Pickup forecasting, in contrast, records decreases in both the highest-value Y *and* the lowest-value Q classes.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
skf	Y	B	M	Q	Total	Y	B	M	Q	Total
0.1	11.6	5.7	7.9	60.8	86.0	11.0	6.4	8.7	59.7	85.8
0.3	12.0	7.9	6.9	53.8	80.6	10.3	8.6	7.9	54.6	81.4
0.5	13.3	8.9	6.6	44.5	73.2	7.9	7.8	5.7	54.3	75.6

Table 6.11. Loads under Variable skf; Booking Curve Detruncation (pbscl=1.0); DF = 0.9

These results may be explained by the differential approach adopted by regression and pickup forecasting: the former assumes an equality of *proportional* booking relationships per interval t among flights, and the latter assumes equality of *absolute* bookings increases from forecast interval t until departure. Since varying *skf* has (on average) no effect on the booking relationships specified by $upbook_t$, the regression curve will not be significantly affected by increasing demand variability. Its forecast depends on the detruncation model chosen, as has already been discussed.

The pickup forecast will be *negatively* affected if increased *skf* reduces the average increase in bookings for HDB flights between any forecast interval t and 0. This is exactly what occurs: The increased number of HDB outlier observations with unusually *low* pickup do not close, and therefore lower the unclosed HDB average estimate of pickup between intervals t and 0. Clearly, the (now increased number of) observations with *high* pickup are more likely to be closed. So without detruncation, pickup forecasts decline relative to regression forecasts.

With the adoption of detruncation methods, outlier HDB observations with high pickup may also be included in forecasts. At higher *skf*, observations w that close are more likely to have exhibited exceptionally high final bookings $B\hat{I}H(0)_w$. This *magnifies* the degree to which

detruncated HDB observations differ from mean final bookings levels for unclosed flights. Since regression forecast *disproportionally* weights outliers in the HDB, the clear effect is increased protection for regression over pickup forecasting. This is true for all detruncators, which explains regression's consistently superior performance at *skf = 0.5*. However, how is it then the case that applying scaling *reduces* revenue differences? If regression benefits because it disproportionally weights HDB outliers, this effect should be *magnified* and not reduced when extreme scaling is applied to closed HDB observations.

	Passenger Loads, Airline A With Regression					*Passenger Loads, Airline B With Pickup*				
pbscl	Y	B	M	Q	Total	Y	B	M	Q	Total
1.0	13.3	8.9	6.6	44.5	73.2	7.9	7.8	5.7	54.3	75.6
0.8	14.2	9.8	10.9	38.6	73.4	8.8	11.9	13.7	41.2	75.5
0.6	15.3	11.7	17.5	29.4	73.8	10.0	17.3	25.0	21.4	73.7

Table 6.12. Loads with Variable Scaling under Booking Curve Detruncation; skf = 0.5; DF = 0.9

To explore this issue further, Table 6.12 has load information for each *pbscl* under high *skf*. The extreme revenue differences under *pbscl = 1.0* are apparently due to the superior load distribution for Airline A with regression forecasting, which is consistent with the outlier hypothesis. As scaling is applied Airline A benefits -- higher-value class loads increase (lined boxes), but Airline B benefits much *more*. Both the decline in its Q class bookings and the increase in higher-value class bookings are more significant. Thus, by *pbscl = 0.6* there is a barely significant revenue difference between the two airlines (see Table 6.10), even though regression has significantly more Y and Q class bookings than the pickup carrier. This result is analogous to the "catch-up" which regression forecasting undergoes when scaling is applied in the base case (see Table 6.4): Revenue results are equivalent, though the load mix is not.

Under booking curve detruncation, then, scaling functions as a *revenue equalizer* for forecasting models: The forecaster which underperforms due to the vagaries of booking curve detruncation gains differentially when scaling is applied. This differential gain may occur because of a limited potential of sell-up among passengers. Table 6.4 (under high DF) indicated that further gains to pickup forecasting from overprotection were limited, since lower-value fare classes were all but closed. In the same fashion, further gains to regression forecasting in Table

6.12 may be limited because underlying demand conditions do not permit significant further sell-up among M and Q class passengers.

From this final analysis of regression versus pickup forecasting, we may conclude the following when systemwide demand variability increases:

- Relative to regression, pickup forecasting underperforms on a revenue basis.
- The difference is mostly eliminated if booking curve detruncation with scaling is the chosen detruncation method.
- Adding scaling tends to equalize differences between forecasters, possibly because of the limited potential of sell-up.

6.1.2. Efficient/Pickup Comparison

The pairwise comparison between the efficient and pickup forecaster proceeds in the same fashion as with regression and pickup (Section 6.1.1), with one significant difference: Since the efficient forecaster requires the use of booking curve detruncation (Section 5.2.3), the alternative detruncators assumed by both forecasting methods are reduced to three, and vary only by choice of scaling *pbscl.*

6.1.2.1. Base Case and High/Low Demand Factor Scenarios

Revenue comparisons under base conditions are given for three demand factors in Table 6.13 below. Pickup forecasting consistently outperforms the efficient forecaster, though scaling again tends to reduce the revenue difference between forecasters (especially at high DF). These results are not encouraging for the efficient forecaster, and are contrary to initial expectations that it would yield superior results because of its "full-information" property (i.e., it uses all available information from the HDB to forecast).

The explanation for efficient forecasting's persistent underperformance and improvement with scaling will, if consistent with the results in Section 6.1.1, show that the efficient forecaster exhibits two central characteristics. First, without scaling it increasingly *underprotects* for high-value fare classes relative to the pickup forecaster as demand factor increases. Second, either it lends outliers a differential weight, or sell-up opportunities with pickup forecasting diminish as scaling is applied -- thus allowing it to "catch up." The presence of the first characteristic is

examined in Table 6.14, which gives loads by fare class at three demand factors when booking curves are not scaled.

	% Rev Diff between Airline with Pickup over Airline with Efficient		
DF	Bk Crv (0.6)	Bk Crv (0.8)	Bk Crv (1.0)
0.7	-0.12%	0.32%	0.12%
0.9	1.25%*	3.35%*	3.02%*
1.2	-0.47%*	3.05%*	8.10%*

Table 6.13. Pickup versus Efficient Forecasting Relative Revenue Performance

These results are entirely consistent with *underprotection* by the efficient forecaster. Both forecasters have approximately equivalent results at DF = 0.7. By DF = 0.9, most of the increase in bookings for Airline B (with pickup) occur in B and M fare classes (lined boxes); Pickup forecasting has predicted more bookings for these classes and thereby limited any increase in Q class. In contrast, Airline A with the efficient forecaster does not protect significantly more seats for B and M classes, causing the bulk of *its* increased loads to occur in Q class (lined box). Under extreme demand conditions, the differences are more significant: The pickup forecaster has tightly constricted Q-class bookings (shaded box), and high demand causes increases in all high value classes. While the efficient forecaster experiences a similar phenomenon, it is clear that it has *underprotected* for high-value classes relative to pickup forecaster. Both the decline in Q class and increase in B and M class bookings are much smaller.

	Passenger Loads, Airline A With Efficient					*Passenger Loads, Airline B With Pickup*				
DF	Y	B	M	Q	Total	Y	B	M	Q	Total
0.7	9.1	4.3	2.9	50.8	67.1	9.1	4.4	2.6	51.2	67.3
0.9	**9.5**	**6.7**	**3.9**	**61.4**	**81.5**	**10.8**	**8.0**	**7.5**	**54.3**	**80.6**
1.2	16.1	22.0	27.7	30.9	93.7	17.8	30.9	30.9	9.1	88.6

Table 6.14. Efficient and Pickup Loads with Variable DF, pbscl = 1.0

The primary reason for the underperformance of the efficient forecaster probably lies not in its "full-information" property, but in the weighting system applied to observations which are truncated. Recall from the description in Section 5.2.3 that the efficient forecaster detruncates all HDB observations which either close or have incomplete booking histories using booking curve detruncation (with scaling as appropriate). A detruncated HDB observation is weighted

according to equation (5.10) by correlation coefficients $CC_i^2(j)$ which measure the consistency of proportional relationships between bookings at truncation interval j and the end of the booking process, calculated over all unclosed flights in the HDB. This process is repeated for each truncated observation. Next, a weighted average of final bookings is calculated over *all* HDB observations. This is then combined in equation (5.12) with a detruncated estimate $\widetilde{BIH}(0)_f$ of bookings on the forecast flight f, based singly on booking curve detruncation applied to present bookings on this flight in the forecast interval t. The weight WP assigned to the estimate $\widetilde{BIH}(0)_f$ is governed by the correlation coefficient between intervals t and 0, as shown in equation (5.11).

The justification for weighing HDB observations with final bookings data that must be inferred *less* than those which can be directly observed[148] is that detruncation (the inference procedure) is inexact. This intuitively appealing system ensures that the most confidence is placed in observations for which we have complete data. However, it also creates a systematic *downward* bias in the weighted average of final bookings on HDB flights. High demand flights are obviously more likely to close. Thus, they will always be weighted *less* than low and moderate demand flights which never close. This effect is magnified if the closure interval occurs earlier in the booking process: The correlation coefficient naturally declines when calculated between earlier booking intervals and interval 0, since there are more intervals over which variation from the proportional relationships specified by $upbook_t$ may occur (see Section 6.1.1.1 and Figure 6.3). As demand for a flight increases, its closure will occur *earlier* (assuming stability of the underlying booking curve). High demand flights are therefore doubly likely to be weighted less than low demand flights: once because they are more likely to close, and again because they are more likely to close *early*.

It is therefore apparent that the efficient forecaster underperforms because it *underprotects*. This is primarily due to the present assumption that the *cause* of an incomplete booking history should not be used in the weighting mechanism. If the efficient forecaster were adjusted to reflect the fact that HDB observations which *close* tend also to be those with high

[148] Recall that a flight g with a complete, unclosed booking history will already have $BIH(0)_g$. In these circumstances the correlation coefficient is 1. See Section 5.2.3.

demands (as opposed to flights which are truncated because they have not yet reached interval *0* in their booking process), these results might be significantly different. Conclusions drawn here are also tempered by the limited consideration of the efficient forecaster in PODS. The forecaster's ability to use the incomplete booking histories of "future flights" which depart *after* the forecast flight was not exploited[149].

In any case, the underprotection argument still has one ambiguity: How can the efficient forecaster "catch-up" in revenues when scaling is applied to booking curves? Certainly it does not super-weight outliers like regression; instead closed observations are weighted *less*. The fact that the efficient forecaster's weighting scheme is not sensitive to *scaling* applied to detruncated observations indicates that scaling unquestionably results in higher forecasts. Obviously, however, this will not occur to the extent of regression forecasting. We are therefore left with the diminishing returns to scaling argument.

	Passenger Loads, Airline A With Efficient					*Passenger Loads, Airline B With Pickup*				
pbscl	Y	B	M	Q	Total	Y	B	M	Q	Total
0.6	23.0	46.2	14.6	1.8	85.7	19.4	50.6	15.2	0.7	85.8
0.8	17.4	29.8	34.8	10.7	92.7	19.2	39.5	26.1	2.5	87.2
1.0	13.1	22.0	27.7	30.9	93.7	17.8	30.9	30.9	9.1	88.6

Table 6.15. Efficient and Pickup Loads with Variable pbscl, DF = 1.2

Table 6.15 shows loads for the three scaling levels at DF = 1.2, where the revenue consequence of not scaling is most apparent. Without scaling, Airline B with pickup forecasting has already induced significant sell-up, thus allowing few Q class bookings. Airline A with efficient forecasting, however, has many Q class bookings. As scaling is performed, Airline B induces significant sell-up from M to B class. Airline A is still significantly behind at *pbscl = 0.8*, but it has achieved substantially equivalent loads when maximal scaling is reached. Airline B has clearly exhausted much of its sell-up possibilities by *pbscl = 0.6*: Q class is entirely closed and M not so far behind. It seems apparent that diminishing returns to more protection for higher-valued fare classes effectively prevents the pickup forecaster from maintaining its revenue superiority over the efficient forecaster. To summarize:

[149] Suppose our forecast flight f is at interval h in the booking process. A "future flight" is one which departs *after* flight f, e.g. $f + 1$. Flight $f + 1$ will be at interval $h + 1$ in its booking process, assuming booking intervals of equal length. The booking histories of future flights are *less complete* than the forecast flight's.

- The efficient forecaster generally underperforms pickup forecasting. This situation worsens as demand factor increases.
- Underperformance of the efficient method is likely due to a weighting system with an inherent downward bias.
- Revenue equality is attained when maximal scaling of booking curves is applied. This is probably due to *diminishing returns* to scaling.

6.1.2.2. High/Low Booking Curve Variability Scenarios

When more variability around the booking curve is exhibited, the relationships specified by $upbook_t$ are less likely to be followed. PODS evaluates this effect by the z-factor variable zf_2, as previously discussed. With increasing zf_2, more flights deviate from the proportional booking curves. Section 6.2.2 explains that in these circumstances the number of closed flights in the HDB increases. These closed flights are disproportionally likely to be "early-low, high-late" bookings flights (i.e., booking curve ***C*** in Figure 5.1). Because closed or truncated flights are insufficiently weighted by the efficient forecaster, its underprotection liability should be magnified as zf_2 increases. Table 6.16 indicates that this is so: Not only does its performance decline precipitously with increasing variability around the booking curve, but the effects are generally worse at high DF.

		% Rev Diff between Airline with Pickup over Airline with Efficient Forecaster		
DF	zf_2	Bk Crv (0.6)	Bk Crv (0.8)	Bk Crv (1.0)
0.9	1.0	0.15%	2.05%*	2.31%*
	2.0	**1.25%***	**3.35%***	**3.02%***
	4.0	2.96%*	4.09%*	3.19%*
1.2	1.0	-1.15%*	0.59%*	5.11%*
	2.0	**-0.47%***	**3.05%***	**8.10%***
	4.0	-0.49%*	8.08%*	12.25%*

Table 6.16. Pickup versus Efficient Forecasting Relative Revenue Performance under Variable zf_2

Underperformance is further ensured by an *intensification* of the weighting bias against high demand HDB observations. A large zf_2 creates numerous deviations from proportional booking relationships specified by $upbook_t$. Thus, an HDB observation g is less likely to have final bookings $BIH(0)_g$ related to bookings-in-hand $BIH(i)_g$ in booking interval j by the

relationship $BIH(0)_g = BIH(j)_g / upbook_j$. Correlation coefficients $CC_g^2(j)$ for all paired intervals therefore decline, further reducing the weight on truncated observations relative to those with complete booking histories. Since high demand situations have more closure by definition, revenue losses in Table 6.16 will increase with demand level.

		Passenger Loads, Airline A With Efficient					*Passenger Loads, Airline B With Pickup*				
pbscl	zf_2	Y	B	M	Q	Total	Y	B	M	Q	Total
0.6	1.0	24.2	35.5	23.1	3.0	85.7	19.4	40.3	24.9	1.0	85.6
	2.0	**23.0**	**46.2**	**14.6**	**1.8**	**85.7**	**19.4**	**50.6**	**15.2**	**0.7**	**85.8**
	4.0	20.9	56.8	8.7	1.0	87.4	18.7	60.1	7.4	0.7	86.9
1.0	1.0	15.6	20.0	29.0	28.7	93.3	17.9	27.3	33.6	9.7	88.6
	2.0	**13.1**	**22.0**	**27.7**	**30.9**	**93.7**	**17.8**	**30.9**	**30.9**	**9.1**	**88.6**
	4.0	9.5	21.4	23.0	39.9	93.8	16.3	31.9	24.9	16.0	89.1

Table 6.17. Efficient and Pickup Loads with Variable zf_2, pbscl = 0.6 or 1.0; DF = 1.2

I examine the load effect of variable zf_2 assuming booking curves with *pbscl = 1.0* and *0.6*. Only the high demand case (DF = 1.2) is considered, since differences are most apparent at this level. As is shown in Table 6.17, *without scaling* the efficient forecaster has an inferior load distribution even with minimal booking curve variation at $zf_2 = 1.0$. It has fewer passengers in all high-value fare classes, which its excess of Q class passengers does not offset. As zf_2 increases, the fare class distribution of both airlines declines. Airline B with pickup is able to increase its Q and B class loads (lined boxes) at the expense of M class (shaded boxes); Airline A, however, loses Y and M class passengers, balancing this with increases only in Q class. At high booking curve variability and without scaling, the efficient forecaster's fare class distribution inferiority is intensified. This is consistent with the theorized weight reduction on HDB high demand observations, causing the seat optimizer to *underprotect* for high-valued fare classes.

In contrast, Table 6.17 indicates that when booking curves are scaled, the efficient forecaster has a superior load distribution (distinctly more Y class bookings) regardless of variation in the booking curve. This results in superior revenue results for efficient forecasting (shaded boxes in Table 6.16) at high demand levels. A curious inconsistency is created: Table 6.16 shows that at extreme scaling (*pbscl = 0.6*), increasing the demand factor removes the revenue superiority of the pickup forecaster, while under moderate or no scaling the superiority is *intensified.*

This revenue shift could be attributed to diminishing returns to scaling. But other facts -- especially the superiority of Y bookings by Airline A (Table 6.17) -- cannot be explained by this reasoning. I argue that two factors interact to favor the efficient forecaster under high DF and scaling conditions. First, Airline B (with pickup forecasting) has substantially limited sell-up possibilities as already discussed. Second, scaling booking curves for detruncation purposes has a disproportionate effect on the efficient forecaster. This may be seen by examination of equation (5.12), which gives the final estimate of unconstrained bookings on the forecast flight *f*. The weight *WP* is assigned to estimated bookings $\widetilde{BIH}(0)_f$ for the forecast flight *f*, where $\widetilde{BIH}(0)_f$ is derived using only advance bookings information $BIH(t)_f$, and booking curve detruncation. The remaining weight *1 - WP* is assigned to a weighted average of final estimated bookings on HDB flights. It is clear from equation (5.11) that the weight *WP* is relatively high, since $WP \geq CC_f^2(t)$. That is, bookings on the flight *f* being forecast are always weighed at least as much as the correlation coefficient between bookings at the forecast interval *t* and final bookings.

The influence of booking curve scaling on pickup forecasts is limited by the number of closed observations in the forecast HDB; each HDB observation is weighted equally. But $\widetilde{BIH}(0)_f$ exerts significant influence on the efficient forecast. If the forecast interval *t* is not too early and there is moderate booking curve variability, $CC_f^2(t)$ is typically above 0.5. This influence increases as zf_2 decreases, since lower variability around the booking curve increases $CC_f^2(t)$ and thereby *WP*. Hence the increasing revenue superiority of the efficient forecaster as zf_2 decreases at high DF (Table 6.16): If demands are high, detruncated $\widetilde{BIH}(0)_f$ and therefore the ultimate efficient forecast are likely to be high. This is not true under moderate demands (DF = 0.9), where $\widetilde{BIH}(0)_f$ will only be high if the flight *f* being forecast exhibits atypically high bookings-in-hand at the forecast interval *t*. In this case, efficient forecasting's sell-up opportunities are limited, and it can at best attain revenue *equality* with pickup forecasting at low zf_2. We may conclude the following:

- Increasing booking curve variability always makes pickup perform better relative to the efficient forecaster.

- Elevated demand levels increase the revenue differences at no or moderate scaling.
- Under extreme scaling, increasing demand reverses the revenue trend: Pickup forecasting is inferior under high demands.
- This is probably due to high weight placed on detruncated observations at high demand factors and low booking curve variability.

6.1.2.3. High/Low System K-factor Variability Scenarios

Like the regression forecaster, the efficient model should not be significantly affected by the variability of total demand between repeated instances of the same flight. This is contrasted with pickup forecasting, which (as explained in Section 6.1.1.3) will be ***negatively*** impacted because it uses average ***absolute*** bookings increases on HDB flights. Table 6.18, giving revenue differences for variable system demand k-factor *skf* at DF = 0.9, generally confirms this hypothesis: The superiority of the pickup forecaster declines as *skf* increases. In keeping with the regression versus pickup comparison, variable *skf* under DF = 1.2 is not considered here.

