RESERVATIONS AND OVERBOOKING

Yuri Aroenber*  

INTRODUCTION  

The Random House dictionary of the English language defines overbooking as the practice of accepting reservations in excess of the number that can be accommodated. This practice has been prevalent in the airline industry for many years. Not so long ago, concerned with the adverse effect of such a disclosure, many airline companies would not even admit that they overbooked. Nowadays, however, after having been granted a legal right to engage in this activity, the airlines are not quite as reluctant to discuss the issue of overbooking, and their own figures loudly speak of the extent of this practice. According to the Air Transport of America, the airlines’ trade association, the industry-wide overbooking rate ranges between 10 to 15 and 20 to 25 percent in peak periods.

In most cases airline overbooking constitutes a deliberate policy designed to cope with the “no-show” problem. The problem is rather serious. By some accounts, on average, as many as 20 percent of reservation holders fail to show up for flights they had booked; with one route registering a staggering 50 to 70 percent no show rate. Under the circumstances, the airlines find that they have no other alternative but to resort to overbooking in order to offset potential losses associated with operating their flights at less than full capacity.

The practice of overbooking is by no means limited to commercial aviation. Although statistical data on the subject are not readily available, there is little doubt that overbooking occurs, among other places, in the hotel and car rental industries.

It is surprising then that, in view of this problem’s apparent relevance, so little has been written about overbooking. Recently Yuri Aroenber (1989), in an attempt to rectify this shortcoming, constructed a model of a risk neutral prize-setting firm facing a service capacity constraint and the possibility that some of its reservation holders may not show up to claim the service. He examined the case (which one may encounter in the hotel industry) where the fee for the firm’s service is due at the time when this service is rendered, demonstrating that the number of reservations which will be accepted is always positively related to capacity. In contrast, this paper deals with the firm whose fee (as is typically the case in the airline industry) is due at the time when reservations are made. Among other things, it is shown that, under these circumstances, a larger capacity does not necessarily warrant a greater number of reservations.

This paper is organized as follows. The next section sets forth a model of overbooking. The sections that follow are concerned with the determination of the optimal number of reservations and the relationship between the parameters of the model and the number of reservations. The last section is reserved for concluding remarks.

*Department of Economics, Brooklyn College of CUNY, Brooklyn, New York 11210
The author would like to thank Paul Goldberg, Robert J. Levison, the anonymous referees and, especially, Frank Bassam for their valuable comments.

THE MODEL

Consider a firm whose service capacity, \( S \ (S > 0) \), is fixed. Reservations for this service are made in advance. Each reservation entitles its holder to one unit of the firm’s service capacity at some future date. Suppose that the firm faces a downward-sloping demand curve for its service and knows with certainty how many reservations (units of service), \( R \), will be made (demanded) by its customers at each per unit price, \( P(R) \), it chooses to specify. Assume that \( S \) is consistent with a point on the elastic portion of this curve and that up to \( S \) the cost incurred by the firm is serving an additional customer is zero. Under these circumstances, had the firm known with certainty that all of its reservation holders will show up to claim the service, it would have accepted \( S \) reservations (and, thus, would have set a price of \( P(S) \)). Suppose, however, that for any \( R \) accepted by the firm, the number of customers who will show up to claim the service, \( 0 \leq N \leq R \), is unknown. Let \( N \) be viewed by the firm as a random variable whose density function is \( f(N) \). Clearly, \( EN \) is nonincreasing in \( R \).

The fee for the firm’s service is due at the time when reservations are made. Suppose that, if a reservation holder fails to show up to claim the service, he is entitled to a refund \( c(R) \), where \( 0 \leq c \leq 1 \). In addition, assume that, in the event the firm chooses to overbook and the number of reservation holders who show up exceeds the available capacity, the firm is required to compensate each customer whom it cannot accommodate by paying him a penalty of \( c(R) \), where \( c > 1 \).

The firm is in risk neutral. Its expected profit function, \( E[R] \), is given by

\[
E[R] = P(R) - cP(R)\int_S^R (R-N)f(N)dN = cP(R)\int_0^S (S-N)f(N)dN - F,
\]

where the first term on the right-hand side of (1) is the total revenue collected by the firm at the time reservations are accepted; the second term is the volume of expected refunds; the third term is the volume of expected compensation to customers whom the firm would not be able to accommodate in the event \( N > S \) and \( F \) is the fixed cost of providing the service. The firm would like to accept that number of reservations which would maximize its expected profits.

