Estimating Air-Cargo Overbooking Based on a Discrete Show-Up-Rate Distribution

Andreea Popescu, Pinar Keskinocak, Ellis Johnson
School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive,
Atlanta, Georgia 30332 {andreeap@isye.gatech.edu, pinar@isye.gatech.edu, ellis.johnson@isye.gatech.edu}

Mariana LaDue, Raja Kasilingam
Sabre Airline Solutions, 1 East Kirkwood Boulevard, Southlake, Texas 76092 {mariana.ladue@sabre.com, raja.kasilingam@sabre.com}

Most airlines overbook their actual capacity (for both passengers and cargo) because part of the booked demand often does not show up at the flight departure. A key element of overbooking is a model that accurately predicts the show-up rate of the current bookings. Given the increasing importance of cargo within their business, most major airlines now scrutinize estimates of show-up rates for cargo bookings. The current practice is to apply the same methodology to the cargo sector as in the passenger business. We investigate the suitability of the current practice, and based on the results, we propose an alternate show-up-rate estimator for cargo and demonstrate its benefits. Tested on real-world data from a major airline as well as on simulated data, the study shows that improved estimation of the show-up rate can improve profits and customer service.

Key words: industries: transportation, shipping; inventory: production.

Histories: This paper was refereed.

With globalization of trade, increasing use of advanced logistics techniques, and the rise of e-commerce, companies increasingly ship freight by air (Air Cargo Management Group 2000). Airline industry forecasts predict that the world air-cargo traffic will expand at an average annual rate of 6.2 percent for the next two decades, tripling current traffic levels (Boeing 2004).

The major players in the air-cargo supply chain are the shippers, the freight forwarders (FFs), and the airlines. Six to 12 months before flight departures, FFs bid for the cargo space airlines offer to accommodate expected demand from shippers; the cargo capacity committed during the auction process is called allotted capacity. The remaining capacity available for free sale, which we refer to as cargo capacity, is open for sale one month before the aircraft’s departure.

Given the potential revenue from air transportation of freight, airlines must manage cargo capacity effectively. Cargo capacity is perishable, can be sold at different prices (depending on the service, for example, express or normal), and is limited. Following their long tradition of using revenue management in the passenger business, airlines naturally seek to adapt the same techniques to cargo. However, the cargo business and the passenger business differ in important ways.

The airlines do not know how much capacity they have available for free sale until the flight departure. FFs intentionally bid on more capacity than they actually need to ensure space on constrained flights because most airlines allow them to return unwanted space at no extra charge. The airlines add the released space to the pool of capacity available for free sale. They typically do not know how much allotted capacity will be unused in advance of the flight departure. In addition, for planes carrying cargo and passengers (combination carriers), the cargo space usually contains both passengers’ baggage and cargo in the same compartment. These factors plus weather (which affects the amount of fuel on board the aircraft) and mail influence how much capacity is available for free sale (Figure 1). Finally, the cargo space is constrained by two dimensions, weight and volume, and the airline typically does not know which dimension is the most restrictive prior to departure.
Popescu et al.: Estimating Air-Cargo Overbooking Based on a Discrete Show-Up-Rate Distribution
Interfaces 36(3), pp. 248–258, © 2006 INFORMS

<table>
<thead>
<tr>
<th>Passengers’ bags, mail, extra fuel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allotments</td>
</tr>
<tr>
<td>Total cargo capacity</td>
</tr>
<tr>
<td>Free sale</td>
</tr>
</tbody>
</table>

Figure 1: The combination passenger-cargo aircraft’s cargo capacity is shared among passengers’ bags, mail, fuel, allotments, and capacity available for free sale.

The booking processes for cargo and passengers are different. The time window during which the airline offers cargo capacity for free sale is shorter than that for passenger capacity; usually no longer than 30 days before departure. Cargo bookings, varying widely in size and volume, come from a fairly small number of customers. A booking may be canceled, rebooked to a different flight, canceled again, rebooked back to the original flight, several times until departure, because airlines typically do not charge for changing reservations.

To hedge against the variability in the amount of cargo actually handed at departure (cargo tendered) and customers’ cancellations, airlines commonly overbook their capacity. Air-cargo overbooking refers to the airlines’ practices to sell more capacity than physically available to compensate for cargo that does not show up at departure. Two important considerations in overbooking are spoilage (demand turned away because the overbooking level was too low, leaving excess capacity at departure) and off-loads (booked demand that the airline cannot accommodate at departure because the overbooking level was too high). Airlines base their decisions on predictions of the show-up rate, the percentage of the demand booked that shows up at departure.