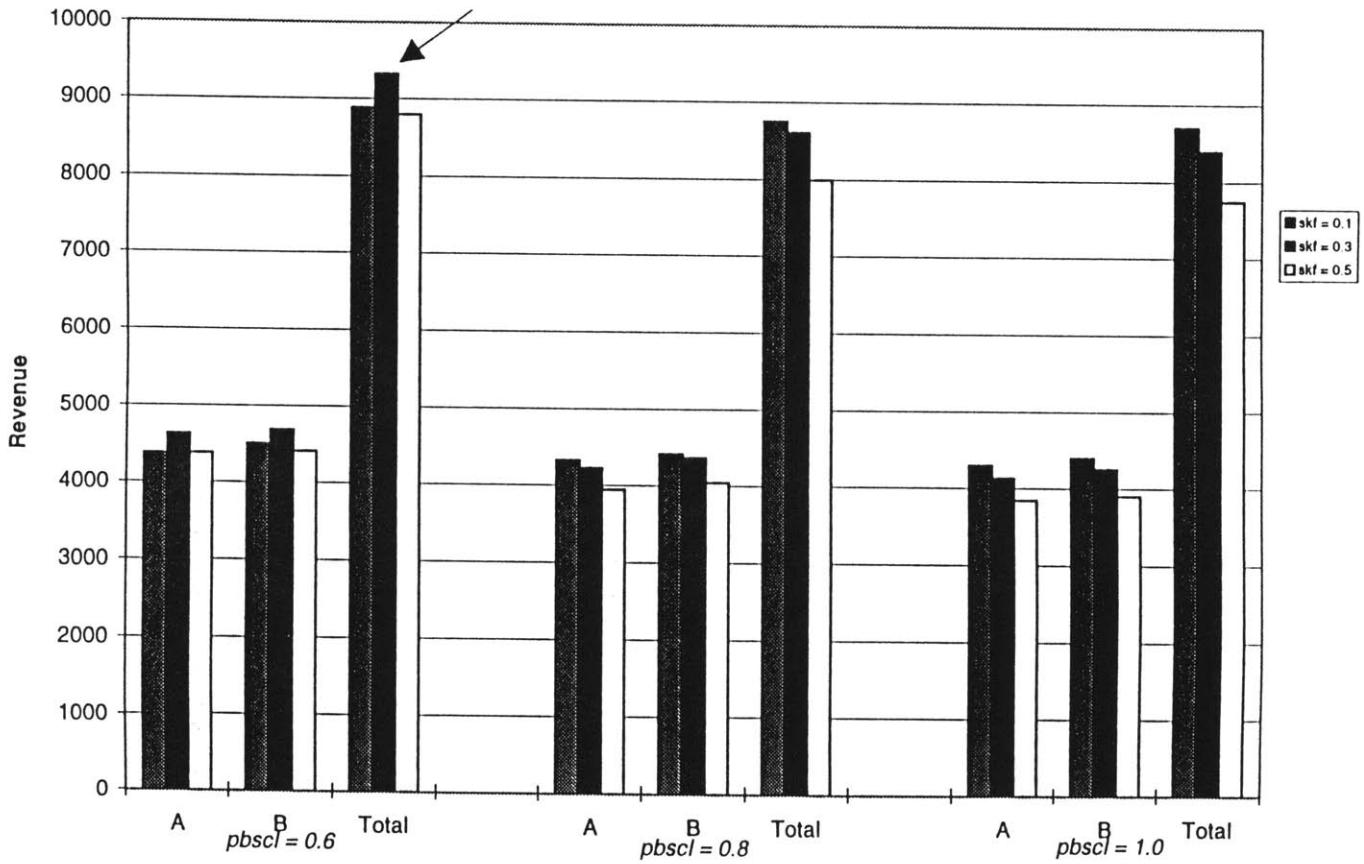
	% Rev Diff between Airline with Pickup over Airline with Efficient		
skf	Bk Crv (0.6)	Bk Crv (0.8)	Bk Crv (1.0)
0.1	2.99%*	2.05%*	2.25%*
0.3	1.25%*	3.35%*	3.02%*
0.5	0.75%*	2.58%*	1.53%*

Table 6.18. Pickup versus Efficient Forecasting Relative Revenue Performance under Variable skf, DF = 0.9

There are two exceptions to the general decline in pickup forecasting revenues with increasing *skf*: The efficient forecaster manages to make revenue inroads at low *skf* when moderate or no scaling is applied (shaded boxes in Table 6.18). I investigate this issue by comparing loads for both forecasters under booking curve detruncation with *pbscl = 0.6* (which exhibits the expected relationships) and *1.0* (which does not) in Table 6.19. It is difficult to generalize from load results. When both airlines scale, substantial gains occur in fare class distribution: Both airlines gain high-value class bookings (lined boxes) at the expense of lower-valued classes. Airline A with the efficient forecaster differentially gains, allowing it to almost eliminate its revenue losses relative to pickup forecasting.

Figure 6.11. Airline and Total Revenues under Variable *skf*.

Airline A has efficient and Airline B pickup forecasting; DF = 0.9



In contrast, without scaling Q class loads still decrease, but there is no significant increase in higher-class bookings. The revenue effect of these different load changes with increasing *skf* is shown in Figure 6.11, which graphs absolute revenue results for each airline and the total market, given a scaling level and *skf*. It is clear from the graph that for *pbscl = 0.8* or *1.0*, increasing *skf* causes revenue declines for both airlines and the total market. However, with extreme scaling, dramatically improved loads when *skf* increases from 0.1 to 0.3 result in *improved* revenues for both airlines and the total market (see arrow in Figure 6.11)! This result is contrary to our expectations: Unless the forecaster and/or detruncator *increases* forecast demands as variability increases, or the seat optimizer increases protections in the face of demand variability, increasing uncertainty should reduce average revenues.

		Passenger Loads, Airline A With Efficient					*Passenger Loads, Airline B With Pickup*				
pbscl	*skf*	Y	B	M	Q	Total	Y	B	M	Q	Total
0.6	0.1	9.3	5.3	4.4	70.2	89.2	13.0	7.3	14.2	48.0	82.4
	0.3	**12.4**	**11.2**	**14.8**	**43.8**	**82.3**	**12.2**	**13.2**	**27.9**	**25.5**	**78.8**
	0.5	11.6	12.9	17.0	33.7	75.1	10.2	15.5	25.5	22.2	73.3
1.0	0.1	9.6	4.9	4.6	68.1	87.2	11.7	5.9	8.4	58.7	84.6
	0.3	**9.5**	**6.7**	**3.9**	**61.4**	**81.5**	**10.8**	**8.0**	**7.5**	**54.3**	**80.6**
	0.5	8.5	7.3	3.7	55.6	75.1	8.7	8.4	5.3	52.7	75.0

Table 6.19. Efficient and Pickup Loads with Variable skf, pbscl = 0.6 or 1.0; DF = 0.9

I will discuss in Section 6.2.3 why seat optimizers usually *decrease* protections as *skf* increases. As was mentioned in Section 6.1.1.3 (and further justified in Section 6.2.3), an *increased* number of closed observations should occur as demand variability increases, simply because a greater proportion of HDB observations have outlying final unconstrained booking demands[150]. This assertion, combined with the efficient forecasting property of placing lower weight on truncated HDB observations, suggests that the forecaster's *downward bias* is magnified as variability increases. Lower high-value class demands are forecast, causing the seat optimizer to *underprotect* for higher-value passengers. Fewer high-value passengers arrive (since their fare classes are more often closed), and revenues decline. Table 6.19 and Figure 6.11 are obviously

[150] It is implicit in this argument that the booking limit for the involved fare class(es) do not vary. But I have just noted the contrary, i.e., booking limits on lower-valued classes *increase* with increasing demand variability. Assuming that variability in booking demand is not correlated among fare classes, this implies that seat availability on higher-valued classes tends to decline, on average. This consideration strengthens the argument which follows.

contrary to this conclusion. Apparently some as yet unknown feature of booking curve scaling operates to increase high-value fare class protections as variability in demand increases. Therefore, under increasing systemwide demand variability:

- The efficient forecaster improves relative to pickup.
- Total loads decline, and the fare class distribution deteriorates (assuming no scaling of booking curves used for detruncation).
- If booking curves are scaled, loads still decline but the fare class distribution inexplicably *improves*.

6.2. Detruncation Model Comparisons[151]

The examination of alternative forecasters in Section 6.1 also included much discussion of the various detruncation methods tested in PODS. Significant theoretical groundwork for the revenue differences I will explore in this section has therefore already been presented. Some of the apparent differences between detruncators have already been mentioned, through as yet different detruncation schemes have not been compared within the *same* PODS case. This section explores differences among detruncation methods, making reference to Section 6.1 (and the discussion of methods in Section 5.3) when appropriate to avoid repetition of theoretical details. I compare detruncation models under the *base case* scenario described in Section 5.1. The two airlines generally differ only in choice of detruncation method, and pickup forecasting is used for all comparisons. All comparisons are between detruncation methods relative to the base case booking curve detruncation without scaling, unless specifically noted otherwise.

6.2.1. Base Case and High/Low Demand Factor Scenarios

Table 6.20 presents comparisons for three demand factors between booking curve detruncation without scaling and other detruncation alternatives. Generally, there are few significant differences between detruncation methods at low demand factors. Differences emerge only at moderate and high demand factors. I will discuss each detruncation method in turn.

[151] Many results in this section are adapted from Skwarek (1996a). Changes have been made where tests were rerun on a newer version of PODS; no significant changes in relative or absolute relationships have been noted over these cases.

	% Rev Diff between Airline with Booking Curve (No Scaling) Detruncation over Airline with Selected Detruncation Method				
DF	None	Bk Crv (0.6)	Bk Crv (0.8)	Projctn	Pickup
0.7	-0.29%	0.20%	-0.23%	-0.41%*	0.00%
0.9	3.62%*	-2.09%*	-2.04%*	-0.02%	2.56%*
1.2	50.29%*	-1.10%*	-1.73%*	-2.63%*	42.01%*

Table 6.20. Booking Curve (No Scaling) Versus Other Detruncators: Relative Revenue Performance at Several Demand Factors

6.2.1.1. Booking Curve (No Scaling)/No Detruncation Comparison

It is clear from Table 6.20 that not detruncating closed observations has serious consequences for an airline. Without detruncation, HDB observations which close are ignored by the forecaster (Section 5.3.1). Since these flights are precisely those which are more likely to close, dramatic underprotection results. Table 6.21 details passenger loads for Airline A (with booking curve detruncation, no scaling) and Airline B (without detruncation). At low demand factors, there is no appreciable difference between the carriers. At moderate demand, both airlines experience substantial increases in loads. Airline A is able to direct most of this increase to higher-valued fare classes, while increases in Airline B loads are nearly all in the low-value Q class. This is consistent with the result shown in Figure 6.5, where booking curve detruncation causes higher estimates of unconstrained demands among HDB flights.

	Passenger Loads, Airline A With Bk Cv (1.0)					*Passenger Loads, Airline B With No Detrunc*				
DF	Y	B	M	Q	Total	Y	B	M	Q	Total
0.7	9.0	4.3	2.6	51.4	67.3	9.1	4.3	2.7	51.3	67.4
0.9	10.7	7.7	6.5	55.8	80.6	9.4	5.5	3.2	63.7	81.9
1.2	25.0	48.4	10.1	1.1	84.6	6.1	7.1	6.0	78.1	97.3

Table 6.21. Loads under Variable DF between Booking Curve (No Scaling) and No Detruncation

Under high demand conditions, the contrast is stark: booking curve detruncation almost completely closes down Q class (shaded box in Table 6.21) and strongly limits bookings in M class, causing the fare mix to shift heavily toward Y and B classes. Average loads increase at the high demand factor, yet on average 15 of 100 seats go empty despite unconstrained average demands of 120 passengers per airline (assuming passengers split evenly between the two). Airline B operates consistently full flights, but nearly all its passengers are low-fare Q class

passengers (see lined boxes). Indeed, its Y class bookings *decline* as the demand factor increases!

This is certainty due to underprotection for high-valued classes and thus a lack of seat availability when high-value passengers arrive. Consider Table 6.22, which yields closure rates (CR) by class for each of the scenarios in Table 6.21. The closure rate by fare class is an output of a standard PODS case. It is defined to be the proportion of observations in the case for which the fare class closes over the portion of the booking curve during which it is available[152]. CR measures the proportion of observations for which seat availability *SA* reaches *0* at some time during the booking process for the given fare class.

Closure statistics increase for one or a combination of three reasons. First, the fare class involved may experience high bookings, so booking limits *BL* become effective. Second, bookings in other fare classes may decrease seat availability *SA* for the fare class. Third, the seat optimizer may adjust *BL* downward if more bookings are expected in a higher-value fare class. Which combination accounts for a given CR must be considered in tandem with load information.

	CR, Airline A With Booking Curves (1.0)				*CR, Airline B With No Detruncation*			
DF	Y	B	M	Q	Y	B	M	Q
0.7	0.13	0.08	0.11	0.07	0.13	0.08	0.10	0.08
0.9	0.35	0.26	0.36	0.52	0.40	0.33	0.38	0.35
1.2	0.57	0.48	0.98	1.00	0.89	0.86	0.89	0.89

Table 6.22. Closure Rates under Variable DF between Bk Crve (No Scaling) and No Detruncation

Table 6.22 indicates that as demand increases, Airline B closes off higher-value fare classes *without* commensurably high bookings, implying *SA* = *0* induced closures as bookings in lower-value fare classes fill up seats on the airplane. In contrast, Airline A significantly limits the increase in high-value class CR via adjustment of *BL* on lower-valued fare classes (see shaded boxes). The effect of this action is seen clearly, as Q-class CR for Airline A approaches 1.00 with few bookings. Airline B's Q-class CR is also elevated, but because of very high bookings in that class as price sensitive passengers denied by Airline A find seats on Airline B. These results are all consistent with results in Section 6.1, which provides theoretical reasons why adoption of *any* detruncation method increases final forecasts of bookings, thus inducing additional protection.

[152] Thus, comparison of CR between fare classes is not appropriate if the fare classes are not defined to be available over the same number of booking intervals.

The dramatically poorer performance of Airline B begs the question of whether seat inventory control without detruncation is better than no yield management system at all. Wilson (1995) established that EMSR seat inventory control was always superior to FCFS or "First Come First Served" control, where no booking limits are set but restrictions built into the fare structure allow limited differential pricing. However, these comparisons assumed the use of booking curve detruncation by the EMSR airline throughout. How much of Wilson's demonstrated revenue improvement is truly attributable to the selected seat inventory control method alone, and how much is due to the added effect of detruncation?

To address this issue, a special three airline case was constructed. One airline has FCFS, one has EMSRb without detruncation, and the third has EMSRb with booking curve detruncation and no scaling. All other base conditions (e.g., DF = 0.9) are retained. Comparison of these three combinations *within* one case is superior to two pairwise comparisons, since stochastic elements vary considerably among successive runs of even the same case[153]. Table 6.23 indicates revenues and seat distributions for this case. The airline with the EMSRb seat optimizer alone yields a 6.02% revenue benefit over FCFS; with booking curve detruncation, a 9.25% improvement is realized (the latter result is substantially similar to results discussed by Wilson[154]). Thus, in this limited scenario, seat optimizing *even without detruncation* is a significant improvement over using only a fare structure to differentiate fares. EMSRb alone accounts for approximately 65% of the previously reported differences between FCFS and full seat inventory control. The *remaining 35%* is due to adoption of booking curve detruncation.

		% Over	*Passenger Loads*				
Control Method	*Revenues*	*FCFS*	Y	B	M	Q	Total
FCFS	$3,890	--	7.1	3.7	2.3	69.4	82.4
EMSRb (No Detrunc)	$4,124	6.02%	9.9	5.5	3.0	63.8	82.1
EMSRb (Bk Crv [1.0])	$4,250	9.25%	11.6	7.8	5.2	55.2	79.8

Table 6.23. Three-Airline Variable Control Method Revenue Results

The results of Tables 6.23 and 6.20 illustrate the importance of detruncation in determining the benefit airlines may expect to achieve with revenue management. An airline

[153] Pairwise comparisons were calculated and yielded results not significantly different than the three-airline single comparison reported here.
[154] Wilson 1995, pp. 91-92.

which invests heavily in a seat inventory control method *without* carefully considering input methodologies (including detruncation) substantially limits its revenue gain. From this section we may conclude:

- Not detruncating yields substantial revenue losses, especially at high DF.
- This is due to underprotection for high-valued fare classes.
- Detruncation accounts for a significant proportion of the gains due to the implementation of a yield management seat optimizer.
- The bulk of the revenue gain, however, is due to the adoption of the optimizer, not the detruncation method.

6.2.1.2. Booking Curve Detruncation With and Without Scaling

The results of Table 6.20 suggest that not scaling the booking curve used for detruncation has a deleterious revenue effect. This is consistent with the discussion in Section 6.1.1.1 on booking curve detruncation, where (given a forecaster) scaling the booking curve results in higher forecasts of unconstrained demand relative to not scaling. Load results for Airline A (which does not scale) and Airline B (which does) for DF = 0.9 and 1.2 are shown in Table 6.24 below. The low demand factor case has been suppressed due to lack of notable revenue differences.

		Passenger Loads, Airline A (No Scaling)					*Passenger Loads, Airline B (Scaling)*				
DF	*B pbscl*	Y	B	M	Q	Total	Y	B	M	Q	Total
0.9	1.0	10.2	7.5	6.5	56.8	81.0	10.3	7.5	6.1	57.2	81.0
	0.8	9.9	7.6	6.6	57.8	81.8	10.9	9.1	10.4	49.9	80.2
	0.6	9.9	8.8	9.6	55.6	84.0	12.2	11.6	20.3	33.4	77.5
1.2	1.0	16.3	29.1	36.7	8.9	91.0	16.3	29.1	36.4	9.3	91.0
	0.8	16.3	31.3	37.4	7.5	92.4	18.5	37.2	29.7	2.4	87.8
	0.6	16.4	35.4	37.7	4.6	94.1	21.1	48.1	12.7	0.6	82.5

Table 6.24. Loads under Variable DF and Scaling Levels for Booking Curve Detruncation

Both airlines have equivalent loads when Airline B does not scale (*pbscl = 1.0*), at both demand factors. However, as Airline B scales, it begins to protect *more* than Airline A. At DF = 0.9, Airline B quickly drops Q class bookings, increasing class M and B bookings significantly. By contrast, Airline A has no significant movement in B class loads, and only modest increases in M loads at the expense of Q bookings. The effects are more extreme with high demands, as Airline B completely closes Q class and significantly limits M class bookings, yielding increases in

Y and B class bookings. Airline A also experiences declines in Q class bookings, with increased B class loads.

Why should Airline A experience this spike in M class loads when *Airline B* adopts more scaling? Since it does not respond by adopting a similar *pbscl,* Airline A should at best experience no effect due to Airline B's scaling, and at worse *lose* higher-class bookings to Airline B as its relative underprotection increases. In the typical situation with asymmetry in relative protection for high-valued fare classes, the carrier which underprotects presumably receives most of the low-value passengers; only the airline with greater protection gains from its actions.

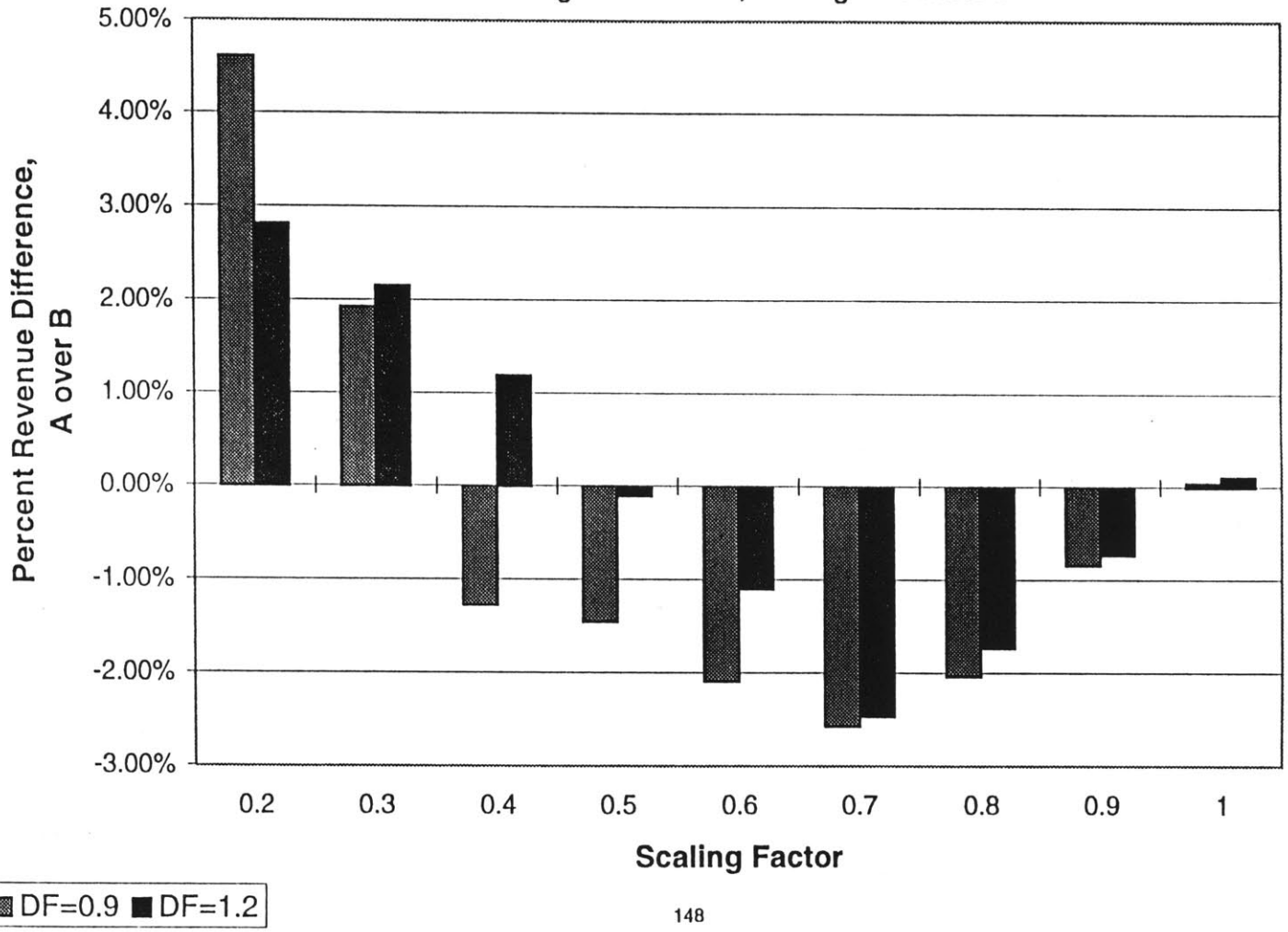
This assertion is *incorrect,* for the simple reason that the competitive nature of airline markets may also work *against* the scaling airline. As it continues to scale, the ever-higher predictions of unconstrained demand result in tighter limits on low-value class bookings. This effort usually nets yet more high-value passengers, since sell-up is induced as low-value availability is constricted. However, the probability that those denied space in the lower-valued classes will refuse to sell-up (and seek accommodation on the airline with relatively less stringent protection) increases commensurably. The observed increases in Table 6.24 suggest that increased protections are beginning to have this effect in M class and perhaps B classes. Such a situation indicates that scaling may at some point have the unintended effect of yielding more benefit to the airline *not* adopting sell-up than the adopting carrier.