Differentiating (1) with respect to \( R \) and using Leibniz’s rule, yields the following first-order condition for maximization:

\[
P'(R) + P'(R)\int_0^S (S-N)f(N)dN - cP'(R)\int_0^S (N-S)f(N)dN = 0,
\]

where

\[
P'(R) = \frac{R}{S} - \int_0^S Nf(N)dN + P'(R)\int_0^S Nf(N)dN
\]

and

\[
P'(R) = \frac{R-S}{S} - \int_0^S (N-S)f(N)dN + P'(R)\int_0^S (N-S)f(N)dN.
\]

According to (2), the firm should accept that number of reservations at which the marginal revenue becomes equal to the expected marginal cost. Observe that the decision whether or not to overbook depends on the properties of the firm’s demand function and the reservation holders’ show up density function. If overbooking is not warranted, expression (2) reduces to

\[
P'(R) + P'(R)\int_0^S (S-N)f(N)dN - cP'(R)\int_0^S (N-S)f(N)dN = 0.
\]

In Appendix I it is shown that this equation cannot hold if \( E[N/R] \) is nondecreasing in \( R \). Thus, if an optimal reservation policy exists, a sufficient condition for overbooking is that the proportion of reservation holders who are expected to show up be nondecreasing in \( R \).

Let \( R^* \) be the optimal number of reservations. To demonstrate that the expected profit function is nondecreasing in \( S \), insert \( R^* \) into (1) and differentiate the obtained identity with respect to \( S \). In view of
(2), the resulting expression reduces to

\[
\frac{dE(M)}{dS} = c_2(R) \int_0^\infty f(N) dN = 0.
\]

Expression (3) asserts that, at the optimum, the benefits derived by the firm from an additional unit of its capacity must be equal to the level of compensation for denying service to a reservation holder times the probability of not being able to accommodate all of its customers who made reservations (the expected marginal penalty for overbooking).

**THE OPTIMAL NUMBER OF RESERVATIONS**

To provide further insight into the firm’s behavior, assume henceforth that \( f(N) \) is uniform. Thus,

\[
f(N) = \frac{1}{R}.
\]

Then, it is not difficult to show that the expected number of no shows is proportional to the number of reservations. It follows that the higher the price per unit of the firm’s service (i.e., the greater the opportunity cost of not claiming the service), the fewer reservation holders are expected not to show up. If it is posited that the firm’s demand function is linear, having the following form:

\[
P(R) = A - BR,
\]

where \( A > 0 \) and \( B > 0 \), the firm’s expected profit function becomes

\[
E(S) = [A - BR]R - c_2(A)BR \int_0^\infty \left[ (R-N)R \right] f(N) dN - c_3[A - BR] \int_0^\infty (N)N \left[ \frac{1}{R} \right] dN - F.
\]

Differentiating (6) with respect to \( R \), equating the resulting expression to 0, and rearranging yields

\[
A - 2BR = \frac{1}{2} c_2[A - 2BR] + \frac{1}{2} \left[A + 2BS - \frac{S^2}{R^2} - 2BR\right].
\]

The second-order condition for maximization requires that

\[
H = c_2A \frac{S^2}{R^2} + \frac{1}{2} (1 + c_3)[R] > 0.
\]

The expression on the left-hand side of (7) is, of course, the marginal revenue, MR, and that on its right-hand side is the expected marginal cost schedule, E(MC), whose the properties may now be examined. Observe first that when \( R \leq S \), E(MC) reduces to \( (1/2)c_2(A - 2BR) \). In other words, as long as the firm does not overbook, it had assured itself of being able to accommodate every customer irrespective of how many show up; and, consequently, its sole cost of accepting one more reservation is the expected amount of additional refunds for those customers who might fail to show up.

To ascertain the shape and curvature of the expected marginal cost curve, differentiate E(MC) with respect to \( R \) to obtain:

\[
\frac{dE(MC)}{dR} = -c_2R + c_2A - c_2B,
\]

\[
\frac{d^2E(MC)}{dR^2} = -2c_2R - c_2A - c_2B.
\]

Thus, the expected marginal cost curve is composed of a linear and a strictly concave segments. As the number of reservations increases, it slopes downward initially, then upward, but it eventually turns downward again when \( R \) becomes sufficiently large. These results can be justified as follows. If \( R \leq S \), the expected marginal cost depends only on two factors: the expected number of reservation holders who might not show up and the amount of refund (which is proportional to the price per unit of service). As the firm accepts one additional reservation (without resorting to overbooking), the expected number of no shows increases by some fraction but the price falls even by a greater proportion; and, thus, E(MC) decreases. When \( R > S \), in addition to the two factors mentioned above, E(MC) is affected by the expected number of customers whom the firm might not be able to accommodate and the penalty (which is proportional to the price per unit of service). As the number of reservations (in excess of the firm’s service capacity) increases, so does the expected number of reservation holders whom the firm might have to deny service; however, the price per unit of service continues to decrease. Now, when the number of reservations slightly exceeds the service capacity, although the price per unit of the firm’s service is relatively high, the expected number of customers who would be "bumped" is extremely low. Thus, under these circumstances, the expected cost of an additional reservation is quite low but increases with the number of reservations. Eventually, however, when the number of reservations is significantly greater than the firm’s service capacity, the expected cost of accepting one more reservation will begin to decrease since (even though the expected number of reservation holders whom the firm would not be able to accommodate and those who might fail to show up is quite large) the price per unit of the firm’s service is very low.