In the passenger sector, the common practice is to formulate the overbooking problem as a newsvendor problem, with the overbooking level selected to minimize the total expected costs of spoilage and off-loads (Weatherford and Bodily 1992). To model the cargo show-up rate, many airlines use the normal distribution, which is a good approximation for passengers (Belobaba 1987). We show that the normal distribution is usually not a good fit for estimating the cargo show-up rate.

The Cargo-Booking Process

The calculation of the overbooking levels is based on show-up-rate estimates. The cargo weight or volume show-up rate is the percentage of cargo weight or volume that shows up at departure out of the total weight or volume of cargo booked at each reading day (RD). For cargo, the booking time window has 30 reading days, which are numbered backwards in time, from 0 (the departure day) to 30. For example, the show-up rate (in percentage) on reading day 21 before departure date $x$ is the amount tendered at departure day $x$ (in kilograms) out of the amount booked on reading day 21 (in kilograms), multiplied by 100.

Following the practice they use for the passenger business, most airlines estimate the cargo show-up rate separately for weight and volume as a normally distributed random variable at the flight-leg level (flight number and origin and destination airports). They feed the estimates to the overbooking module, which sets the level of capacity authorized for sale. In each reading day, the airline accepts demand if capacity remains after subtracting the current bookings (accepted bookings that have not been cancelled) from the authorized capacity (the capacity available for free sale multiplied by the overbooking level). The show-up rate changes from reading day to reading day, and the airlines have to make sure they capture these changes and use correct levels for overbooking when selling cargo space. Usually, they do not monitor the booking process over the entire booking period but rather only on specific reading days, which they consider as significant based on historical booking activity.

During the summer of 2003, Georgia Institute of Technology and Sabre Holdings teamed up to investigate the efficiency of the air-cargo overbooking model used by some major airlines. The project’s objective was to determine the appropriateness of using a normal distribution for estimating the air-cargo show-up rate distribution, to find a better fit if necessary, and to
study its impact on the overbooking levels. The first challenge we had to overcome was the data collection. The cargo business is a fairly new candidate for providing additional revenue for the airline industry, and there is still a lot of manual handling involved; the orders come in through different channels (agencies, Internet, freight forwarders), and most of them are not properly captured in the airlines’ systems.

Because of the nature of the data the airline collected, we focused on estimating the weight show-up rate only. Luo et al. (2005) justified the use of a common distribution for the show-up rate for weight and volume by conducting statistical tests on real-world data.

We reevaluated the level of the authorized capacity over the following pre-specified 15 reading days: RD30, RD28, RD21, RD14, RD10, RD9, . . ., RD1, and RD0.

**The Data**

For seven combination passenger and cargo flights, we have show-up rate data for a 16-month period. For each flight number and departure date, we have 15 show-up rates, corresponding to each reading day. We used the first 12 months of the data to estimate the show-up-rate distribution, two months of data to forecast the fitted distribution, and the last two months to validate and compare our results with the normal distribution. Hence, we used 80 percent of the data for training and the remaining 20 percent for testing, a common practice among marketing and neural networks researchers.

**Nonparametric Distribution Estimation and Forecasting**

For the flights we analyzed, we observed that the distribution of the show-up rate follows different shapes and skewness. We fitted the following continuous parametric distributions to all analyzed flights in the study for all reading days for one year of historical departures: normal, gamma, beta, Weibull, lognormal, and exponential. We used three goodness-of-fit tests for each distribution: Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling. The hypothesis that the sample was from any of the specified distributions was rejected in approximately 90 percent of all cases.

These results motivated us to fit a nonparametric distribution to the show-up rate. If the probability distribution function were from a known parametric family (for example, Gaussian), we would have to estimate the finite-dimensional parameters that characterize that particular distribution (for example, the mean and the variance). Without the parametric assumption, the problem is known in statistics as the nonparametric estimation problem.

One of the easiest nonparametric estimators is the histogram, which is obtained by dividing the data range into equal intervals (bins) and counting the number of observations that fall into each interval (Figure 2). Each bin is represented by the midpoint of the corresponding interval, with probability equal to the ratio of the number of observations it contains and the total number of observations.

The effectiveness of the histogram estimator depends on the number and size of the bins. In general, the higher the number of bins, the higher the probability of capturing noise rather than the characteristics of data. On the other hand, a too low number of bins will represent the data poorly. We first developed an estimator with equal-size bins. To choose the optimal number of bins to minimize the error of mapping the data into the class intervals (bins), we implemented Birge and Rozenholc’s method (2002). This equal-size-bin estimator performed slightly better than the normal estimator in terms of mean off-loads and spoilage. The main disadvantage was that it used equal-size bins, which led to bins containing no points for some of the reading days, where the show-up rates were clustered together around some values.