This issue is addressed further in Figure 6.12, which shows the revenue difference of Airline A (which does not scale) over Airline B (which does) for *pbscl* from 0.2 to 1.0, under DF = 0.9 and 1.2. Clearly, Airline A earns less than Airline B when the latter chooses *pbscl* between *0.5* to *1.0*. However, scaling choices below this level confer more revenues to Airline A as passengers unwilling to sell up to the extent required by Airline B flock to it. Airline A's gain is not singly restricted to Q-class bookings. Gains occur in all but the highest-value classes as its yield management system induces sell-up in recognition of high-demand conditions for its capacity[155]. If Airline B desires to maximize its revenue superiority over the competition, it should restrict *pbscl* to a relatively modest 0.7. Note that conditions deteriorate faster under high demand conditions. With high market demands, significant sell-up is *automatically* induced via

[155] Obviously, the sell-up is less extensive than that which Airline B would impose. Otherwise passengers have no incentive to switch to Airline A.

Fig. 6.12. Percent Rev. Difference with Variable Scaling, Airline A over B

No scaling for Airline A; scaling for Airline B



the *BL* adjustment process of the seat optimizer. Further increases via *pbscl* are therefore limited. This issue will be discussed further in Section 6.3 below. In summary,

- Not scaling under booking curve detruncation imposes revenue losses because higher-value classes are underprotected.
- If only one airline adopts scaling, the ability of the adopting carrier to induce higher revenues is limited. More revenues will be given to the non-adopting competitor at extreme *pbscl*, as passengers refuse to sell-up.

6.2.1.3. Booking Curve (No Scaling)/Projection Detruncation Comparison

Table 6.20 indicates no large differences between projection and booking curve detruncation without scaling, except at high demand factors. Under high demand conditions, projection detruncation becomes approximately revenue equivalent to booking curve detruncation with extreme scaling. How is it that projection detruncation is unremarkably better until high DF, when it then outperforms most alternative methods? I have discussed a possible answer in Section 6.1.1.1 above, i.e., the creation of extreme outliers with high demand. This is justified by equations (5.15) and (5.16), which derive the projection detruncation estimate of unconstrained "pickup" from the forecast interval t until departure on a constrained HDB observation. Here we desire to forecast unconstrained bookings on a particular forecast flight f which is currently at interval t in its booking history, and at least one of the HDB observations w closes in an interval *past t*. In equation (5.15) as the difference cl (bookings from forecast interval t until closure on w) less μ_t (mean "pickup" on unclosed HDB flights between t to flight departure) decreases, the projection estimate of "pickup" and thus final bookings $B\hat{I}H(0)_w$ on w dramatically increases -- especially for a low value of τ. Do high demand conditions cause cl to approach μ_t?

	Passenger Loads, Airline A With Bk Cv (1.0)					*Passenger Loads, Airline B With Projection*				
DF	Y	B	M	Q	Total	Y	B	M	Q	Total
0.7	9.0	4.3	2.6	51.4	67.3	9.2	4.3	2.7	51.2	67.4
0.9	10.3	7.3	5.9	57.4	80.9	10.6	6.8	4.5	59.6	81.4
1.2	16.4	32.7	37.3	6.9	93.4	20.7	41.8	20.2	3.7	86.4

Table 6.25. Loads under Variable DF between Bk Crve (No Scaling) and Projection Detruncation

Certainly. With moderate demand, μ_t is well below average booking limits at forecast interval t. Relatively low cl will induce closure, since the seat optimizer (expecting moderate

high-value bookings) allows low-value bookings to fill the plane. In contrast, high demands cause μ_t to increase, but this increase is limited by the fact that unclosed observations must have incremental bookings *below* the remaining capacity on the plane. At the same time, bookings *cl* which induce closure are significantly higher (since high demand situations cause the optimizer to protect more seats for higher-valued fare classes). Therefore, the difference μ_t - *cl* declines as demand factor increases. It is clear that projection detruncation will detruncate closed HDB observations to high (outlier) estimates of unconstrained bookings. This is demonstrated in Table 6.25, where both detruncation methods dramatically reduce Q-class bookings (see shaded boxes) and increase higher-valued fare class (lined boxes). Airline B with projection scores significantly higher gains in Y and B fare classes, while Airline A's booking curve detruncation increases M class loads.

6.2.1.4. Booking Curve (No Scaling)/Pickup Detruncation Comparison

The poor showing of pickup detruncation in Table 6.20 suggests that it is hardly better than no detruncation at all. Comparison of load results when Airline B has pickup (Table 6.26) versus no detruncation (Table 6.21) indicate that this is exactly the case. As DF increases, the increase in low-class bookings without any substantial increase in high-value class bookings is exactly similar. Curiously, pickup detruncation even yields less M class passengers than without detruncation at all, though this is compensated for by higher Y class bookings.

	Passenger Loads, Airline A With Bk Cv (1.0)					*Passenger Loads, Airline B With Pickup*				
DF	Y	B	M	Q	Total	Y	B	M	Q	Total
0.7	9.0	4.3	2.7	51.2	67.2	9.1	4.3	2.7	51.0	67.1
0.9	10.4	7.5	6.1	56.7	80.7	9.8	5.6	3.2	63.0	81.6
1.2	22.4	45.4	15.1	2.2	85.0	8.9	7.0	3.4	77.6	96.9

Table 6.26. Loads under Variable DF between Bk Crve (No Scaling) and Pickup Detruncation

The reason for pickup detruncation's remarkably substandard revenue performance has been explained in Section 6.1.1.1 and Figure 6.7. Assuming that the average increase in bookings for *unclosed* HDB observations would have been obtained on closed HDB flights completely ignores the fact that fare classes close precisely because they have high demand. High and low demand flights *do not* exhibit the same absolute booking characteristics!

6.2.1.5. Detruncation "Zero-Sum" Analysis

The comparisons in Table 6.23 between EMSRb with and without detruncation, and FCFS are possible because detruncation of HDB observations which are closed is *not* a requisite element of seat inventory control. Unlike forecasting, seat optimizers operate sufficiently (albeit suboptimally) without detruncation. This *optional* nature of detruncation permits a "zero-sum" analysis of detruncation. That is, since airlines operate in competitive contexts, do the gains to detruncation come simply at the expense of competitors? Is detruncation adoption *zero-sum*, so that total market revenues do not vary depending on whether one or both carriers obtain it?

Figures 6.13 and 6.14 give for DF = 0.9 and 1.2 the percentage revenue improvement for each airline and the total market over the case where neither detruncates, for each detruncation pair. According to the first two columns of Figure 6.13 for DF = 0.9, when Airline A adopts booking curve detruncation, it earns an approximate 3% revenue improvement assuming Airline B does not. Limited losses of less than 1% are imposed on Airline B, yielding a total market improvement of approximately 1.5%. If Airline B responds by introducing projection detruncation, it earns approximately the same as Airline A, or a 2% improvement. As I have discussed, Airline B does little better by adopting pickup detruncation than not detruncating at all. Finally, Airline B may achieve superior revenues by adopting booking curve detruncation and also scaling, which also improves Airline A's revenue results. Both airlines earn a 2% revenue improvement by both adopting booking curve detruncation. Therefore, detruncation is *not* zero-sum: Total market revenues increase in both the single and joint adoption cases, though the airline adopting first (Airline A in Figure 6.13) may lose some of its gain when the competition catches up. These events are repeated at high DF (Figure 6.14), though the proportionate improvements over not detruncating are *significantly* higher.

6.2.2. High/Low Booking Curve Variability Scenarios

The revenue performance measures in Table 6.20 were also calculated under low and high zf_2, giving the sensitivity of the compared detruncation methods to variation in bookings received around the booking curve. This process was repeated for moderate and high demand factors in

Figure 6.13. Revenue Change Over No Detruncation, DF = 0.9

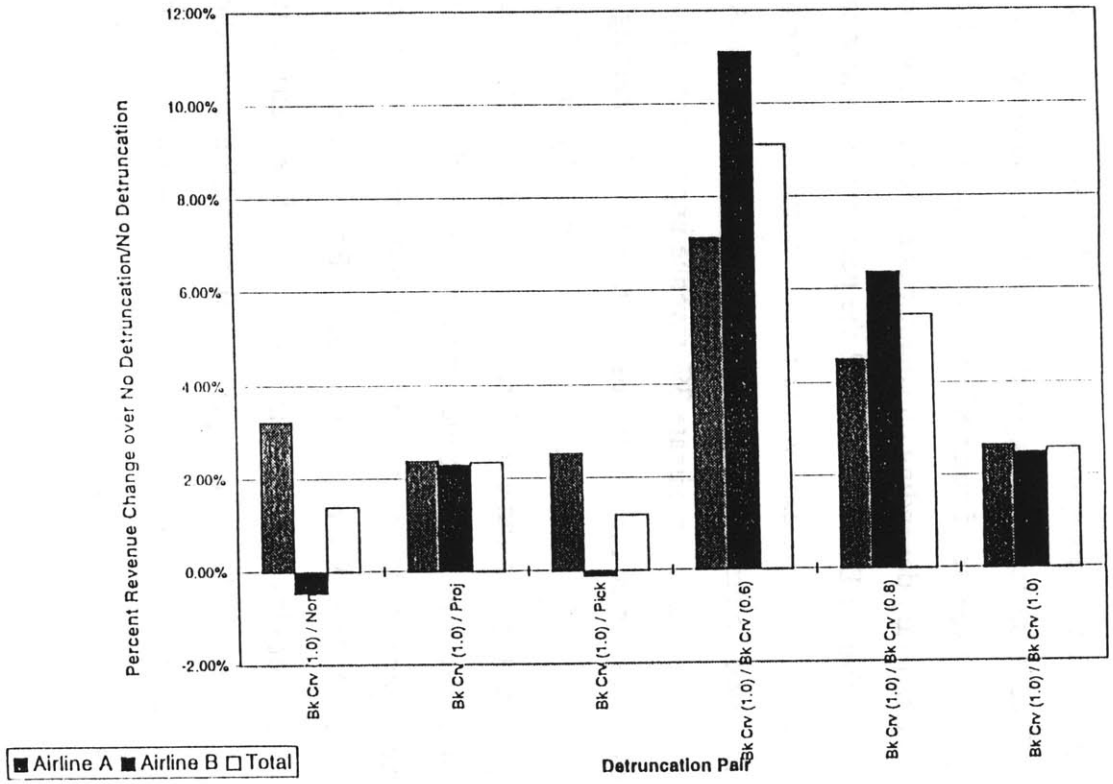


Figure 6.14. Revenue Change Over No Detruncation, DF = 1.2

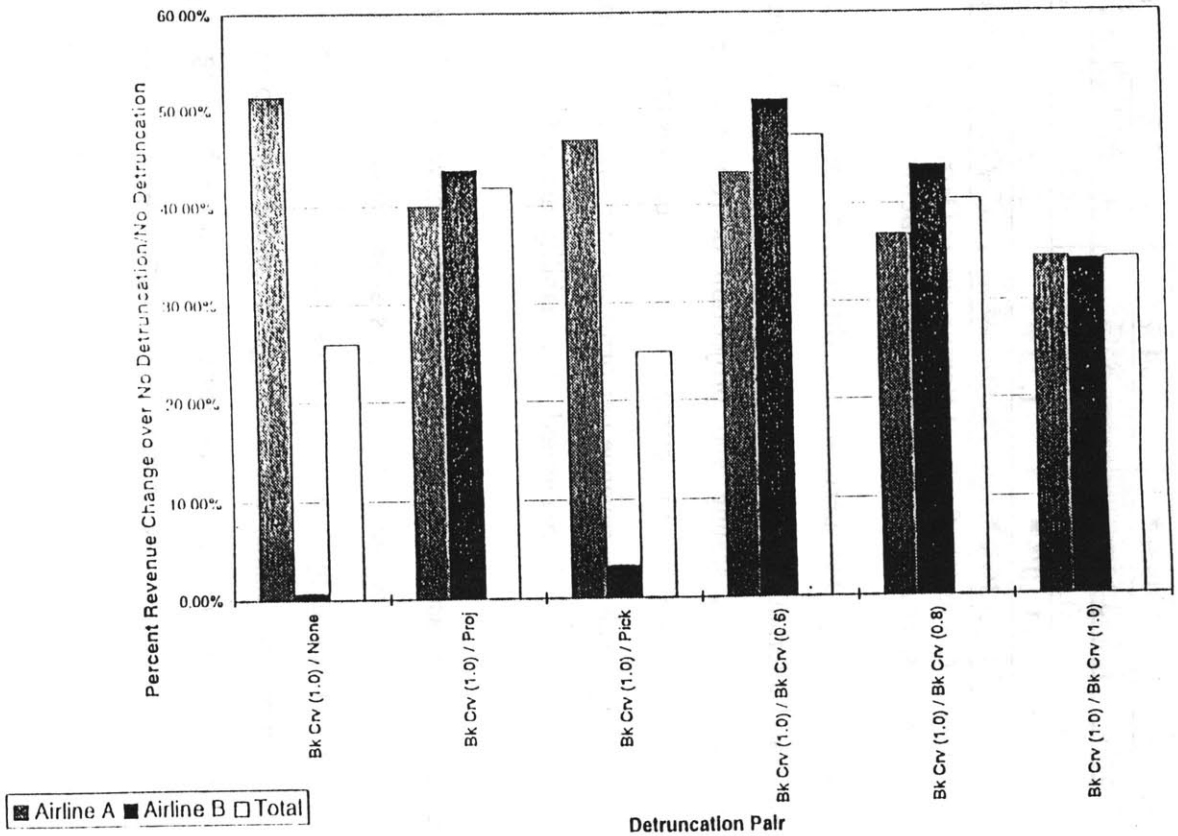


Table 6.27; DF = 0.7 has been excluded as demand levels are not sufficiently high to cause observable differences.

		% Rev Diff between Airline with Booking Curve (No Scaling) Detruncation over Airline with Selected Detruncation Method				
DF	zf_2	None	Bk Crv (0.6)	Bk Crv (0.8)	Projctn	Pickup
0.9	1.0	3.21%*	-0.59%*	-1.83%*	0.77%*	2.80%*
	2.0	**3.62%***	**-2.09%***	**-2.04%***	**-0.02%**	**2.56%***
	4.0	3.88%*	-3.58%*	-1.85%*	-0.33%*	3.27%*
1.2	1.0	46.22%*	0.19%	-0.59%*	-1.57%*	37.15%*
	2.0	**50.29%***	**-1.10%***	**-1.73%***	**-2.63%***	**42.01%***
	4.0	49.9%*	-5.10%*	-5.00%*	-5.77%*	46.19%*

Table 6.27. Booking Curve (No Scaling) Versus Other Detruncators: Relative Revenue Performance at Several Z-Factors

Table 6.27 indicates generally that performance differences between booking curve detruncation without scaling and alternatives vary in consistent directions as zf_2 increases. Thus, since booking curve detruncation with scaling and projection detruncation are superior at moderate zf_2, this superiority *increases* as booking curve variability increases (see shaded boxes). In contrast, the superior protection levels and revenue performance of booking curve detruncation without scaling over pickup detruncation is magnified as variability around the booking curve increases.

These facts suggest an outlier-based analysis of these results. I have established in Section 6.2.1 that revenue and load differences between detruncators arise because their generation of outliers and subsequent protection levels by fare class varies. Further, Section 5.3.2 discussed why flights with low early bookings and higher late bookings are more likely to close than those with typical proportional booking patterns, or high bookings *early* in the booking process. The effect of increasing zf_2 is to increase the number of flights which deviate from the $\frac{1}{upbook_t}$ proportional relationships. This was illustrated in Figure 5.1, where additional variation around the typical booking curve ***B*** increases the likelihood of *outlier* observations with high early bookings ***A*** and high late bookings ***C***.

Since closed observations are more likely to occur among flights with booking curves similar to ***C***, it follows that increasing zf_2 generates a higher proportion of closed observations in

the dataset. Thus, the revenue consequences of adopting an inferior detruncator increase: Its systematic underprotection will be applied to more flights. This situation is illustrated in Table 6.28, where loads under booking curve detruncation without scaling are compared with *pbscl* = *0.6* scaling at DF = 1.2. Even without significant booking curve variation (zf_2 = *1.0*), Airline A (without scaling) attains a different fare class distribution than Airline B with scaling. However, they are approximately revenue equivalent.

	Passenger Loads, Airline A With Bk Cv (1.0)					*Passenger Loads, Airline B With Bk Cv (0.6)*				
zf_2	Y	B	M	Q	Total	Y	B	M	Q	Total
1.0	17.3	30.0	40.0	5.8	93.1	20.6	38.5	23.6	0.8	83.5
2.0	16.4	38.4	37.7	4.6	94.1	21.1	48.1	12.7	0.6	82.5
4.0	13.4	37.8	39.8	4.9	95.9	21.8	55.4	5.4	0.4	83.0

Table 6.28. Loads under Variable zf_2 between Bk Crve (No Scaling) and Bk Crve (pbscl = 0.6)

As variation around the booking curve increases, Airline A experiences a significant decline in Y class bookings and a commensurate increase in B class loads, suggesting inferior protection for Y class due to underpredicted final demands. Other fare classes are not affected. In contrast, Airline B maintains approximately the same Y class bookings but sees a substantial transfer of M class bookings to the higher-value B class. In so doing Airline B gains a significant revenue advantage over its competitor. These results are consistent with the underprotection of inferior detruncators as the number of closed observations in the HDB increases. We may conclude that:

- Increasing variability around the booking curve leads to more closed observations.
- Inferior detruncators will cause greater revenue losses at high zf_2, since they are applied to an increased number of observations.

6.2.3. High/Low System K-factor Scenarios

Comparison of detruncation methods under variable *skf* allows the determination of how well each method responds to underlying stochastic variation in demand between repeated observations of a flight. Revenue results for each detruncation method relative to booking curve without scaling are shown, for *skf* = *0.1* through *0.5*, in Table 6.29. Low demand factors have again been excluded, since the absence of effective booking limits on the seat inventory

management system tends to eliminate revenue differences due to alternative methodologies. The high demand factor case also will not be discussed for brevity.

	% Rev Diff between Airline with Booking Curve (No Scaling) Detruncation over Airline with Selected Detruncation Method				
skf	None	Bk Crv (0.6)	Bk Crv (0.8)	Projctn	Pickup
0.1	2.18%*	0.22%	-0.27%*	0.41%*	1.53%*
0.3	**3.62%***	**-2.09%***	**-2.04%***	**-0.02%**	**2.56%***
0.5	1.73%*	-6.22%*	-3.40%*	-1.74%*	1.27%*

Table 6.29. Booking Curve (No Scaling) Versus Other Detruncators: Relative Revenue Performance at Several skf

Table 6.29 clearly indicates two dichotomous groups. When booking curve detruncation without scaling (the base case detruncator) is compared with pickup detruncation or no detruncation, an apparently inconsistent trend with increasing *skf* results: The revenue superiority of booking curve is maximized at moderate variability, *declining* as *skf* increases. The second group is composed of detruncation methods which clearly outperform unscaled booking curve detruncation, to an increasing degree as demand variability increases.

We may explain these results with Figure 6.15, which graphs the distribution of unconstrained demand for a moderate and high *skf* distribution, as represented by the probability density functions $M(\cdot)$ and $H(\cdot)$ respectively. Theoretically, varying *skf* has no effect on mean unconstrained bookings B_h, though the likelihood of receiving B_h decreases significantly, since $M(B_h) > H(B_h)$[156]. Now suppose a booking limit *cl* has been imposed on this fare class. For simplicity, *cl* is invariant over the booking period, between flights, and various *skf* values.

Then the *mean* unconstrained demand for observations which close increases with *skf*, so $\underline{cl}_H > \underline{cl}_M$. Therefore, detruncators which generate *higher* unconstrained estimates of demand for closed flights will perform better in a high *skf* environment. This is exactly what occurs for detruncators in the second group, and is consistent with the result of Section 6.1.1.1 that booking curve detruncation without scaling generally yields lower estimates of unconstrained demand than these detruncation methods. Figure 6.15 also indicates that the likelihood of closure increases

[156] Though Figure 6.15 is drawn and treated as a continuous distribution, the distribution of final bookings *BIH(0)* for a flight is obviously discrete.

with *skf*, as $\int_{x=cl}^{\infty} M(x)dx < \int_{x=cl}^{\infty} H(x)dx$. Since detruncation is applied over a greater proportion of the HDB, inferior detruncators should therefore perform *worse*. This is contrary to the results of the first group, where (inferior) pickup detruncation and no detruncation partially recover their revenue losses at high *skf*.