To determine the optimal number of reservations, reconsider equation (7). It reveals that, since \( S < \sqrt{A/c_2} \) as long as \( R < S \), the MR curve is above the E(MC) curve, whereas, at \( R = \sqrt{A/c_2} \) the MR curve is below the E(MC) curve. It follows that the optimal number of reservations, \( R^* \), consistent with the point at which these curves must intersect is such that

\[
S < R^* < A/c_2.
\]

It can be established that \( R^* \) is the only positive root to equation (7).

Expression (10) asserts that, faced with the no show problem, notwithstanding that it has been compensated in advance and reservations in excess of service capacity may entail penalties, the firm above its break-even point. In addition, (10) shows that \( R^* \) is consistent with a point on the elastic portion of its demand curve. Clearly, the firm would not operate on the inelastic portion of its demand curve because along this portion of the curve it is always possible, by raising the price, to augment the total revenue and to reduce the expected cost.

The determination of the optimal number of reservations is illustrated in Figure 1. The intersection of the expected marginal cost curve and the marginal revenue curve occurs at point E. The firm, therefore, accepts the number of reservations, \( R^* \), consistent with this point, thus, overbooking by \( R^* - S \) units and charges its customers the price of \( P(R^*) \) per unit of its service. It should be noted that at the point of intersection the slope of the E(MC) curve can be positive (as shown in Figure 1) or negative (zero). In the next section it will be shown that this issue is of particular importance.

**STRATEGIC OVERBOOKING**

To examine the effects of changes in the refundable portion of the fee and the penalty rate for denying service on the number of reservations, differentiate (7) with respect to \( c_2 \) and \( c_2 \) to obtain:

\[
\frac{dR^*}{dc_2} = -\frac{1}{2}\left[ A - 2BS - \frac{S^2}{R^2} - 2BR\right] < 0
\]

and

\[
\frac{dR^*}{dc_2} = -\frac{1}{2}\left( A + 2BS - \frac{S^2}{R^2} - 2BR\right) < 0.
\]
While inspecting the relevant (R > S) statement of (9), one infers that, if (15) holds, the slope of the E(MC) curve is negative. It follows that a necessary (but not sufficient) condition for an increase in $S$ to lead to a reduction in the number of reservations is that the slope of the expected marginal cost curve at the point of its intersection with the marginal revenue curve be negative. It is useful to investigate under what circumstances this slope is negative.

Equating $[E(MC)]/dR$ (associated with $R > S$) of (9) to 0, one observes that the E(MC) curve consistent with some reservation $S$ reaches its maximum point when the number of reservations, $R^*$, is such that

$$
R^* = \frac{c_A}{c_s + c_B} S^0.
$$

Inserting $R^*$ into $E(MC)$ and $MR$ and subtracting the former from the latter, yields

$$
V = MR(R^*) - E(MC)(R^*) = \frac{1}{2} [12 - (c_s + c_B)] A - c_B S^0
$$

Now, if $V > 0$, then up to $R^*$ the expected marginal cost curve consistent with this $S$ must lie below the marginal revenue curve. I.e., if $V > 0$, the point where intersection occurs must be located on the downward-sloping portion of the E(MC) curve. While inspecting (17), one observes that $V > 0$ if the sum of $c_s$ and $c_B$ is about 2 or smaller and $S$ is relatively low. If, on the other hand, this sum is somewhat greater than 2, $V < 0$ for all values of $S$.

The preceding analysis warrants the conclusion that a small increase in capacity may lead to a reduction in the number of reservations if the refundable portion of the fee to a reservation holder who fails to show up and the penalty rate for denying service are moderate and the original level of capacity is low. How can this result be justified? If the refundable portion of the fee and the penalty rate are not substantial and its capacity is very low, were the firm to accept only a small number of reservations, the benefit of a relatively moderate expected number of customers who might be entitled to a refund or compensation would be overshadowed by the detriment of foregoing considerable revenue associated with the "unplugged" elastic portion of its demand curve. Under the circumstances, the firm elects to accept a number of reservations which is well in excess of its capacity and, thus, expects not to be able to accommodate some of its customers. In the range of low capacity levels, however, as capacity gradually increases, the expected cost of having to refund (some of) the money and to deny service to its reservation holders begins to outweigh the revenue gains, and the firm finds that it is more advantageous to reduce the number of reservations.