To capture the data distribution more accurately, we next used a histogram estimator with variable size bins, which uses the equal-size-bin estimator as a starting point.

We use wavelet methods (Vidakovic 1999) to denoise the signal defined by the bin frequencies of the equal-size-bin estimator. The idea is that the wavelet coefficients correspond to the details of the signal. The method considers the small details to be noise and deletes or smoothes them out without substantially affecting the main features of the original signal. After denoising, the signal may comprise several adjacent bins with the same number of data values,
Comparing the Discrete Distribution and the Normal Distribution

Using percentiles, we compared the proposed discrete distribution and the normal distribution used by the airline. Percentiles are position measures, describing where a specific data value falls within the data set or the distribution range. We computed nine percentiles (10, 20, …, to 90), using the statistical tool SAS. While we could compute the percentiles directly for continuous distributions (for example, normal), for the discrete distribution we used the value of the midpoint of each interval, for which the cumulative distribution function is closest to the considered percentile.

The mean absolute error for the discrete distribution ($MAE_{\text{discrete}}$) is consistently between 10 and 50 percent lower than the mean absolute error for the normal distribution ($MAE_{\text{normal}}$) for each reading day (Figure 3). The results encouraged us to proceed with studying the impact of the new estimator on overbooking.
The overbooking model some major airlines use is a newsvendor problem (Winston 1993) with service-level constraints and upper and lower bounds for the authorized capacity. The service level, or failure rate, is defined as the ratio between the expected value of the off-loads and the expected value of the show-ups. The airlines impose the failure rate constraint to discourage too high overbooking levels, so that they can meet the service levels promised to the customers (appendix).

**Impact of the Show-Up-Rate Estimation on Costs/Profits**

We compare the tendered amount of cargo at departure with the real capacity available on reading day zero (departure day). The closer the tendered amount was to the real capacity, the better the overbooking policy.

Two factors determine the tendered amount of cargo at departure:

1. The overbooking levels per reading day, which are directly related to the show-up-rate estimators; and
2. The estimate of the capacity available for free sale.

To compare the influence of the show-up-rate estimators on profits, we simulated the cargo-booking process, which can be summarized as follows. In each reading day,

1. We calculate current bookings based on previous bookings and the cancellation rate,
2. We calculate overbooking levels and hence the authorized capacity,
3. Demand arrives,
4. We accept demand according to the space available after subtracting current bookings from the authorized capacity, and
5. We update current bookings to take account of the newly accepted demand.

We considered two demand scenarios. In the first scenario, we modeled the demand arrivals as normally distributed random variables. For cancellations, we used two random variables: a uniformly distributed random variable to model the probability of cancellations occurring on a certain day, and a normally distributed random variable to model the magnitude of cancellations.

In the second scenario, we used real-world demand data. Only truncated demand data was available, however, because most companies do not record lost sales. By truncated demand, we mean the demand the airlines satisfied, not including the demand lost because of insufficient capacity. Hence, the truncated demand is a lower bound on the actual demand. Although not equal to the real demand, the truncated demand captures the dynamics of the booking process, that is, cancellations and still periods.

Truncated, or censored, data is common in the airlines’ passenger business. Weatherford and Pölt (2002) analyzed six methods used to uncensor passenger demand data: three so-called naïve methods, and three more sophisticated methods. These methods work on data that contain an indicator as to whether a particular fare class was open or closed to bookings at the specified time. The three naïve methods are

—(N1) To use all data and ignore whether bookings were open or closed;
—(N2) To use only open observations and toss out the closed ones; and
—(N3) To replace closed observations with the larger of the following: the actual observations or the average of the open observations.

The method we used to uncensor the data is close to (N3): for the days the capacity was completely utilized, we added a normally distributed random variable with a probability of 0.5, because we do not know which observations were open and which
were closed. Although other methods to uncensor data exist, we would have had to test them empirically. (N3) is a reasonable trade-off between complexity and performance, as Weatherford and Pölt (2002) pointed out.

To calculate the overbooking levels in both scenarios, we used a failure rate of 10 percent, a lower and upper bound of 100 percent and 200 percent of the physical capacity, respectively, and a ratio of 4 to 1 for spoilage and off-loads costs. In the air-cargo industry, spoilage is more costly than off-loads. At departure, airlines generally have a good mix of general and time-sensitive cargo. When there is less cargo than capacity at departure, the aircraft flies partially empty, which translates into lost opportunity. Most airlines have a cost ratio of 1 to 3 or 4 for off-loads versus spoilage.