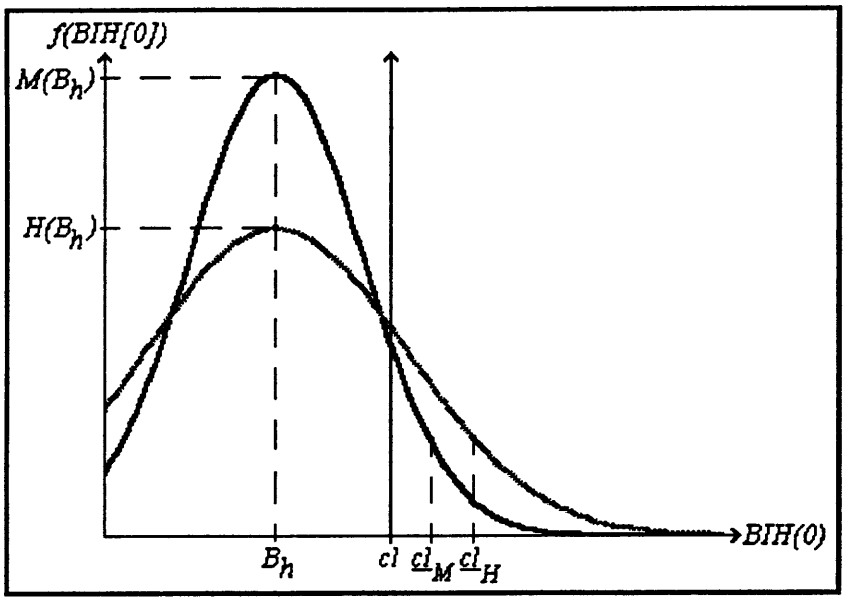


Figure 6.15. Probability Distribution of Bookings under Moderate and High skf

To gain insight into this incongruity, Table 6.30 compares loads when Airline A has booking curve detruncation without scaling, and Airline B utilizes no method. Contrary to Figure 6.15, both carriers lose significant bookings as system variability increases. Airline A loses significant Y and M class bookings, while Airline B's losses occur primarily in Q class (see shaded boxes). Airline A's superiority declines at high *skf* because it loses more high-value class bookings.

	Passenger Loads, Airline A With Bk Cv (1.0)					*Passenger Loads, Airline B With No Detrunc*				
skf	Y	B	M	Q	Total	Y	B	M	Q	Total
0.1	11.7	6.0	8.6	58.0	84.2	9.8	4.5	3.4	69.8	87.5
0.3	10.7	7.7	6.5	55.8	80.6	9.4	5.5	3.2	63.7	81.9
0.5	8.4	6.9	4.1	55.6	74.9	8.0	5.9	2.8	58.6	75.2

Table 6.30. Loads under Variable skf between Bk Crve (No Scaling) and No Detruncation, DF=0.9

The loss in total bookings and disproportionate losses in high-value bookings may be explained by a structural component of the seat optimizer which causes protection levels to *decline* as demand variability increases. Given a constant mean bookings level, many seat optimizers including EMSR protect fewer seats for high-valued fare class(es) as their variability increases, assuming that the ratio of fares between adjacent fare classes is greater than 0.5 (which is reasonable in present airline practice)[157]. Such a procedure is intuitively justified by the marginal protective logic of most seat optimizers. Recall from Section 3.3.2.2 that EMSRb jointly protects seats for higher-value fare classes until the *expected* marginal revenue from receiving the next passenger in one of these classes is equated with the expected marginal revenue from receiving the first passenger in the next lowest fare class[158]. Clearly, increasing the variation in demands for higher-value fare classes causes the probability of receiving n passengers to begin declining (from 1) at a lower value of n, yielding lower expected marginal revenues[159].

	CR, Airline A With Bk Cv (1.0)				*CR, Airline B With No Detrunc*			
skf	Y	B	M	Q	Y	B	M	Q
0.1	0.25	0.14	0.30	0.63	0.36	0.28	0.38	0.33
0.3	0.35	0.26	0.36	0.52	0.40	0.33	0.38	0.35
0.5	0.40	0.33	0.37	0.42	0.42	0.36	0.39	0.36

Table 6.31. Closure Rates under Variable skf, Bk Crve (No Scaling) and No Detruncation; DF=0.9

This is confirmed in Table 6.31: As *skf* increases, Airline A loses bookings in Y and M class while its closure rates *increase* (see lined boxes). Such an event occurs in one of two instances. The first is if the seat optimizer sets inadequately low protection levels for the (presumably high-value) fare class, causing it to close with fewer bookings. Second, the seat optimizer will reduce booking limits on a (presumably low-value) fare class if additional bookings are expected in higher fare classes. Since the second case is not relevant, we may conclude that the seat optimizer increasingly *underprotects* high-valued fare classes as *skf* increases. A similar effect occurs for Airline B, though somewhat diminished (lined boxes). This causes its revenue

[157] Belobaba (1987), pp. 153-157.

[158] Since this latter expectation is always approximately one if there are sufficiently few classes, the expected revenue from receiving the first passenger in the next lowest fare class is simply the fare for that class.

[159] However, this more rapid onset of probability decline is balanced by a less steep decline with higher values of n. This is the reason for the fare ratio > 0.5 condition. Belobaba (1987), p. 155.

"comeback." It is, however, ironic that in this case *underprotection* (which is usually associated with higher low-value class loads) is paired with *lower* total loads and lower loads in Q class.

When variability in total demand increases between observations, the following occurs:

- The mean unconstrained demand on closed observations increases. Therefore, detruncation methods with larger estimates of unconstrained bookings perform better.
- Most seat optimizers decrease protection levels. Total bookings tend to decline, while closure rates for high-valued fare classes increase.

The comparative discussion of detruncators is concluded. I will complete this chapter with results on attempts to induce sell-up by modifying seat protection levels.

6.3. Sell-Up Analysis[160]

Throughout the discussion in Sections 6.1 and 6.2, *revenue superiority* was consistently shown to be obtained by using forecasters and detruncators which yielded higher estimates of demand, thereby causing the seat optimizer to protect more seats for high valued classes. This resulted in the earlier closure of low-valued classes, causing some passengers who would otherwise have traveled in a lower fare class to *sell-up*, or buy a more expensive ticket than the passenger would have selected had the low fare class been available. The discussion also indicated that certain limitations to sell-up exist; passengers are increasingly less likely to buy up to progressively higher-valued fare classes as low-valued classes close. This section focuses on PODS tests of an explicit attempt to induce sell-up by adjusting booking limits in each class. It uses a method developed by Belobaba and Weatherford, and discussed in Section 3.3.2.2. The approach of Sections 6.1 and 6.2 (with a more theoretical explanation for revenue differences between input methodologies) is inappropriate for this section because we are not comparing among alternative methods with differing assumptions about demand processes. Here I will focus on a descriptive comparison of observed revenue effects at various assumed sell-up levels, and intuitive explanations for these effects.

Recall from the discussion of simulation context in Section 5.4 that tests of the EMSRb sell-up model are performed over a variety of sell-up rate (*SU*) estimates between all adjacent fare

[160] Many results in this section are adapted from Skwarek (1996b).

class pairs. Estimated *SU* rates are the assumed probability that a passenger able to meet the restrictions in his or her initially desired fare class (but denied) will buy a seat in the next most expensive fare class. Three sell-up scenarios are created: *no* adjustment (i.e., *SU = 0.0), single* adjustment, and *joint* adjustment. In these experiments, Airline B always adjusts for estimated *SU*, while Airline A does or does not depending on the scenario. In the *joint* adjustment case, both carriers adopt the same estimated *SU*, implying revenue and load symmetry. Realizing that the competitive context of PODS ensures that an airline's action has potentially significant effects on the competitor, choosing estimated *SU* to maximize individual gains may possibly lead to important competitor gains as well. I therefore consider *three* possible objectives for Airline B, the airline always adjusting for sell-up:

- **Objective I**: Maximize *difference* between adjusting and non-adjusting airline revenues. This is represented in the graphs which follow by a thin solid arrow over the estimated *SU* rate where the objective is achieved.
- **Objective II:** Maximize *improvement* in adjusting airline's revenues over *SU = 0.0* when only it adjusts for *SU*. This is the myopic individual maximization perspective, and is represented by a thick dashed arrow. Realistically, it is also the most likely to be considered by airlines.
- **Objective III:** Maximize *improvement* in an airline's revenues over *SU = 0.0* when *both* carriers adjusts for *SU*. This is the appropriate goal for a collusive combination (if it were possible, and legal) to maximize joint revenues; it is represented by a thick solid arrow. While I treat this as a collusive case, collusion *need not* be occurring if both carriers have adopted *SU*. Joint adjustment and common *SU* estimates may easily result from individual profit-maximizing behavior.

Further, two *critical points* for revenue results occur because of the competitive interaction of the two airlines. These are:

- *CR1*: Revenue benefit to adjusting airline equals benefit to non-adjusting airline. Airline B (the adjusting airline) must choose an estimated *SU* lower than *CR1*, otherwise one benefits the competitor more than oneself.
- *CR2:* Revenue benefit to non-adjusting airline equals individual benefit with joint adjustment. The colluding airlines must choose an estimated *SU* lower than or equal to *CR2* if the objective is to maximize joint revenues. Otherwise, it is individually optimal for an airline *not* to incorporate *SU* estimates assuming the other does, and the collusion fails.

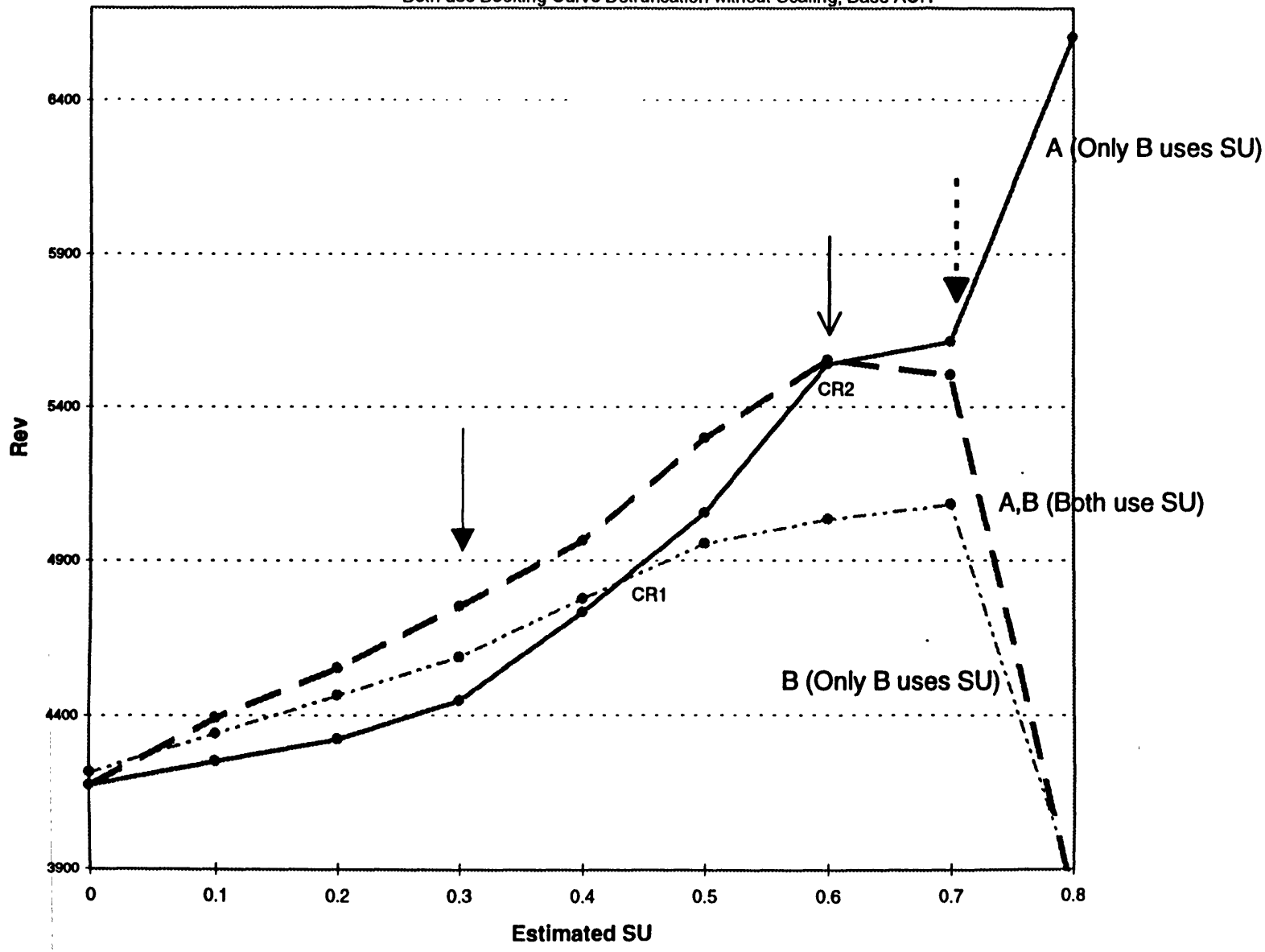
It is important to emphasize that the actual success of a given estimated *SU* rate depends on the actual (unknown) willingness of passengers to sell-up, and the degree to which the present fare structure is imperfect. First, the results of this thesis include underlying assumptions about passenger willingness to sell-up. Recall from Section 4.5.2 that within the PODS framework a passenger eliminates paths and fare classes whose absolute fare is greater than the passenger's maximum willingness to pay (MWTP). This MWTP is derived from a multiple of the lowest fare in the market (set by passenger type), and is stochastically given by the acceptable cost ratio or ACR (Section 5.1.3). MWTP screening is the primary mechanism by which passengers exclude expensive fare products, though subsequent choosing among qualifying paths naturally selects the least expensive option when monetary costs of all travel attributes have been affixed to each path and fare class option (see Section 4.5.2). Base case ACR is 5.0 for nondiscretionary and 2.0 for discretionary travelers. Thus, the former will always take any fare class, while the latter will sell-up to all but Y class, on average. The second factor influencing the performance of adjusting for estimated *SU* is the degree to which the present fare structure *imperfectly* segments passengers by passenger type. If there are no misidentified passengers (i.e., all purchase tickets in the most expensive fare class they are willing to accept when $SU = 0.0$), estimating any $SU > 0.0$ will cause losses for the adjusting airline(s). Results in this thesis are specific to the environment assumed by PODS.

6.3.1. Base Case and High Demand Factor Scenarios

Figure 6.16 shows revenue results for estimated *SU* rates between 0.0 and 0.8 for three cases: Airline A's individual revenue when only Airline B adjusts (thick solid line), Airline B's individual revenue when only it adjusts (thin dashed line), and revenues per carrier when *both* adjust (thick dashed line). Clearly, revenues experience gradual improvement until an extreme estimated *SU* probability of 0.7 used in the model, in which case revenues for the adopting airline(s) declines precipitously. When Airline A does not adjust for estimated sell-up, it nonetheless gains substantially from Airline B's actions. Since $CR1 \approx 0.45$, Airline B is clearly limited in its ability to induce sell-up. Further overprotection causes passengers to move *en masse* to Airline A, giving it far superior revenues. A relatively high $CR2 \approx 0.6$ indicates that airlines

Figure 6.16. Revenues with Increasing SU, DF=0.9

Both use Booking Curve Detruncation without Scaling; Base ACR



adjusting for estimated sell-up have a significant range of estimated *SU* rates where *joint* profitability is possible.

A "jointness" effect also prevents Airline B from gaining all the revenue benefit of sell-up unless Airline A cooperates and also adjusts for estimated *SU*. Joint revenues are significantly higher than revenues when only Airline B adjusts over the entire range of *SU*. This is consistent with the hypothesis that sell-up depends critically on the *availability of alternatives* as a prominent force governing passenger sell-up likelihood. If both airlines adjust for *SU*, passengers may no longer escape to a low-priced alternative, and so must pay more consumer surplus to the colluding airlines. Objective performance expressed in percentage terms in Table 6.32 indicates that maximal revenue improvements are achieved only with joint adjustment (Obj. III), yielding a 32.98% revenue improvement for each adjusting carrier. Boxes outlined in Table 6.32 are the three *direct* objectives of the adjusting airline(s), as outlined above. Other percentage changes are calculated for analysis of competitive impacts.

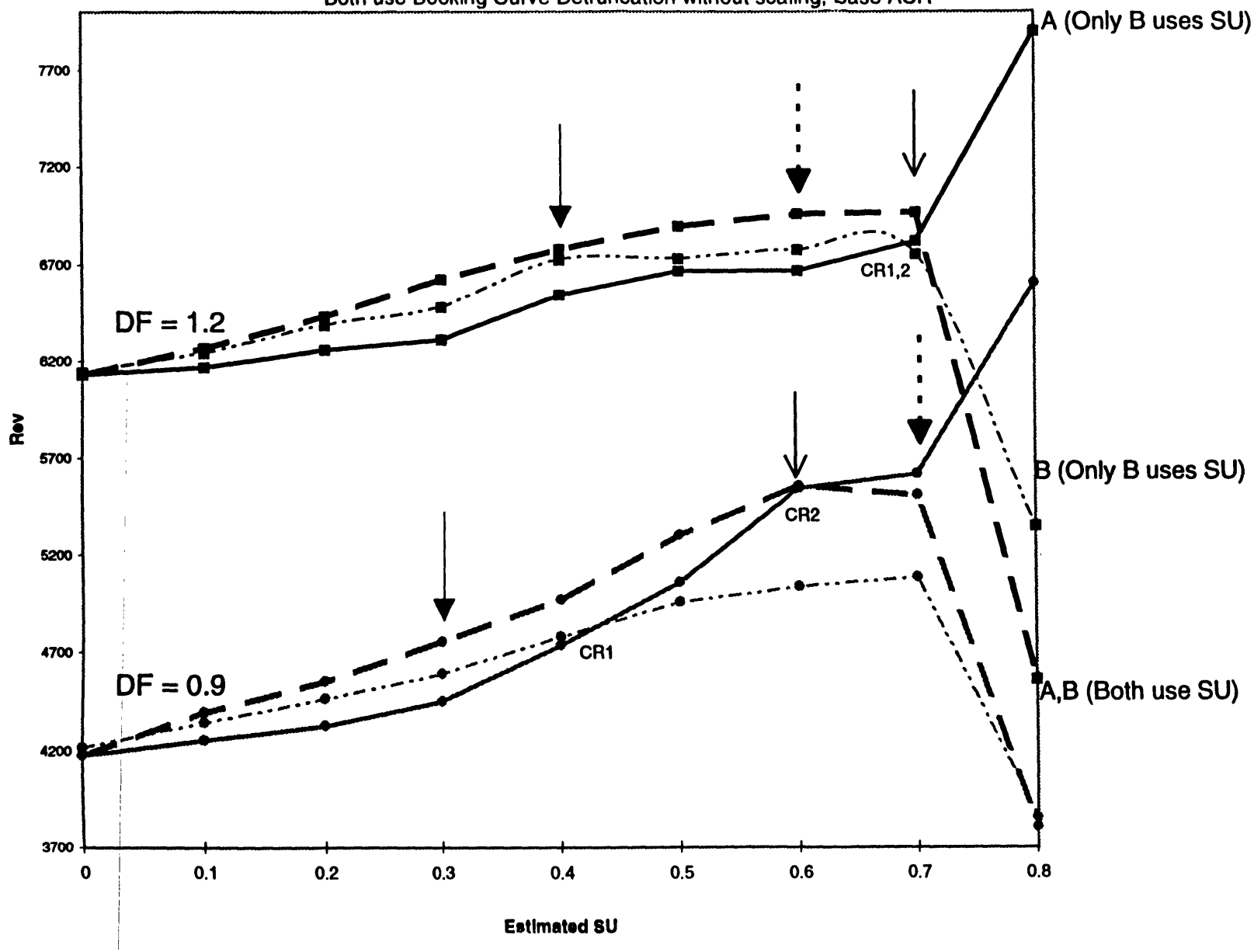
Base Case Best Estimated SU Rates and % Revenue Improvement					
Adoption By	Obj.	Best Est. SU Rate	B % Over A	Airline B % Over SU = 0.0	Airline A % Over SU = 0.0
B Only	I	0.3	3.17%	8.80%	6.49%
B Only	II	0.7	-9.46%	20.54%	34.44%
Both	III	0.6	N/A	32.98%	32.98%

Table 6.32. Sell-Up Objective Performance, Base Case. DF = 0.9, Base ACR.

The best Airline B can individually do to distinguish its revenue results from its competitor (Obj. I) is to assume a modest estimated *SU = 0.3*, yielding a 3.17% revenue difference over Airline A, and a 8.80% improvement over not adjusting for estimated SU. In this case Airline A still gains 6.49% by doing nothing! If Airline B adopts a simple myopic "best improvement" philosophy, its revenues can improve 20.54% with an estimated *SU = 0.7*, but it will underperform Airline A by 9.46%. Two reasons suggest that these revenue results are likely *overestimates* of the benefits to consideration of sell-up: First, this simulation considered *very few alternatives* relative to the actual airline marketplace, where an O/D pair may be served by many carriers using connections through their respective hubs. Second, our input values on ACR

Figure 6.17. Revenues with Increasing SU, DF = 0.9 and 1.2

Both use Booking Curve Detruncation without scaling; Base ACR



(passenger price sensitivity) may be biased upward. I will examine the too-few alternatives problem in Section 6.3.4 and the price sensitivity issue in Section 6.3.2.