The relationship between the firm’s service capacity and the optimal number of reservations can now be summarized. If the refundable portion of the fee and the penalty rate are moderate, starting an extremely low capacity level, the number of reservations decreases initially as capacity increases.

### TABLE 1

<table>
<thead>
<tr>
<th>Case</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service capacity</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Number of reservations*</td>
<td>39.5</td>
<td>35.1</td>
<td>32.2</td>
<td>30.8</td>
<td>39.0</td>
<td>41.2</td>
<td>43.5</td>
<td>45.7</td>
<td>47.0</td>
</tr>
</tbody>
</table>

*Rounded-off to the nearest one-tenth.
TABLE 2
Service Capacity and the Optimal Number of Reservations for $P(R) = 100 - R$, $c_i = 0.50$, and $c_i = 2.50$

<table>
<thead>
<tr>
<th>Case</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service capacity</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Number of reservations*</td>
<td>7.5</td>
<td>14.4</td>
<td>30.3</td>
<td>26.0</td>
<td>30.9</td>
<td>35.4</td>
<td>39.6</td>
<td>43.3</td>
<td>46.8</td>
</tr>
</tbody>
</table>

*Rounded off to the nearest one-tenth.

Eventually, however, when capacity is sufficiently high, further increases warrant a larger number of reservations. If, on the other hand, the refundable portion of the fee and the penalty rate for bumping a reservation holder are substantial, it is never sensible to underbook by a very large margin. Under these circumstances, the number of reservations the firm chooses to accept is always positively related to its capacity level. Tables 1 and 2, which report the results of numerical simulation for the cases where the refundable portion of the fee and the penalty rate are moderate and substantial, respectively, confirm these conclusions.

The case of an increase in capacity which, in the presence of relatively small $c_i$ and $c_{i+1}$, leads to a reduction in the number of reservations is illustrated in Figure 2. There, given two capacity levels, $S_i$ and $S_{i+1}$, the smaller one, $S_i$, calls for a greater number of reservations, $R^*_i$, and, thus, a lower price, $P(R^*_i)$, and a much greater overbooking margin, $R^*_i S_i$, than the larger one whose respective configurations are $R^*_{i+1}$, $P(R^*_{i+1})$, and $R^*_i S_{i+1}$.

CONCLUDING REMARKS
This paper examined the reservations policy of a risk-averse price-setting firm facing a service capacity constraint and the possibility that some of its reservations holders may not show up to claim the service. It was concerned with the case where the fee for the firm's service is due at the time when reservations are made and is partially refundable to those customers who fail to show up. It was shown that, even though the firm has been compensated in advance and there are potential penalties associated with accepting reservations in excess of service capacity, the firm always has an incentive to overbook. In addition, it was demonstrated that an increase in the refundable portion of the fee or in the penalty rate for failing to accommodate a reservation holder always results in a reduction in the number of reservations and the larger the service capacity, the smaller the rate of overbooking. More important, the analysis revealed that, if the refundable portion of the fee and the penalty rate for bumping a reservation holder are substantial, an increase in capacity always warrants a greater number of reservations; however, interestingly enough, if the refundable portion of the fee and the penalty rate are moderate, such an increase does not necessarily lead to an augmentation in the number of reservations. It is hoped that this paper will attract further interest in the theory of overbooking; interest which, in view of this paper's proliferation, appears to be long overdue.

APPENDIX 1
Observe that

$$P(R) + P'(R)|R + c_2 (P(R) + P'|R) + \int_{R}^{R^*} N(R)|N| \ dN| = P(R) + \int_{R}^{R^*} N(R)|N| \ dN|$$

$$= (1 - c_2) \{P(R) + P'(R)|R + c_2 \{P(R) + P'(R)|R + \int_{R}^{R^*} N(R)|N| \ dN| - (1 - c_2) \{P(R) + P'(R)|R$$

$$+ c_2 \{P(R) + P'(R)|R - \frac{\partial E[N]}{\partial R} - P'(R)|R^{1/2} \frac{\partial E[N|R]}{\partial R} \} \}$$

where

$$\frac{\partial E[N]}{\partial R} = \frac{d}{\partial R} \int_{R}^{R^*} N(R|R) \ dN| = \int_{R}^{R^*} N(R|R) \ dN|$$

and

$$\frac{\partial E[N|R]}{\partial R} = \frac{d}{\partial R} \int_{R}^{R^*} N(R|R) \ dN| = \frac{1}{R} \{R(R + \int_{R}^{R^*} N(R|R) \ dN| - \int_{R}^{R^*} N(R|R) \ dN| \}$$