We implemented the overbooking policy used by several major airlines (appendix), using the normal and the discrete show-up-rate estimators. This resulted in two different authorized capacity levels and, consequently, in two different streams of accepted demand, that is, current bookings per RD, for each simulation run. We called the current bookings resulting from the normal and discrete estimators 

$$T_{\text{normal RD}} = CB_{\text{normal RD}} \times SR_{\text{actual RD}}$$

and

$$T_{\text{discrete RD}} = CB_{\text{discrete RD}} \times SR_{\text{actual RD}}$$

We compared the tendered cargo ($T_{\text{normal RD}}$ and $T_{\text{discrete RD}}$) with an ideal solution and with the real capacity at departure. We obtained an ideal solution from the deterministic version of the process: if we knew all the demand that would show up in advance, then we would accept demand per reading day up to the estimated capacity at departure. The tendered demand in this case, $T_{\text{ideal RD}}$, is the accepted demand per reading day.

For the normally distributed demand scenario, we conducted experiments for all combinations of low, medium, and high mean demand as a percentage of the capacity and coefficient of variation (standard deviation over mean).

We ran 500 simulations for each of the nine experiment settings (Figure 4) and obtained similar results. For all instances, we have the following results:

1. In a comparison with the ideal solution
   - The mean absolute error of $T_{\text{normal}}$ (compared to $T_{\text{ideal}}$) at departure was approximately seven percent higher than the mean absolute error of $T_{\text{discrete}}$.
   - The standard deviation of the error was approximately two percent higher for $T_{\text{normal}}$ than for $T_{\text{discrete}}$.

2. In a comparison with the real capacity
   - The mean absolute error of $T_{\text{normal}}$ (compared to the real capacity at departure) was approximately four percent higher than the mean absolute error of $T_{\text{discrete}}$.
   - The standard deviation of the error was approximately one percent higher for $T_{\text{normal}}$ than for $T_{\text{discrete}}$ (compared to the real capacity at departure).

The differences between the comparisons with the real capacity and the ideal solution at departure result from the inaccuracy of the capacity estimate per reading day. Consider a simple example, in which the estimated capacity for any given reading day $j > 0$ exceeds the real capacity at departure and the demand is greater than the estimated capacity in any given reading day. In this case, even if we do not overbook and accept as much demand as the estimated capacity, we still end up with demand that cannot be accommodated at departure.

We reran the simulations assuming perfect forecast of cargo capacity at departure and found that the

<table>
<thead>
<tr>
<th>Mean demand</th>
<th>low</th>
<th>medium</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand CV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>60%</td>
<td>80%</td>
<td>95%</td>
</tr>
<tr>
<td>medium</td>
<td>60%</td>
<td>80%</td>
<td>95%</td>
</tr>
<tr>
<td>high</td>
<td>60%</td>
<td>80%</td>
<td>95%</td>
</tr>
</tbody>
</table>

Figure 4: In our demand scenarios, we vary the mean and the coefficient of variation (CV) of demand from low to high. For example, the medium-high cell represents instances where the mean demand is 95 percent of the capacity and the CV is 0.4.
impact of a poor capacity forecast on the business is considerable. The mean absolute error between the tendered cargo and the real capacity at departure is on average 25 percent higher and the standard deviation of the error 10 percent higher for the normal estimator for all instances.

The mean off-loads (accepted demand that cannot be accommodated at departure) are on average significantly higher (45 percent) for the normal estimator, and the normal estimator results in off-loads 10 percent more often than the discrete estimator. The discrete estimator results in spoilage about 25 percent more often than the normal estimator, but the mean spoilage is about 10 percent lower for the discrete estimator. For cargo, the total quantity of spoilage, and not the frequency, is the leading factor for costs (or lost profits). Hence, the higher spoilage frequency does not affect the gain from its considerably lower mean.

When we ran the simulations using the altered real-world truncated demand, the results were consistently better in terms of mean absolute error, mean spoilage, and frequency. The mean absolute error and spoilage were on average 14 percent and 22 percent higher, respectively, when we used the normal estimator. The off-loads were statistically equal when the added normal variable for untruncating demand had a high mean and variance, and the discrete estimator resulted in off-loads five percent lower in mean than the normal estimator when the added normal variable had a low mean and variance.