When the base case scenario of Figure 6.16 is repeated at a high DF (DF = 1.2), we have Figure 6.17 (which graphs the results of Figure 6.16 along with the high demand case). The impacted demand conditions increase the total number of price insensitive nondiscretionary travelers in the market. However, most of the revenue benefit of this effect is *already* gained by the seat optimizer (compare the revenue conditions under *SU = 0.0* for DF = 0.9 and 1.2, and the slopes for the revenue graphs as estimated *SU* increases). Thus, *proportionate* revenue improvements under any objective are lower at higher demand. This is seen by comparing the shaded boxes of Table 6.33 with Table 6.32: The revenues Airline B may expect to gain via individual adjustment and myopic individual maximization decline from 20.54% to 10.31%. Joint improvements experience similarly steep declines, from 32.98% to 13.64%.

The declining revenue improvements in high demand situations suggest that the forecaster, detruncator, and seat optimizer[161] used in these tests respond exactly as expected: Booking limits on low-value classes are adjusted downward, protecting seat inventories for later-arriving high-value passengers. However, it also indicates that without adjustment of bookings limits for *SU* or another method to induce sell-up (see Section 3.3.2), *underprotection* is a significant problem on the many flights in the airline industry with moderate demand levels.

DF = 1.2 Best Estimated SU Rates and % Revenue Improvement					
Adoption By	Obj.	Best Est. SU Rate	B % Over A	Airline B % Over SU = 0.0	Airline A % Over SU = 0.0
B Only	I	0.4	2.78%	9.44%	6.65%
B Only	II	0.6	1.60%	10.31%	8.74%
Both	III	0.7	N/A	13.64%	13.64%

Table 6.33. Sell-Up Objective Performance, DF = 1.2; Base ACR

While the proportional improvements to sell-up adjustment are less with high demand, the optima and critical points all generally shift out to higher estimated *SU*. When only Airline B adjusts for estimated *SU*, Airline A experiences lower revenues over a *wider SU* range. This may be explained with reference to Figures 6.18 and 6.19, which give loads for both airlines by fare

[161] Recall the following base case settings: The forecasting method is pickup, the detruncation method is booking curve without scaling, and the seat optimizer is EMSRb.

Figure 6.18a. Airline A Ttl and Per Class Loads, DF=0.9, Base ACR
Only Airline B Adopts Scaling.

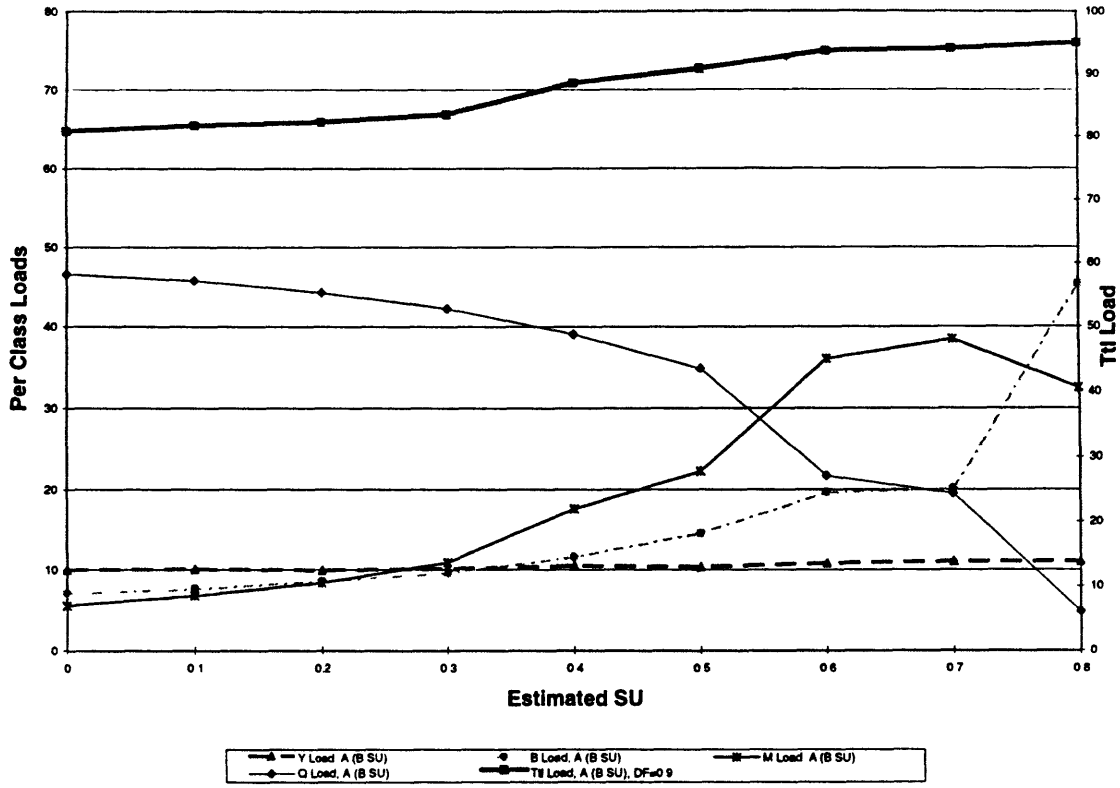


Figure 6.18b. Airline B Ttl and Per Class Loads, DF=0.9, Base ACR
Only Airline B Adopts Scaling.

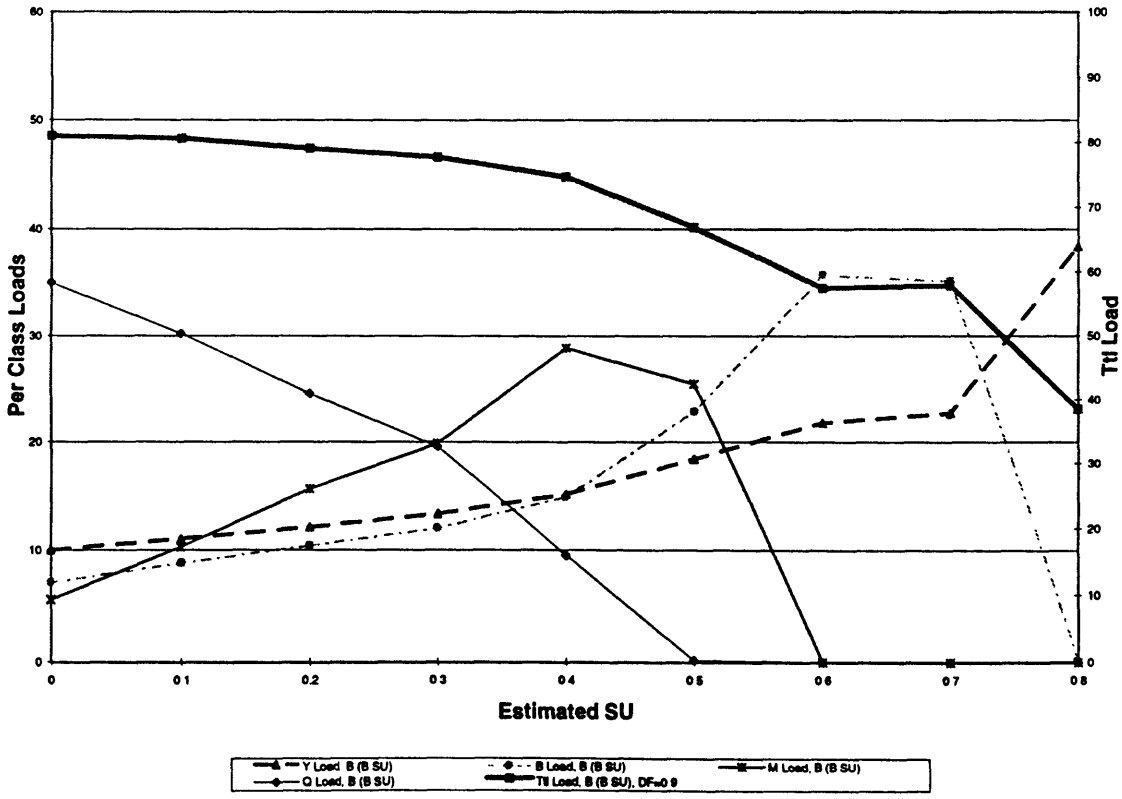


Figure 6.19a. Airline A Ttl and Per Class Loads, DF=1.2, Base ACR
Only Airline B Adopts Scaling.

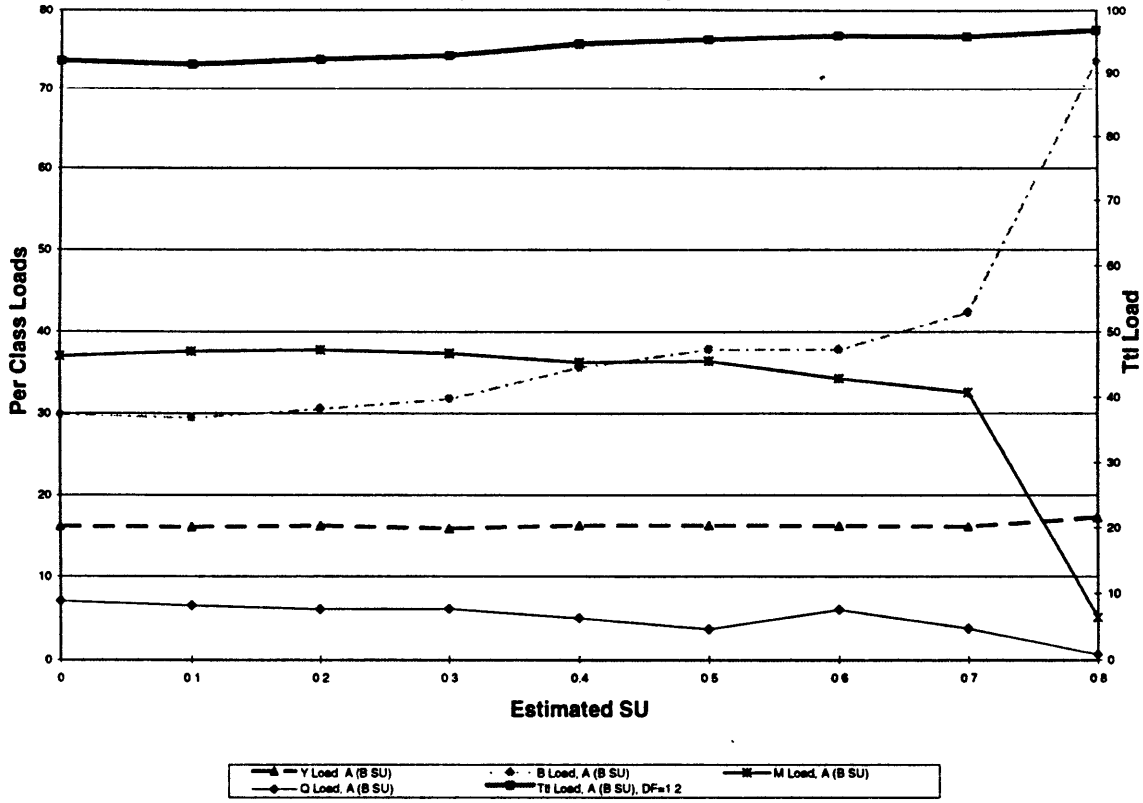
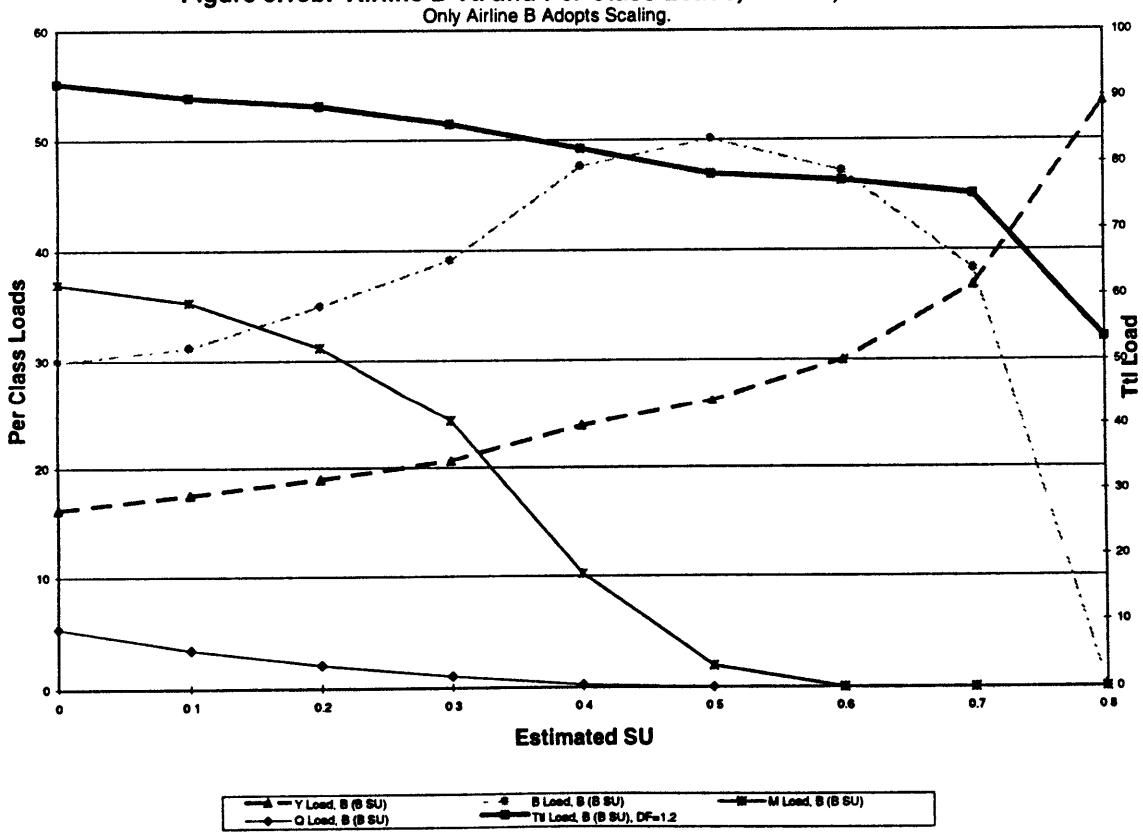


Figure 6.19b. Airline B Ttl and Per Class Loads, DF=1.2, Base ACR
Only Airline B Adopts Scaling.



class and demand level when only Airline B adjusts booking limits for sell-up estimates. Under high demand conditions, Airline A is at near maximum capacity even at low estimated *SU* (Figure 6.19a). As Airline B increases its estimated *SU*, increasing numbers of M class passengers who are unwilling to sell-up are being "dumped" (Figure 6.19b) -- But Airline A's constant M loads indicate that the class is already closed. Relatively few of these passengers are willing to sell-up to B class, so Airline A is constrained from taking advantage. Only when Airline B's *SU* is extreme and even B class passengers are being dumped (see Figure 6.19b) can Airline A accept these passengers and score significant revenue differentials.

In contrast, under low demand conditions Airline A quickly gains the M class passengers denied by Airline B (compare Figures 6.18a and 6.18b). Interestingly, Airline A manages even to gain a limited number of high-value B class passengers despite *not* adjusting for *SU*. Airline A experiences gains not only from the accommodation of passengers denied by Airline B, but also market-wide positive externalities induced by Airline B's extreme protections at high estimated *SU*. In the basic moderate and high-demand scenarios, we may conclude:

- There are significant revenue benefits to be gained from adjusting the EMSRb seat optimizer for estimated *SU* rates. However, a "jointness" effect requires all airlines in a competitive marketplace to adopt this policy for full benefits to occur.
- Under individual adjustment, an airline can easily give more revenue to the non-adjuster (who does nothing) by overprotecting.
- Sell-up opportunities generally diminish at high demand factors, since in these cases the seat optimizers protect more for higher-valued seats, and passengers are already selling up to high-valued fare classes in quantity.

6.3.2. High Price Sensitivity Scenarios

As I have discussed in Section 6.3, crucial to the revenue improvements explored in Section 6.3.1 is PODS' underlying assumptions about passenger price sensitivity. Price sensitivity involves passenger response to two specific market events: First, the price for his or her initially desired fare product increases; second, the initially desired product is unavailable and more expensive alternatives are offered. This section tests an assumption of *higher* price sensitivity or *lower* willingness to pay in the latter case. As previously discussed, price sensitivities are represented by ACR. These tests examine the impact on the profitability of incorporating

estimated *SU* when ACRs are reduced by 25%. Low ACRs are then 3.75 for nondiscretionary and 1.5 for discretionary passengers. In the low ACR case, the former will still take the most expensive fare but discretionary travelers, on average, will take only the lowest two fare classes (the PODS base case fare structure is given in Table 4.1). Low ACR is tested at both DF = 0.9 and 1.2.

Results at low ACR are graphed against base conditions in Figure 6.20 for DF = 0.9. Clearly, most optima and critical points contract. Faced with closure of lower-priced fare options, passengers will sooner decide not to travel. Comparison of the top panel of Table 6.34 for DF = 0.9 and Table 6.32 clearly indicates that options for revenue improvement with low ACR are limited. The myopic individual profit maximization objective (Obj. II) for Airline B when only it adjusts for estimated *SU* records a paltry 6.83% (from 20.54) individual improvement. The benefits of collusion decline from a 32.98% revenue improvement under base ACR to 13.10%. Curiously, low ACR *increases* the maximum revenue difference Airline B is able to maintain against Airline A, from 3.17% to 4.61%. This is confirmed in Figure 6.20, where the distance between Airline B and Airline A revenues *increases* at low ACR, so that Airline B no longer experiences an improvement with *joint SU* adjustment. Diminishing returns to sell-up are clearly in effect here: Airline B has already exhausted most gains to modifying booking limits at relatively low *SU* estimates. Higher estimates cause rapid revenue deterioration, and the jointness effect is lacking.

Low ACR Best Estimated SU Rates and % Revenue Improvement, Variable DF						
DF	Adoption By	Obj.	Best Est. SU Rate	B % Over A	Airline B % Over SU = 0.0	Airline A % Over SU = 0.0
0.9	B Only	I	0.3	4.61%	4.89%	0.29%
	B Only	II	0.5	0.93%	6.83%	5.87%
	Both	III	0.5	N/A	13.10%	13.10%
1.2	B Only	I	0.3	6.47%	6.74%	0.92%
	B Only	II	0.4	5.34%	7.94%	3.14%
	Both	III	0.5	N/A	9.81%	9.81%

Table 6.34. Sell-Up Objective Performance at Low ACR and Variable DF

A high demand situation (DF=1.2) combined with low ACR in Figure 6.21 create a somewhat intermediate condition, with countervailing pushes on revenue performance. High