Since $S$ is consistent with a point located on the elastic portion of the demand curve and the firm does not overbook, $P(R) + P'(R|R) > 0$. Then, if $\frac{\partial E[N]}{\partial R} > 0$ (and, thus, $\frac{\partial E[N|R]}{\partial R} > 0$), one must conclude that expression (A1.1) is positive.
APPENDIX 2

Expression (13) is positive if

\[ A > B \text{ in (A.2)} \]

and, consequently, if

\[ 2 - c_i < \frac{[2 - c_i] R^*}{R^* - S} \]  \( \text{(A.2.1)} \)

To show that (A.2.2) must hold, assume the contrary, i.e.,

\[ 2 - c_i > \frac{[2 - c_i] R^*}{R^* - S} \]  \( \text{(A.2.3)} \)

Then, in accordance with (7), one must conclude that

\[ A + 2 B S - A \frac{R^*}{R^*} - \frac{[2 - c_i] R^*}{R^* - S} \]
\[ \geq 2 - c_i \frac{(A - 2 B R^*)}{R^* - S} \]
\[ = A - 2 B R^* + \frac{1}{2} (2 - c_i) A \frac{S}{R^*} \]  \( \text{(A.2.4)} \)

which is a contradiction. This completes the proof.

NOTES

4. Since the advent of all-line deregulation, several arrangements—from the nonrefundable ticket plan (i.e., \( c_i = 9 \)) to its fully refundable counterpart (i.e., \( c_i = 0 \)—have been available to travelers. For a discussion of the consequences of airline deregulation see Moor (1968) and Winston and Mitter (1987).
5. In the interest of expository simplicity, the possibility that, in case the service capacity exceeds the number of reservation holders who showed up, the firm can offer the unused capacity to standby customers at a discount will be disregarded.
6. It can be shown that, if the fee is nonrefundable, the firm would always overbook.

REFERENCES


INTRODUCTION

Economists have long advocated price discrimination as an efficient means of making higher education accessible to all social and economic classes. That is, they argue that postsecondary schools should charge higher tuition to those groups with a relatively low price elasticity of demand and target student aid to those groups whose demand is more elastic. The alternative policy of keeping tuition or net price low across the board (a practice followed by most state systems of higher education) is less efficient since it provides a subsidy to students who would have attended school without any assistance. The success of this approach relies on the ability to identify groups whose enrollment decisions are price sensitive from those whose enrollment decisions are price insensitive. Numerous empirical studies have shown that low-income students are more responsive to costs than high-income students.7 line with these findings most federally funded student aid programs are targeted toward students from families at the lower end of the income scale. Most college-administered scholarship programs also tend to allocate a larger fraction of both internal private funds and external public funds to low-income groups (Blakemore and Low 1983a; Cates and Robinson 1982, Ventry 1983). There is little professional and public consensus, however, on how financial aid awards should be packaged. Over the past two decades there has been a dramatic shift in the composition of federal aid away from grants toward subsidized loans. A study prepared for the Joint Economic Committee of Congress (Hausen 1986) reported that borrowing under federally subsidized programs increased by nearly five-fold in a ten-year period from $2 billion in 1975-1976 to $9.8 billion ten years later. This change represented, in part, an increase in the number of students eligible for these loans and, in part, a shift in federal spending away from grants toward loan subsidies. Over the same period federal expenditures on grants fell from 6.8 billion to $4.8 billion. The study also found that much of the increased borrowing was undertaken by students from low-income families. In 1983, 40 percent of all Guaranteed Student Loan borrowers came from families earning below $15,000, well below the median family income of $24,500.

This growing reliance on loans has raised public concern over decreased access to higher education among students from low-income families as well as among female and minority students. Deborah J. Carter, a co-author of a recent study on trends in minority enrollments (Wilson and Carter 1988) has suggested that the decline in the college participation rates of black male high school graduates could be attributed, in part, to "a shift in Federal financial aid from mostly grants to mostly loans." (New York Times 1989, p. A37). As loans represent a claim on uncertain future income they may be viewed as a significantly less desirable form of aid for all students but especially for students from families who lack the financial resources to assist with the repayment. Women and minorities who, on average, face lower

7 I would like to thank Edward Cavin, Jonathan Sandy, and two anonymous referees for their helpful comments and suggestions on an earlier draft.