Figure 6.20. Revenues with Increasing SU, Variable ACR

Both use Booking Curve Detruncation without Scaling; DF = 0.9

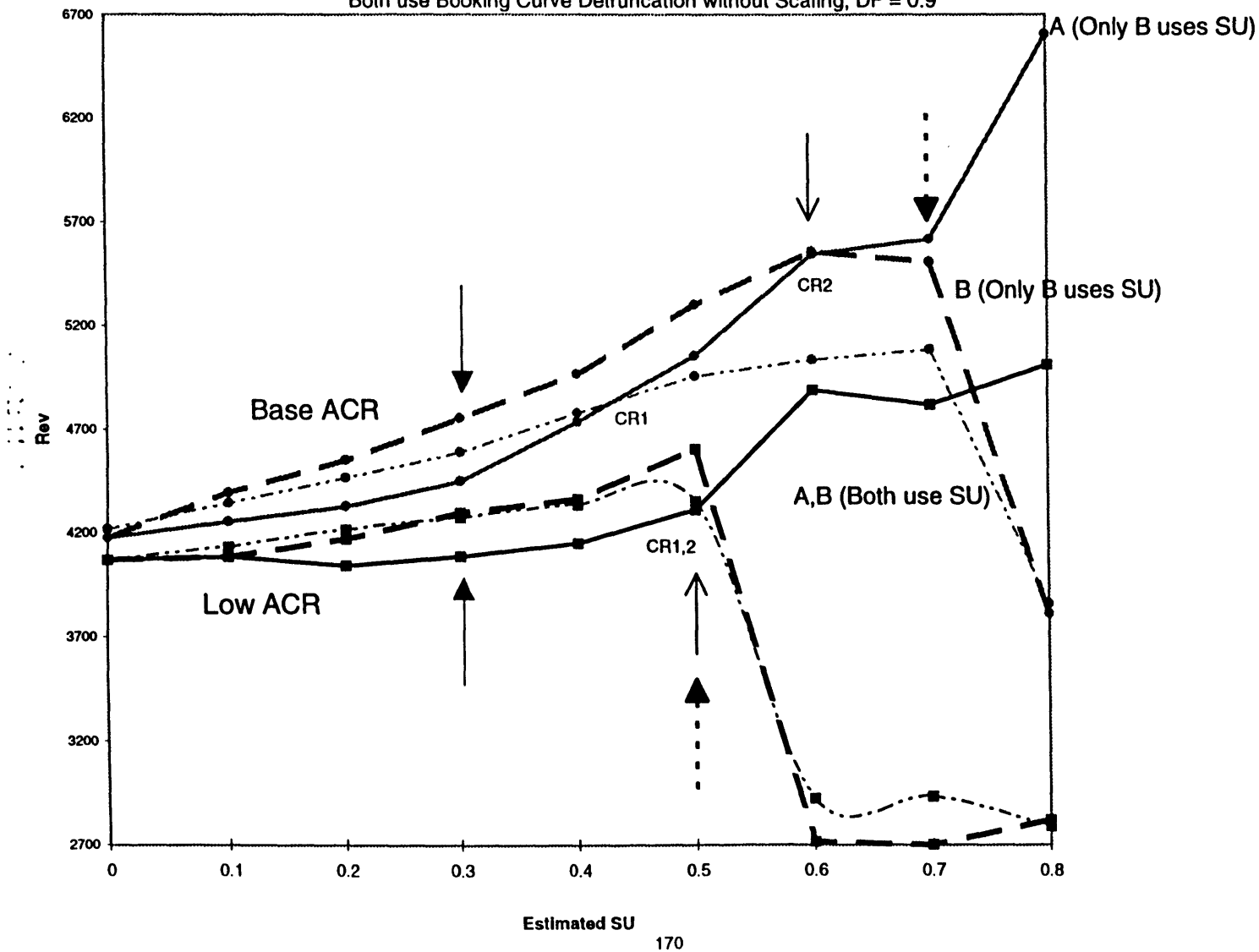
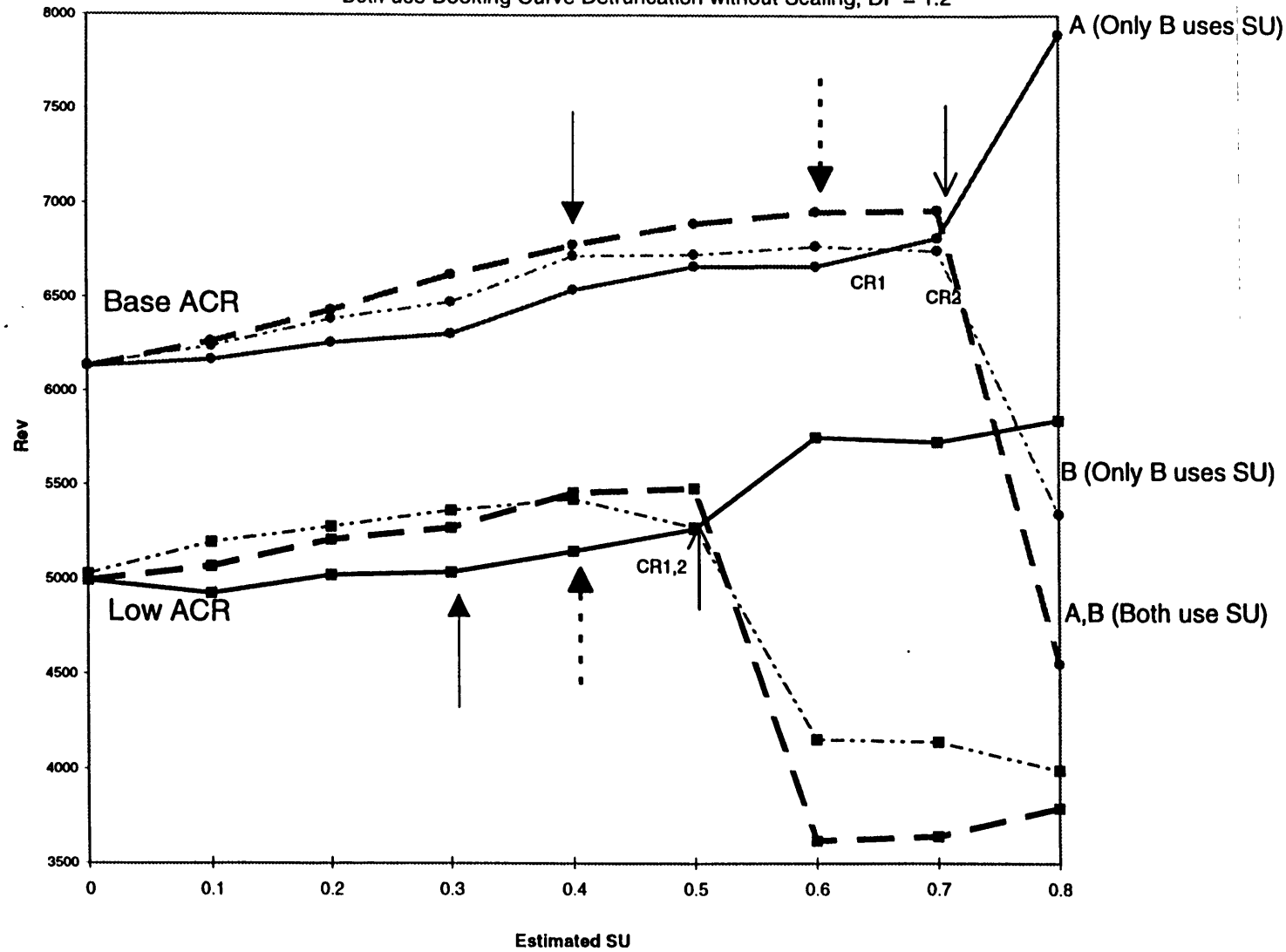


Figure 6.21. Revenues with Increasing SU, Variable ACR

Both use Booking Curve Detruncation without Scaling; DF = 1.2



demand increases sell-up potential, while high price sensitivity decreases it. Comparison between Figures 6.21 and 6.20 immediately reveals a dramatically enlarged revenue difference at any estimated *SU* rate between the base and low ACR scenarios. It follows that high demand factors *magnify* the negative effect on sell-up potential of high price sensitivity.

This result is consistent with observed closure properties at each DF (recall that sell-up may occur only with closure). At moderate demands, closure occurs a relatively low proportion of the time. Thus, since sell-up occurs infrequently, decreasing ACR has little effect: Underlying demand conditions do not require passengers to pay as much as they might. At high DF, however, closure occurs often in low-value classes as the seat optimizer protects for high-valued passengers. The number of *misidentified* passengers declines significantly. More passengers' willingness to pay is correctly matched with the appropriate fare class as they must sell-up to the extent they are willing. Decreasing ACR here substantially affects revenues, since the passengers formerly willing to pay more now refuse to travel. The distinct revenue inferiority at low ACR and high DF is not remedied by increasing estimated *SU*: Few passengers will sell-up.

Table 6.34 indicates that under low ACR, increasing demand factor allows Airline B to *increase* its revenue improvement over not adjusting for *SU* when it is not colluding (lined boxes), while both the revenue improvement when the airlines collude and Airline A's benefit to non-adjustment declines (shaded boxes). Airline B's earnings improvement over not adjusting for *SU* (Obj. II) increases from 6.83% to 7.94%. More importantly, Airline A is reduced from being only 0.93% to 5.34% behind Airline B. This effect is almost certainly due to high demand conditions. Adjusting for *SU* estimates enables Airline B to save seats for the few passengers willing to pay higher-value fares. Airline A cannot take advantage of this situation because high demands fill its capacity, and it lacks the *SU* adjustment mechanism to respond with equivalent protection. Thus, Airline B achieves revenues which are *superior* to the joint collusive outcome over a significant *SU* estimate range. Collusion in this case is *not* in Airline B's interest. As the innovator with the *SU* methodology, allowing its competitor to also adjust for *SU* estimates would lower revenue results. Both the airlines would adjust for the few passengers willing to sell-up, yielding higher refusals of low-valued passengers and lower revenue results. To summarize,

- Higher price sensitivity dramatically reduces the revenue benefits to both individual and joint adjustment of the seat optimizer for estimated *SU* rates.

- The revenue penalty of higher price sensitivity is magnified at high demand levels. This occurs because passengers' sell-up willingness is less likely to be exploited under moderate demands.
- However, the benefit to individual adoption *increases* at high demand factor, such that collusion may no longer be in the adopter's interest.

6.3.3. Booking Curve Scaling Scenarios

I have described in Section 5.3.2 why scaling under booking curve detruncation is necessary: Flights which close tend generally to have different booking curves than those which do not. Section 6.1.1.1 detailed why scaling the booking curve tends to make outliers of detruncated HDB observations, thereby increasing forecasts of unconstrained bookings. This effectively induces sell-up via additional protection. Combining scaling with sell-up may yield insights on the point at which scaling ceases to compensate for booking curve estimation issues and instead becomes a vehicle for sell-up. These tests assume both airlines adopt scaling between 0.2 and 1.0 (in increments of 0.2). Moderate demand (DF = 0.9) and base ACRs are assumed.

Figure 6.22a gives revenue results for each *pbscl* when only Airline B adopts estimated *SU* rates into the seat optimizer (representation of maxima and critical points was suppressed for clarity). When the scaling decreases from 1.0 to 0.2 and *SU* = *0.0*, revenues for both carriers consistently increase. As expected, Airline B's range of revenue superiority as it adopts estimated *SU* declines; Airline A is able to earn significantly superior revenues quickly as passengers' sell up propensity has been exhausted. After *pbscl* = *0.6*, nonzero *SU* results in no differential improvement for Airline B, and soon strict losses. The shaded areas in Figure 6.22a represent the range of superior revenues for Airline B. Besides clearly illustrating the diminishing returns to protection argument, this shows *pbscl* = *0.6* to be an absolute lower bound on the *pbscl* level which is no longer compensating for the booking curve bias but instead inducing sell-up (since attempting to induce more by individually adopting estimated *SU* has no effect). In these base-case ACR tests, the *pbscl* which only adjusts for the booking curve is in fact likely significantly higher than *0.6*.

Figure 6.22b provides analogous results when both airlines adopt. Inexplicably, more extreme scaling always yields additional revenues: Rather than inducing losses, very low *pbscl* = *0.2* simply causes higher estimated *SU* to have zero revenue effect. All curves tend to converge at estimated *SU* = *0.6*, after which there is precipitous decline. To gain insight into this effect, load

Figure 6.22a. A, B Single Rev. With Increasing SU, Variable Scaling

Both use Booking Curve Detruncation; DF = 0.9; Base ACR

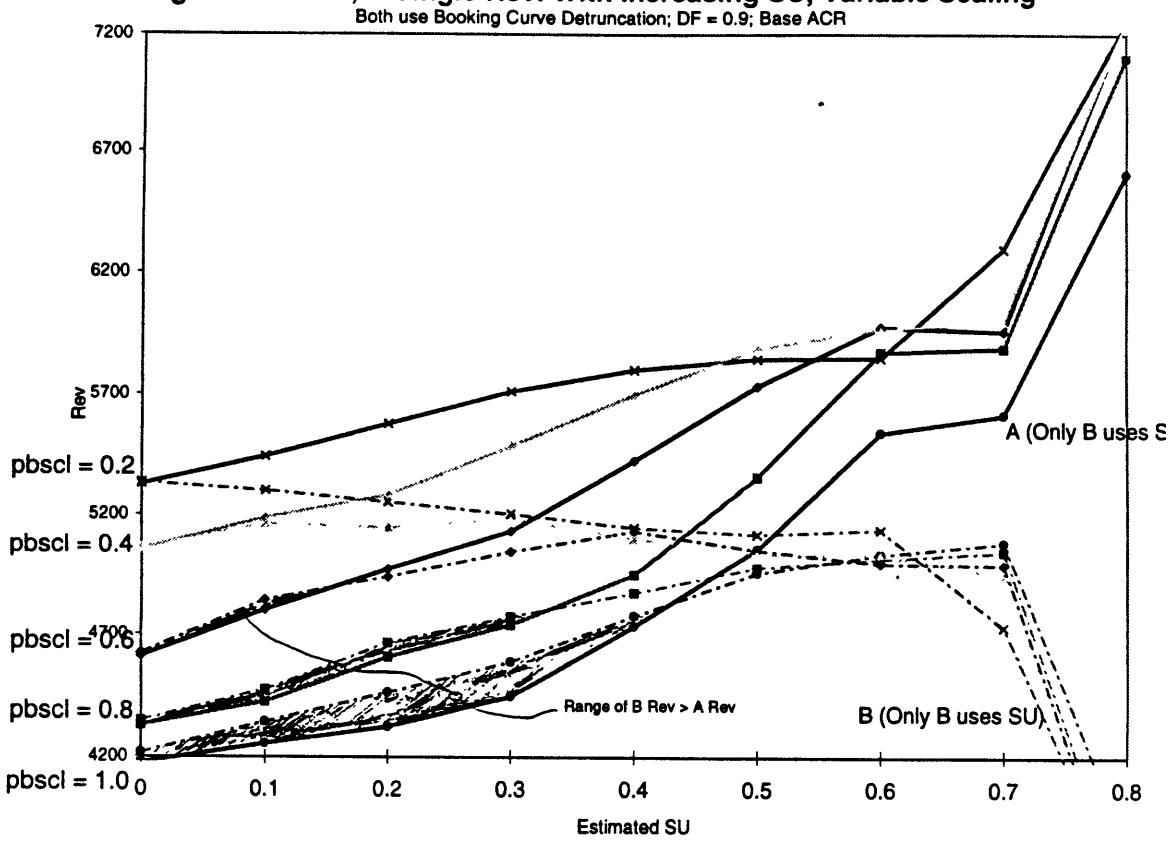


Figure 6.22b. A, B Joint Rev. With Increasing SU, Variable Scaling

Both use Booking Curve Detruncation; DF = 0.9; Base ACR

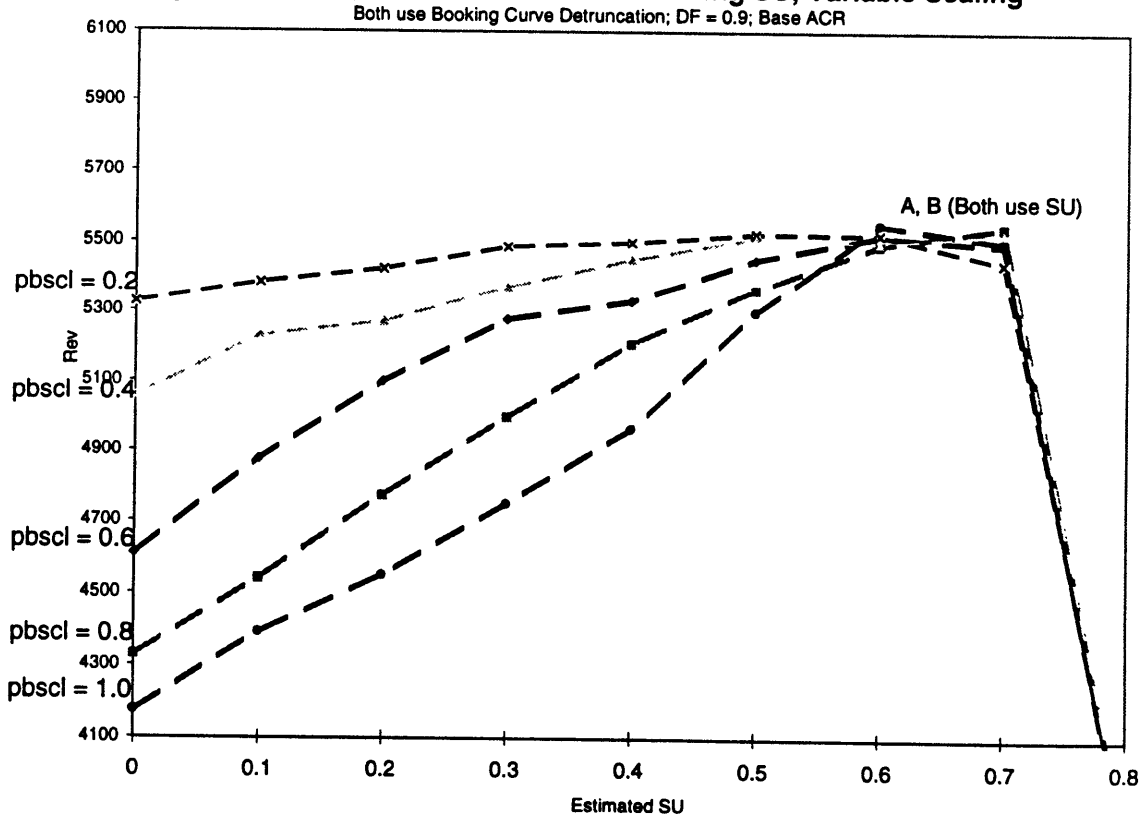
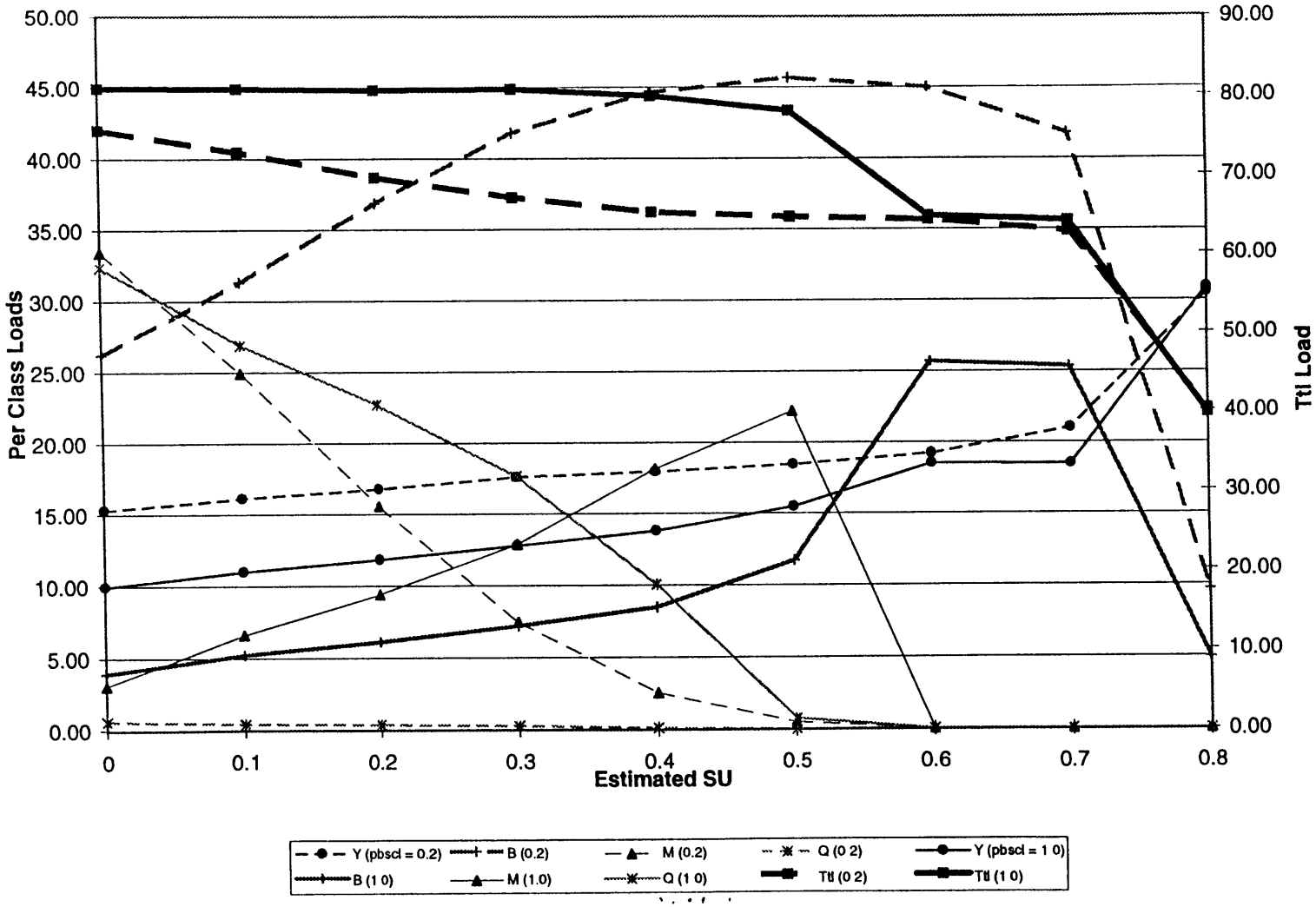


Figure 6.23. Loads under Joint Adjustment with Increasing SU

Both use Booking Curve Detruncation; pbscl = 0.2 or 1.; DF = 0.9; Base ACR



totals and by fare class for *pbscl* = *0.2* and *1.0* are graphed in Figure 6.23 for the joint case. In this case, both airlines have symmetric load conditions. Each fare class has the same symbol and line type in Figure 6.23, except that *pbscl* = *0.2* has dashed lines while *pbscl* = *1.0* has solid lines.

I will examine the results by fare class. As estimated *SU* increases, total loads under both scaling procedures declines, but more significantly for *pbscl* = *0.2*. The airlines then attain exactly equal loads for $SU \geq 0.6$. Scaling induces a superiority in Y class loads at all estimated *SU*, but this difference declines to approximate equality for $SU \geq 0.6$. B class loads are substantially different for most *SU* values: Scaling yields significantly higher loads, which increase with estimated *SU* until a tapering off. Dramatic declines occur after *SU* = *0.6* both with and without scaling as all but the higher fare classes close. A similar pattern is repeated for every fare class: Generally, scaling yields an improved load distribution. But after *SU* = *0.6*, regardless of scaling level, *all* fare classes but Y close and equivalent revenue results are obtained. This explains the convergence in revenues of Figure 6.22b. It also explains why progressively higher *pbscl*, by itself, does not decrease revenues: Assuming both airlines adopt equal *pbscl*, no scaling level is sufficiently high to cause closure of B or even M class.

Variable pbscl Best Estimated SU Rates and % Revenue Improvement											
Objective:	*pbscl=1.0*		*pbscl=0.8*		*pbscl=0.6*		*pbscl=0.4*		*pbscl=0.2*		Base
	Best SU	%Ovr Base	Best SU	%Ovr Base	Best SU	%Ovr Base	Best SU	%Ovr Base	Best SU	%Ovr Base	
I	**0.3**	**3.2%**	0.2	1.3%	0.1	0.9%	0.0	0.2%	0.0	0.2%	A Rev
II	**0.7**	**20.5%**	0.7	16.0%	0.4	11.0%	0.3	2.4%	0.1	-0.8%	BRev(SU=0.0)
III	**0.6**	**33.0%**	0.7	28.1%	0.6	19.7%	0.7	9.3%	0.5	4.8%	A/BRev(SU=0.0)

Table 6.35. Sell-Up Objective Performance under Variable pbscl; Base ACR; DF = 0.9

Table 6.35 gives objective performance data for *pbscl* between *0.2* and *1.0*. Because of space considerations, only the airline(s) direct objectives (outlined boxes in Tables 6.32-6.34) are included. Consistent with the results in Figures 6.22a and 6.22b, best estimated *SU* and percentage improvements all decline precipitously as *pbscl* increases. Airline B is completely unable to attain any revenue difference with its competitor past *pbscl* = *0.6* (Obj. I), while even under collusive conditions, by *pbscl* = *0.2* only a 4.8% improvement may be had over the base case where estimated *SU* = *0.0*. This confirms that successively lower *pbscl* completely removes

an incentive for individual adjustment for estimated *SU* rates, and significantly limits gains to joint adoption. In summary,

- Increased scaling limits additional individual and joint sell-up possibilities, since scaling the booking curve exerts an equivalent effect (i.e., increased protections for higher-valued fare classes).
- Benefits due to joint adjustment converge (regardless of *pbscl*) as higher estimated *SU* are adopted. This occurs because all but the highest-value classes are closed.

6.3.4. Additional Frequency Scenarios

The last sell-up tests I will discuss involve the effect of an increased number of *alternative* paths. This section is most relevant to industry practice, since many O/D pairs in the United States are served by a variety of airlines -- though not even one may serve the city-pair non-stop. Instead, passengers connect between flights at intermediate hubs which nearly every major airline has constructed. As the number of alternative flights on competitors increase, clearly an airline's ability to induce sell-up is limited: Passengers instead switch to the alternatives. Adjusting for assumed sell-up probability can only be effective if passengers are somewhat *captive*, or without available alternatives. It follows that *moderate* demand levels (e.g., DF = 0.9) are an essential requirement for more competing frequencies to impact Airline B's attempts to induce sell-up. At high demand levels, "alternative" flights are always full, and do not enlarge the choice set of passengers subject to Airline B's estimated *SU* policy.

The effect of more alternative paths on the *adopting* carrier is not clear. How should additional frequencies have any effect if the adopting airline adjusts for sell-up on all its flights, and the same unconstrained demand level obtains on all flights in the market? On the other hand, any one passenger who arrives at a given time in the booking process will have a larger number of available alternatives with the additional frequencies, assuming the booking patterns in the added flights are similar. To answer these questions, three cases are constructed:

Case 1: Airline A has 1/2/4 frequencies; B has 1. Airline B adjusts for SU.
Case 2: Airline A has 1/2/4 frequencies; Airline B has 1. Both adjust for SU.
Case 3: Airline A has 1 frequency; B has 1/2/4. Airline B adjusts for SU.

All frequencies are exactly equal with respect to attributes which might influence passenger choice (e.g., time of departure, airline preference), and reported results are on a *per-flight* basis. Thus, in Case 2 Airline A and Airline B should achieve approximately equal per flight results, because PODS presently does not include an adjustment for the S-curve effect indicating super-proportional market share for airlines which dominate the market[162]. Moderate demands of DF = 0.9 and base case ACRs are assumed.

Figure 6.24a graphs results for Case 1. Airline B's per-flight revenues clearly decline as Airline A's frequencies increase. The slope of Airline B's revenue lines with increasing estimated *SU* decline significantly, suggesting that efforts to induce further sell-up will certainly fail as more alternative paths become available. Surprisingly, Airline A's per flight revenues *also decline*, even though DF is constant. Airline A's gain is likely limited because its failure to estimate *SU* allows passenger switching among its flights, thereby reducing passenger captivity on its flights as well. Hence, sell-up occurring under moderate demands by the action of the seat optimizer alone is *also* limited as alternatives become available. Objective results listed in Table 6.36 indicate that under the two individual objectives (I and II) for Airline B, significantly lower proportional gains are achieved. The best it can individually gain declines from 20.5% if Airline A has one frequency to 5.6% if it has two, and only 3.5% if the competition has four frequencies.

Variable Airline A Freq. Best Estimated SU Rates and % Revenue Improvement							
Objective:	*A: 1 Freq*		*A: 2 Freq*		*A: 4 Freq*		Base
	Best SU	%Ovr Base	Best SU	%Ovr Base	Best SU	%Ovr Base	
I	**0.3**	**3.2%**	0.1	1.6%	0.1	1.9%	A Rev
II	**0.7**	**20.5%**	0.4	5.6%	0.2	3.5%	BRev(SU=0.0)
III	**0.6**	**33.0%**	0.6	32.0%	0.7	32.3%	A/BRev (SU=0.0)

Table 6.36. Sell Up Objective Performance under Variable Airline A Frequency. Base ACR; DF = 0.9

The joint adjustment in Case 2 is illustrated in Figure 6.24b. Despite the increasing frequencies by Airline A, when both airlines adjust for estimated *SU* there is *no appreciable effect* of additional frequencies on revenue achievement. This is confirmed in Table 6.36, which

[162] The mechanism currently exists via modification of coefficient of preference on airlines. These tests assume the same value for this variable, regardless of frequency.

Fig. 6.24a. Single Per-Flight Rev. with Increasing SU, Variable A Freq

B has 1 frequency. Both use Booking Curve Detruncation without Scaling; DF = 0.9

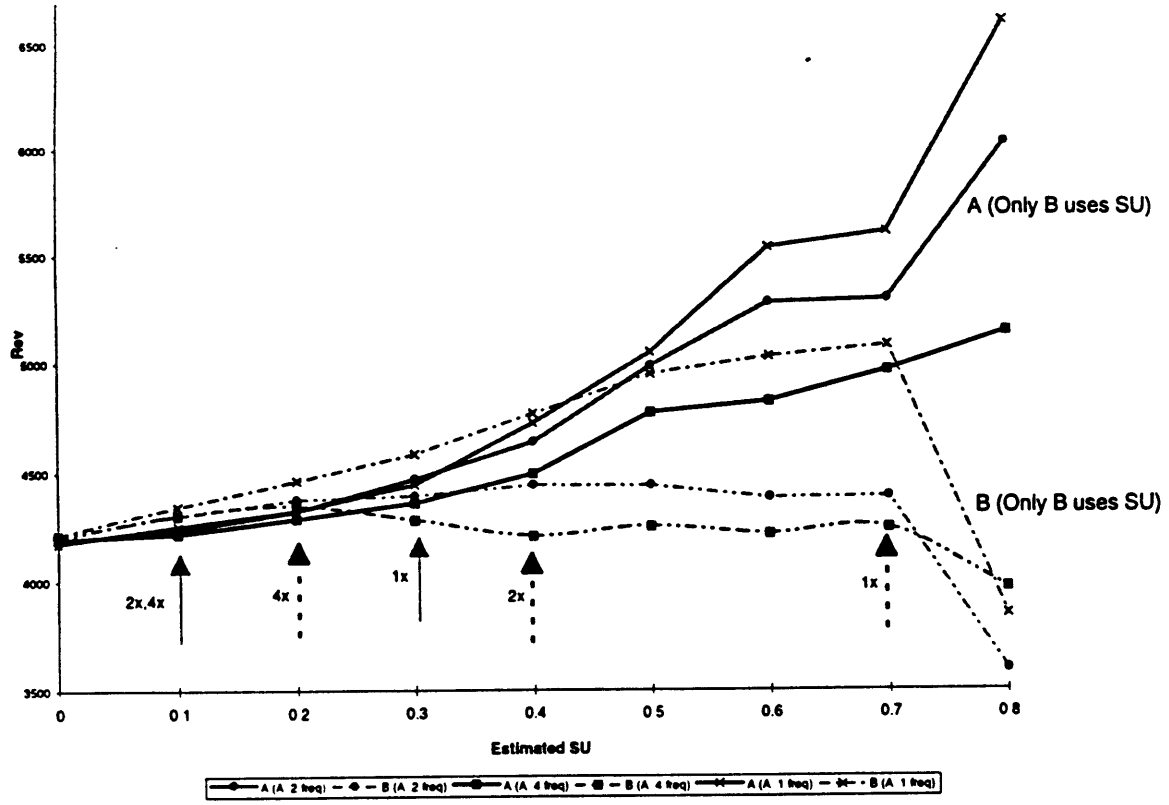


Fig. 6.24b. Joint Per-Flight Rev. with Increasing SU, Variable A Freq

B has 1 frequency. Both use Booking Curve Detruncation without Scaling; DF = 0.9

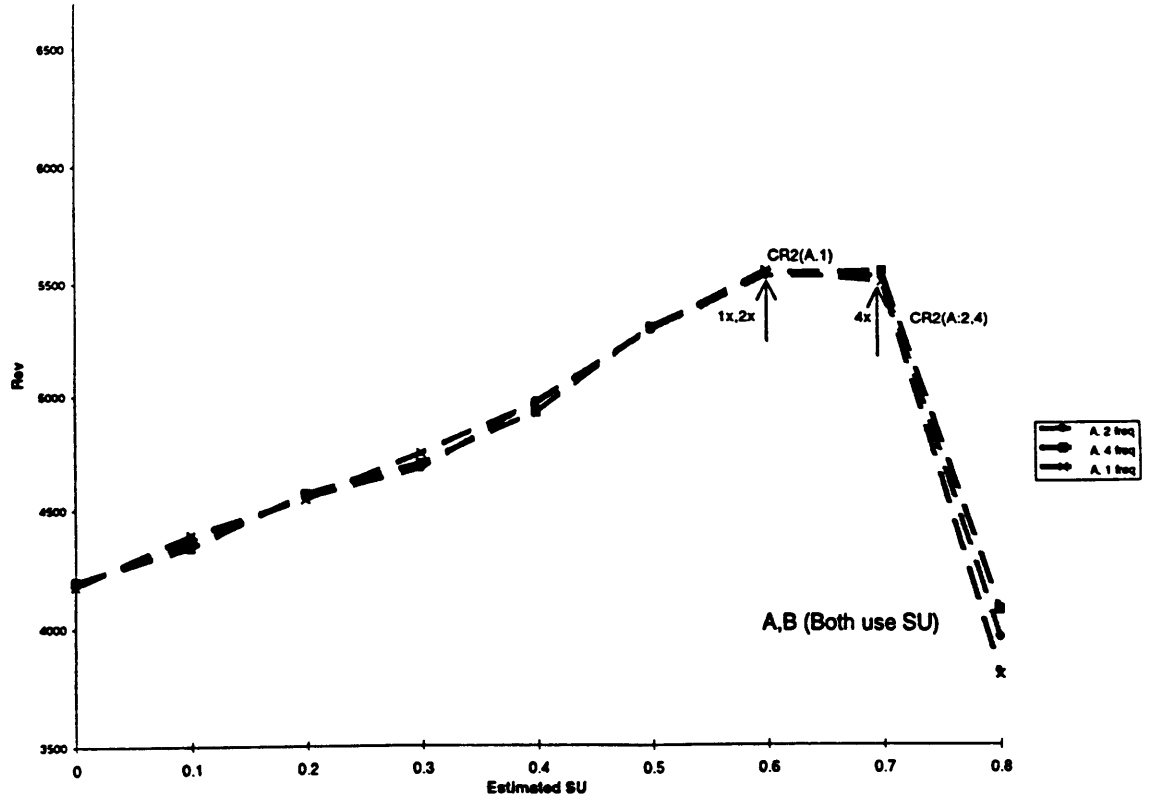
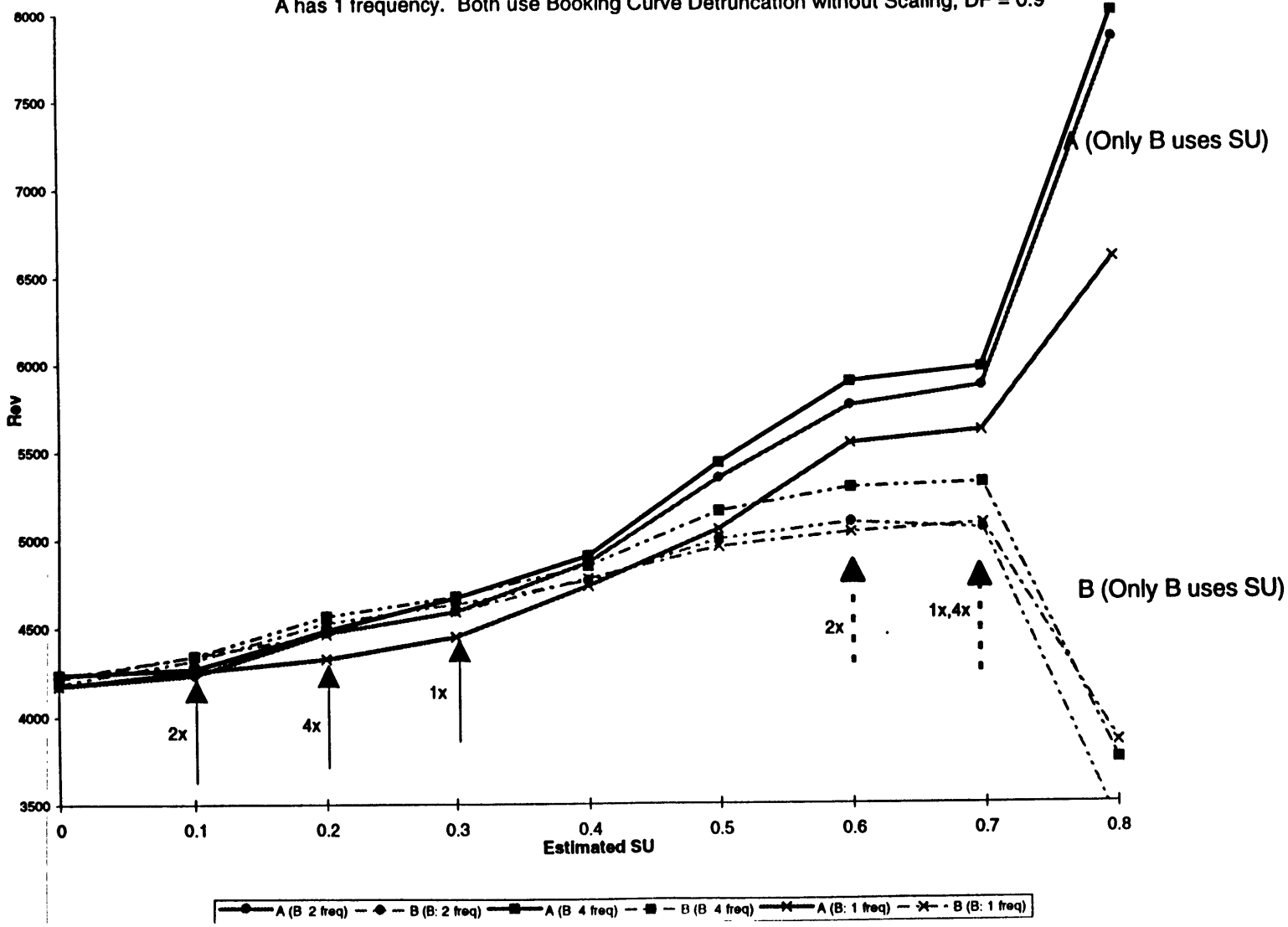


Fig. 6.25. Single Per-Flight Rev. with Increasing SU, Variable B Freq

A has 1 frequency. Both use Booking Curve Detruncation without Scaling; DF = 0.9



indicates that the maximize joint gains objective (Obj. III) suffers no significant loss in percentage improvement as frequencies increase (shaded boxes). Such a result underscores the revenue benefits to collusion when multiple travel alternatives exist.

What about the intermediate case, when only Airline B introduces estimated *SU* but also has the majority of frequencies? Results for this Case 3 are given in Figure 6.25. Here the position of revenue lines are *reversed* from Figure 6.24a. Not only do Airline B's revenues per flight at any estimated *SU* increase as it introduces more frequencies, but Airline A *also* gains more on its one frequency as its competitor increases service. Airline A's benefit may be due to the effect of receiving passengers unwilling to sell-up from *four* flights on Airline B. Figure 6.25 indicates that a *dominant* airline can achieve substantial benefits in attempting to induce sell-up *despite* the noncooperation of competitors. However, this comes at the cost of providing some benefit to the airline which does *not* adopt sell-up. These effects are also shown in Table 6.37, where Airline B's myopic individual improvement objective (Obj. II) yields increasing benefits as frequencies increase. If Airline B wishes to maximize its revenue superiority over the competition (Obj. I), it is *less* able to do so: This objective declines from 3.2% difference with Airline A when Airline B has one frequency to 1.8% per flight if it has four (shaded boxes).

Variable Airline B Freq. Best Estimated SU Rates and % Revenue Improvement							
Objective:	*B: 1 Freq*		*B: 2 Freq*		*B: 4 Freq*		Base
	Best SU	%Ovr Base	Best SU	%Ovr Base	Best SU	%Ovr Base	
I	**0.3**	**3.2%**	0.1	2.1%	0.2	1.8%	A Rev
II	**0.7**	**20.5%**	0.6	21.5%	0.7	25.9%	BRev(SU=0.0)

Table 6.37. Sell Up Objective Performance under Variable Airline B Frequency. Base ACR; DF = 0.9

We may conclude that attempts to induce sell-up beyond that provided by use of the seat optimizer are most successful if airlines in the marketplace collude. This thesis has demonstrated that revenue and load performance for an airline are strongly governed by input methodologies chosen by other airlines in the market. Because of this, it is sensible that explicit agreement among airlines on which seat optimizing and input methodologies to use is collusive and probably

illegal[163]. Hence, the *joint* adjustment case seems somewhat unlikely in practice. The most realistic application of the Belobaba/Weatherford strategy of modifying sell-up rates is in markets over which a carrier is the dominant airline. To summarize this multiple frequency section:

- Opportunities for inducing sell-up exist via booking limit modification are limited if an airline is a weak competitor, with few frequencies in the market. The non-adjusting airline gains *more* than the adjusting airline at low estimated *SU* rates.
- Opportunities are substantially enhanced if an airline is the dominant player in the market. However, the non-adjusting carrier will gain as a result of this pricing action.
- Revenues are highest if airlines *collude* and jointly adjust for the estimated *SU* methodology. But this is illegal, and therefore unlikely.

This completes the results section of the thesis. Chapter 7 now concludes with a comprehensive summary of significant findings.

[163] For example, agreement to adopt in common a sell-up inducement methodology like the Belobaba/Weatherford *SU* adjustment mechanism could be interpreted as an attempt to *fix prices*. This mechanism affects the closure properties of fare classes. Common adoption causes mimicry of available prices (i.e., airlines will tend to close fare classes simultaneously), assuming the carriers involved face similar demand conditions.

VII. Summary and Future Directions

7.1. Synopsis of the Thesis

7.1.1. Definitions

In Chapter 1, this thesis defined the *seat inventory control process* to be the method by which airlines create and then offer for sale various fare products in limited quantity. These products are targeted to match the different characteristics and sensitivities of the passenger types in the markets served by the airline. The purpose of this thesis was to examine the revenue impacts of alternative input methodologies for the *forecasting, detruncation,* and estimated *sell-up adjustment* steps of seat inventory control.

Forecasting is an essential part of the control process. Its purpose is to estimate mean unconstrained booking demand and the associated standard deviation by fare class. This information is then used by the seat optimizer to set booking limits for each fare class on a future flight. Inadequate forecasting will cause the seat optimizer to set booking limits incorrectly, resulting in less than optimal revenues.

Detruncation is also necessary when some of the historical data used by the forecaster is constrained. This occurs when one or more fare classes on a flight receive enough bookings that the booking limit is reached, and no more bookings in the affected fare class are permitted. The bookings which would have occurred *had space been available* are not known, and must be estimated through detruncation. Failure to detruncate or inadequate detruncation cause the forecaster to underestimate booking demand, resulting in the seat optimizer allocating too few seats to the affected fare classes. This is true because flights with high bookings are either absent from the data used by the forecaster (with no detruncation) or have estimates of final unconstrained bookings which are too low (with inadequate detruncation). Revenue losses result.

Finally, explicit consideration of *sell-up* (passenger willingness to buy a more expensive fare product when the initially desired product is unavailable) may allow airlines to extract more revenues from passengers. This occurs because the fare structures used by airlines to segment passengers are imperfect (i.e., some passengers will be able to take a less expensive fare than is targeted to them), and passengers will only sell up when they are denied the opportunity to take a

cheaper fare. Modifying booking limits to *induce* further sell-up than already occurs with the existing fare structure and seat optimizer may lead to revenue improvements.

7.1.2. Review of Past Comparative Studies and Models

After a discussion in Chapter 2 of whether the multiple fares offered by airlines are "discriminatory" in an economic sense, the thesis proceeds in Chapter 3 with an analysis of techniques for each of the three input methodologies, and a critical review of past comparative studies.

Most comparative studies of forecasters have been based on measures of *forecast error*, or the difference between predicted and actual bookings in a fare class over a data set of flights. I argue that this comparative approach causes inherent *biases* which disfavor particular forecasters, since either or both the forecast dataset and the base for the measurement of forecast error are usually *constrained.* Further, the goal of achieving zero forecast error is not only impossible to achieve, but misguided in principle: The assumption that zero error for each fare class maximizes revenues is almost certainly incorrect. It is commonly assumed that demands by fare class are independent. In fact, demands are *interdependent,* so setting booking limits *too low* in a fare class (causing a negative forecast error) may cause passengers to switch to a higher-valued fare class, thus *increasing* revenues. I argue that a more appropriate comparative metric is *revenues.* This thesis adopts the revenue comparison approach, using the Passenger Origin and Destination Simulator (or PODS), a comprehensive simulator of passenger behavior and seat inventory control developed by the Boeing Commercial Airplane Group.

Until this thesis, there have been no known comparative studies of detruncators. A revenue approach is also adopted for this comparative study of detruncation mechanisms. Finally, no comparative studies of alternative sell-up mechanisms exist. Most of the sell-up methods presented in the literature and discussed in Chapter 3 are *ad hoc* methods, and so are not amenable to systematic comparison. Therefore, this thesis abandons the comparative approach for sell-up analysis and instead focuses on the revenue effects of one formal approach to inducing additional sell-up.

7.1.3. Models Tested in This Thesis

Chapter 4 contains a limited discussion of the construction and assumptions of the PODS simulation. More detailed information is found in Wilson (1995). Chapter Five discusses the "base-case" simulation environment assumed in PODS tests, and examines the models compared and tested in the thesis. The base-case environment includes only two airlines each operating a single non-stop flight in one isolated market. Both airlines use the EMSRb seat optimizer. Three forecasting models are tested: Regression, pickup, and the "efficient" forecaster developed at Boeing. Regression estimates a linear relationship between bookings-in-hand on a flight being forecast at a particular booking interval, and bookings-to-come until departure on that flight[164]. This estimation is done utilizing data on historical (previously departed) flights; generally, a *positive* relationship is estimated between these variables. The pickup forecaster, on the other hand, disregards bookings-in-hand on the forecast flight and assumes that bookings-to-come or "pickup" will be a simple average of bookings-to-come on historical flights. The efficient forecaster is a hybrid model which assumes that bookings-to-come are related to bookings-in-hand on the forecast flight *and* historical flights.

Four different detruncation methods are tested: No detruncation, booking curve detruncation, projection detruncation, and pickup detruncation. Without detruncation, data for fare classes on flights which close are excluded from historical data. Booking curve detruncation estimates unconstrained bookings on closed flights by assuming that the proportional relationship between bookings-in-hand at the booking interval of closure and bookings-in-hand at departure is *constant* over all flights in the fare class of interest. Moderate or extreme "scaling" of the booking curve by constant factors is also tested to adjust for the questionable validity of this assumption. Projection detruncation estimates unconstrained bookings assuming that the conditional probability of receiving more than the forecast bookings (given closure) should be a specified constant. Similar to pickup forecasting, pickup detruncation assumes that bookings-to-come between the closure interval on the closed flight and departure is best approximated by average bookings-to-come over the same intervals on unclosed flights.

Finally, the sell-up methodology tested is due to Belobaba and Weatherford (1996). Their construction takes an input estimated sell-up rate for each fare class (or proportion of passengers

[164] It is proved in Chapter 5 that the estimated relationship may also be between bookings-in-hand and final bookings on the forecast flight, without changing the regression forecast.

in the fare class who will sell-up when it is closed), and modifies EMSRb booking limits accordingly. A large range of estimated sell-up rates is tested in the thesis.

7.2. Principal Findings

7.2.1. Forecasting Methods

Regression and the efficient forecaster were each compared relative to "base-case" pickup forecasting. Between regression and pickup forecasting, which method was superior depended significantly on the detruncator assumed, as well as the demand factor. At low demand levels, there was never a large difference between forecasters regardless of detruncation method. Without detruncation, both forecasters were approximately equal regardless of demand. With booking curve detruncation and no scaling, pickup forecasting performed significantly better than regression as demand increased. This was less true if the booking curves used for detruncation were also scaled. When extreme scaling was applied, there was no difference between the pickup and regression forecaster at any demand level. Under projection detruncation, regression became increasingly superior as demand increased. With pickup detruncation, there were no significant differences between regression and pickup forecasting, regardless of demand level.

Regression's improved performance at high demand factors (under detruncators with relatively *high* estimates of unconstrained bookings on closed flights) was attributed to its disproportional weighting of outliers. Tests were performed under conditions of increasing "booking curve variability," which increases uncertainty in the proportion of total bookings which will have been received by any particular booking interval. The performance of regression forecasting declines relative to pickup forecasting as this uncertainty increases, especially at high demand factors. This is true unless booking curve detruncation with extreme scaling is used, in which case pickup performs equally with regression forecasting. Final tests compared the regression and pickup forecaster when variability in demand between successive flights increases. Pickup forecasting performed worse than regression as demand variability increases under all detruncation methods but booking curve detruncation with scaling. Such a result is explained by the increased presence of outliers as demand variability increases. Regression disproportionally weights outliers, and so forecasts demand for *more* seats for high-valued fare classes than pickup forecasting in these circumstances.

The second comparison was between efficient and pickup forecasting. The efficient forecaster consistently underperformed pickup forecasting, worsening as demand level increases. I argue that this results from the efficient forecaster's system of weighting observations used to make its forecast. Its underperformance is magnified as booking curve variability increases, mostly because the weighting bias is magnified as proportional booking relationships are less predictable. However, the efficient forecaster generally improves relative to pickup forecasting as demand variability increases. While results are not entirely consistent, this effect is probably the result of pickup forecasting's inability to disproportionally weight outlier observations.

Under a wide range of market conditions, the efficient forecaster is inferior on a revenue basis to pickup forecasting. Pickup forecasting usually performs at least as well as the regression forecaster, and significantly better in certain circumstances (e.g., high demand or booking curve variability conditions). One pointed exception is a high systemwide demand variability scenario, in which case regression forecasting is superior. A qualified revenue ranking of these three forecasters places pickup forecasting first, regression forecasting second, and the efficient forecaster third.

7.2.2. Detruncation Methods

Detruncation alternatives were compared against the "base-case" booking curve detruncation without scaling under base-case environmental conditions and the same alternatives discussed in Section 7.2.1. Low demand factors eliminated any difference between detruncation methods. This will occur because there are very few flights to detruncate when demand levels are low relative to capacity. No detruncation is substantially inferior to booking curve detruncation without scaling as demand factor increases. Curiously, pickup detruncation is not much better. This is attributed to underprotection by pickup detruncation, which assumes that the *absolute* bookings increase on closed flights is approximated by the average increase on *unclosed* flights. Obviously, flights close because of high demands, which by definition receive *more* bookings than (unclosed) low and moderate-demand flights!

Booking curve detruncation *without* scaling underperforms booking curve detruncation *with* moderate or extreme scaling because the latter two methods protect more seats for higher-valued fare classes, thus inducing sell-up. Revenue losses due to not scaling increase with

demand level. However, there are limitations to the sell-up which may be induced by scaling: More extreme scaling of booking curves for detruncation eventually causes revenue *declines* as passengers refuse to sell up. Projection detruncation and booking curve detruncation without scaling are approximately revenue equivalent except at the highest demand levels, in which case the former estimates higher unconstrained demand on closed flights, thereby gaining revenue superiority.

When booking curve variability increases, any revenue difference is intensified. Thus, the performance of booking curve detruncation without scaling declines relative to booking curve detruncation with scaling, but does increasingly better than pickup detruncation. This effect is magnified at high demand factors. Outliers probably explain these results: More outliers are generated as booking curve variability increases. Thus, the "inferior" detruncator (which estimates lower final bookings on closed flights) will be applied to a greater proportion of observations used by the forecaster. These effects are generally repeated as variability in demand increases between successive flights. Booking curve detruncation with moderate or extreme scaling and projection detruncation perform better, because more outliers are generated under high demand variability conditions.

Previous PODS reports on the revenue benefits of a seat optimizer (i.e., Wilson [1995]) always assumed that base-case booking curve detruncation was automatically adopted along with a seat optimizer. Since this thesis indicates that *no detruncation* exhibits very poor performance relative to the base-case detruncator, there was some concern that most of the revenue benefit previously attributed to use of a seat optimizer was in fact due to use of detruncation. To address this issue, a special case with three airlines -- one with FCFS (first-come-first-served, where different fare products are offered but no booking limits are set), one with EMSRb without detruncation, and one with EMSRb *and* base-case detruncation -- was constructed. Results indicated that adoption of a seat optimizer alone accounts for about 65% of Wilson's reported revenue improvement of EMSRb over FCFS. The remaining 35% is due to adoption of a detruncation method.

7.2.3. Sell-Up Analysis

The final tests performed in this thesis involve the Belobaba and Weatherford modification of booking limits strategy for inducing sell-up. Base-case booking curve detruncation without scaling and the pickup forecaster were used in sell-up tests unless otherwise noted. It is essential to note that quantified revenue improvements here are very sensitive to the assumptions PODS incorporates about passenger willingness to pay and the degree of imperfection in the base-case fare structure.

Given this caveat, there are generally *significant* benefits to be gained from incorporating sell-up estimates into the seat optimizer. These effects are magnified if collusion occurs in the market, and both carriers adjust for estimated sell-up rates. If only one airline in the market adjusts, its opportunities are somewhat limited: At relatively low estimated sell-up rates, it can give *more revenue to the competitor* (passengers refuse to sell-up and instead divert) than it earns from passengers selling up. Under high demand conditions, there are fewer opportunities for further sell-up. Seat optimizers already protect more seats for high-valued fare classes, and many passengers are already paying as much as they are willing.

The benefits to sell-up estimation under increased passenger price sensitivity was also tested. As expected, benefits substantially decline for the airline(s) estimating sell-up under high price sensitivity. Revenue losses due to this effect are significantly larger for all estimated sell-up rates under *high demand* conditions. This is another consequence of the fact that under moderate demand levels most sell-up opportunities are not being exploited, since closure is relatively rare.

Tests were also performed when the booking curve detruncation was *scaled.* Scaling the booking curve for detruncation limited the additional gains occurring with estimated sell-up rates, to an increasing degree as the scaling level became more extreme. This result confirms that scaling booking curves for detruncation has the *same effect* as introducing estimated sell-up rates: Protection levels for high-valued fare classes increase, resulting in higher loads in these classes and higher revenues.

Final tests for estimated sell-up rates involved additional frequencies in the market, where total market demand was a constant proportion of capacity as additional flights were added. Base case conditions with two airlines and moderate demand levels were assumed. It was hypothesized that sell-up depends on passenger captivity, which is clearly reduced if more alternative frequencies are available in the market. Sell-up tests were divided into three cases. In the first

case, the airline which does not adjust for estimated sell-up adds more frequencies. The second examined more frequencies by one airline when *both* airlines in the market jointly adjust. In the third case, the airline which does adjust adds additional frequencies.

First case results were as expected: The airline adjusting for estimated sell-up gained significantly less as competitive frequencies increased. However, the per-flight revenues for the *non-adjusting* carrier also declined. I argue that the negative impact of reduced passenger captivity is *not* isolated to the airline adjusting the seat optimizer to cause additional sell-up: The sell-up induced by the fare structure and seat optimizer *without adjustment* is also limited by additional alternatives. This suggests that the *per-flight* revenue benefit to adoption of a seat optimizer is reduced when multiple frequencies are available in the market.

In the second case, joint adjustment for estimated sell-up rates completely eliminates any per-flight revenue loss for either airline as the number of frequencies in the market increase. Collusion is completely able to thwart any negative "additional alternatives" effect because *joint* estimation of sell-up rates limits availability on *all* flights in the market. Third case results indicate that when the airline which adjusts for estimated sell-up adds frequencies, its per flight revenues increase -- though the airline which does not adjust for estimated sell-up also gains on its one flight. This indicates that a *dominant* airline may gain significant revenue benefits by adjusting the seat optimizer for estimated sell-up, even when smaller players in the market do not cooperate.

7.3. Future Directions

This thesis illustrates the versatility of the PODS simulator. As the first output of a joint MIT/Boeing collaboration on the revenue benefits of seat inventory control, Wilson (1995) used PODS to argue that adoption of seat optimizers significantly increases an airline's revenues. This second thesis has utilized PODS to argue that an airline's revenue performance is also significantly impacted by the *input methodologies* for seat inventory control used by airlines. Forecasting and detruncation method choice, and the modification of booking limits to induce sell-up all have significant impacts on revenues.

However, both these theses are constrained by the simplistic market environment assumed in these simulations. Only one market with non-stop flights by two airlines is considered[165].

[165] Wilson (1995) also tests a simple three-market scenario.

Among many other possible variables, preference variables and departure times were equal for these airlines. In fact, most markets for air travel in the United States (where "market" is defined between an origin and destination) have many alternatives, with varying degree of passenger allegiance to particular airlines based on frequent flyer plan membership, ubiquity of flights out of the origin, etc. A simulation environment which adjusts to these market realities would significantly improve the applicability of PODS results to actual airline practice.

The simple market environment creates another significant issue. Thus far, the MIT/Boeing collaboration has avoided analysis of seat optimization algorithms under *full network* market conditions. This is presently the most prominent issue in seat inventory control, and adds an entirely new level of complexity. Seat optimizers which have thus far been covered (EMSRa and EMSRb) are *leg-based* approaches which take forecasts and set booking limits for each leg independently.

But airline passengers today travel largely over interconnected *networks*, with trips involving multiple legs. A superior seat optimizer in this environment should theoretically forecast and set booking limits between each origin and destination in the network. Yet this solution is frustrated by data and computational limitations of large network optimization. Many "optimal" and heuristic models have been advanced, which purport to achieve approximately the revenue benefit of full network optimization *without* falling prey to its limitations. At present, no comprehensive, comparative revenue simulation of these alternative network seat optimizers exists. Applying PODS to this issue (due to commence shortly) will mark a significant step forward in the science of revenue management.

Bibliography

Alamdari, F.E. and Black, I.G. (1992). "Passengers' Choice of Airline Under Competition: The Use of the Logit Model." Transport Reviews 12, pp. 153-170.

Areeda, P. (1981). Antitrust Analysis: Problems, Text, Cases. Third Edition. Boston, MA: Little, Brown and Company.

Belobaba, P. (1987). Air Travel Demand and Airline Seat Inventory Management. Flight Transportation Laboratory Report R87-7. Cambridge, MA: Massachusetts Institute of Technology.

Belobaba, P. (1992). Optimal vs. Heuristic Methods for Nested Seat Allocation. Presentation to AGIFORS Reservations Control Study Group Meeting. Brussels, Belgium, May 1992.

Belobaba, P. and Weatherford, L. (1996). "Comparing Decision Rules that Incorporate Customer Diversion in Perishable Asset Revenue Management Situations." Decision Sciences 27 (2), pp. 343-363.

Ben-Akiva, M., and Lerman, S. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: MIT Press.

Ben-Akiva, M., et al. (1987). "Improving Airline Passenger Forecasts Using Reservation Data." Presentation at Fall 1987 ORSA/TIMS Conference, St. Louis, MO.

The Boeing Company (1988). 1988 Domestic Fare Survey. Renton, WA: The Boeing Company.

Borenstein, S. (1983). Price Discrimination in Free-Entry Markets. Ph.D. Dissertation, Department of Economics, MIT. Cambridge, MA: Massachusetts Institute of Technology.

Botimer, T. (1994). Airline Pricing and Fare Product Differentiation. Ph.D. Dissertation, Department of Civil and Environmental Engineering, MIT. Cambridge, MA: Massachusetts Institute of Technololgy.

Brumelle, S. et al. (1990). "Allocation of Airline Seats Between Stochastically Dependent Demands." Transportation Science 24, pp. 183-192.

Cohen, A. (1959). "Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated." Technometrics 1(3), pp. 217-237.

Cohen, J., and Cohen, P. (1975). Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences. New York, NY: John Wiley & Sons.

Crew, M. and Kleindorfer, P. (1986). The Economics of Public Utility Regulation. Cambridge, MA: MIT Press.

Curry, R.E. (1990). "Optimum Seat Allocation with Fare Classes Nested by Origins and Destinations," Transportation Science 24, pp. 193-204.

Duncanson, A.V. (1974) "Short-Term Traffic Forecasting." Proceedings of the Fourteenth Annual AGIFORS Symposium. [?]: AGIFORS.

Federal Aviation Administration (1995). Annual FAA Forecast Proceedings, 1995. United States Department of Transportation, FAA, Office of Aviation Policy. Washington, DC: Government Printing Office.

Ferea, J. (1996). Revenue Management Impacts of Multiple-Hub Codeshare Alliances. Flight Transportation Laboratory Report R96-x [*forthcoming*]. Cambridge, MA: Massachusetts Institute of Technology.

Frank, R. (1983). "When are Price Differentials Discriminatory?" Journal of Policy Analysis and Management 2 (2), pp. 238-255.

Hansen, M. (1990). "Airline Competition in a Hub-Dominated Environment: An Application of Noncooperative Game Theory." Transportation Research (B) 24, pp. 27-43.

Harris, P., and Marucci, G. (1983). "A Short-Term Forecasting Model." Proceedings of the Twenty-Third Annual AGIFORS Symposium. [?]: AGIFORS.

Hirschleifer, J., and Glazer, A. (1992). Price Theory and Applications. Fifth Edition. Englewood Cliffs, NJ: Prentice Hall.

Holm, C. (1995). Airline Overbooking Performance Measurement. Flight Transportation Laboratory Report R95-6. Cambridge, MA: Massachusetts Institute of Technology.

Hopperstad, C. (1991). Efficient Forecasting: 1/92 Equation Simplifications. Boeing Commercial Aircraft Company Internal Document. Renton, WA: Boeing.

Hopperstad, C. (1995). An Alternative Detruncation Method, Nov. 95. Boeing Commercial Aircraft Company Internal Document. Rendon, WA: Boeing.

Hopperstad, C. (1996). Passenger O/D Simulator -- PODS v.5. Boeing Commercial Aircraft Company Internal Document. Renton, WA: Boeing.

Hunt, A. (1994). "Assault on the Airline Industry: Private Antirtust Litigation and the Problem of Settlement." Journal of Air Law and Commerce 59 (4), pp. 983-1022.

Kanafani, A. K. (1983). Transportation Demand Analysis. New York, NY: McGraw-hill.

L'Heureux, E. (1986). "A New Twist in Forecasting Short-Term Passenger Pickup." Proceedings of the Twenty-Sixth Annual AGIFORS Symposium. [?]: AGIFORS.

Lee, A. (1990). Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process. Flight Transportation Laboratory Report R90-5. Cambridge, MA: Massachusetts Institute of Technology.

Nelson, W. (1982). Applied Life Data Analysis. New York, NY: John Wiley & Sons.

Pfeifer, P. (1989). "The Airline Discount Fare Allocation Problem." Decision Sciences 20, pp. 149-157.

Sa, J. (1987). Reservations Forecasting in Airline Yield Management. Flight Transportation Laboratory Report R87-1. Cambridge, MA: Massachusetts Institute of Technology.

Scandinavian Airlines System (1978). SHOPS: a Short-Term Passenger Forecasting System. Stockholm, Sweden: SAS.

Schneider, H. (1986). Truncated and Censored Samples from Normal Populations. New York, NY: Marcel Dekker, Inc.

Skwarek, D. (1996a). "Revenue and Traffic Impacts of Alternative Detruncation Methods." Reservations and Yield Management Study Group Proceedings. Zurich, Switzerland: AGIFORS.

Skwarek, D. (1996b). "Competitive Impacts of Incorporating Sell-Up Estimates Into Revenue Management." MIT/Industry Cooperative Research Program in Air Transportation, 1996 Annual Meeting. Flight Transportation Laboratory Report R96-?. Cambridge, MA: Massachusetts Institute of Technology.

Swan, W. (1978). A Systems Analysis of Scheduled Air Transportation Networks. Flight Transportation Laboratory Report R79-5. Cambridge, MA: Massachusetts Institute of Technology.

Tan, J. (1994). Leg-Based Heuristic Methods to Network Seat Inventory Control. Flight Transportation Laboratory Report R94-7. Cambridge, MA: Massachusetts Institute of Technology.

Taneja, N. K. (1978). Airline Traffic Forecasting. Lexington, MA: Lexington Books.

Tirole, J. (1988). The Theory of Industrial Organization. Cambridge, MA: MIT Press.

Williamson, E. (1992). Airline Network Seat Inventory Control: Methodologies and Revenue Impacts. Flight Transportation Laboratory Report R92-3. Cambridge, MA: Massachusetts Institute of Technology.

Wilson, J. (1995). The Value of Revenue Management Innovation in a Competitive Airline Industry. Flight Transportation Laboratory Report R95-8. Cambridge, MA: Massachusetts Institute of Technology.