For the real-world demand data, if we used a cost of $1.6 for unit spoilage and $0.4 for unit off-loads, typical for the South America to United States market, the average savings from using the discrete estimator for a combination carrier with 300 flights per day and an average cargo capacity per departure of 13,000 kilograms was $16,425,000. The discrete estimator resulted in significantly lower mean spoilage, that is, better utilization of capacity, and no increase in off-loads, leading to high savings in costs, increased profits, and improved customer satisfaction. Lower spoilage translates into more customers served promptly, and lower off-loads means that the airline turns down fewer customers. Hence, better utilization of the cargo capacity improves the service the airline offers to customers, which is important in the competitive market of air-cargo transportation.

We also found that forecasting capacity at departure plays an important role in cargo overbooking. If capacity estimates fluctuate over the reading period, spoilage or off-loads will occur even in the ideal setting when we know all demand in advance. Misestimation of capacity at departure results in poor utilization, which means unavoidable monetary losses because of the lost opportunity to satisfy more demand. Companies should invest in forecasting cargo capacity at departure, because without accurate forecasts, any improved overbooking procedure would fail to improve the utilization of cargo capacity.

Conclusions
The discrete estimator for the show-up rate outperforms the normal estimator in various aspects. The overbooking levels using the discrete estimator prove a better approximation of the capacity at departure in terms of mean absolute error between the tendered cargo and the real capacity at departure, standard deviation of the error, spoilage, and off-loads. For a set of real-world demand data, the average yearly savings from the discrete estimator for a combination carrier with 300 flights per day and an average cargo capacity per departure of 13,000 kilograms was $16,425,000. The discrete estimator resulted in significantly lower mean spoilage, that is, better utilization of capacity, and no increase in off-loads, leading to high savings in costs, increased profits, and improved customer satisfaction. Lower spoilage translates into more customers served promptly, and lower off-loads means that the airline turns down fewer customers. Hence, better utilization of the cargo capacity improves the service the airline offers to customers, which is important in the competitive market of air-cargo transportation.

Appendix
The Histogram Estimator with Varying Size Bins
Wavelet methods have been applied successfully to density estimation (Vidakovic 1999) because of their ability to filter out noise. Generally speaking, a wavelet basis is a collection of functions obtained as translations and dilations (shift and scale) of a scaling function $\phi$ and a mother wavelet $\psi$. Once the mother wavelet $\psi$ is fixed, dilations and translations of the function $\psi_k(x) = \text{const} \cdot \psi(2^k x - k)$, define an orthogonal basis in $L^2(R)$ (space of integrable functions) together with the scaling function $\phi$; that is, any element of the space can be represented as a linear
combination of the basis functions. Chui (1992) provided a general exposition of the wavelet theory.

We chose $\psi$ as the simplest of wavelets, the Haar wavelet, which is a step function taking values 1 and $-1$ on $[0, 1/2)$ and $(1/2, 1]$. The scaling function for the Haar wavelet is the unity function on the interval $[0, 1)$: $\phi(x) = 1 \ (0 \leq x < 1)$.

In general, for a data vector $y = [y_0, y_1, \ldots, y_{2^n-1}]$ of length $2^n$ associated with a piecewise constant function $f$ on $[0, 1]$, the wavelet decomposition of $f$ has the form

$$ f(x) = c_0\phi(x) + \sum_{j=0}^{n-1} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}(x), $$

with $c_0$ and $d_{jk}$ being the wavelet coefficients.

We chose the function $f$ to be the observation count associated with the bins calculated by Birge and Rozenholc (2002); if the number of bins from the procedure is not a dyadic (power of two) number, we set it to the closest higher dyadic number. We used a quadratic variance-stabilizing transformation of the observation count per bin to improve the performance of the wavelet estimator (Vidakovic 1999).

We used the following notation:

- $N$ number of data points.
- $X_1, \ldots, X_N$ data sample.
- $D$ number of bins calculated based on Birge and Rozenholc (2002) and adjusted to the closest higher dyadic number.
- $f = [f_1, \ldots, f_D]$ observation count per bin.

The steps of the procedure are as follows:

*Step 1.* Determine $D$; if $D = 2^n + c \leq 2^{n+1}$, with $c > 0$, set $D = 2^{n+1}$.

*Step 2.* Apply the following variance-stabilizing transformation to the bin count: $\sqrt{f_i + 3/8}$.

*Step 3.* Decompose the transformed observation count $f$ via forward wavelet transform.

*Step 4.* Threshold the wavelet coefficients to filter out noise.

*Step 5.* Recover the denoised signal $f$ via inverse wavelet transform.

*Step 6.* Calculate midpoints and probabilities based on $f$.

For Step 1, see Birge and Rozenholc (2002); the method is fairly straightforward to implement. Steps 3 and 5 refer to the Haar wavelet transform; most statistical packages have it already implemented.

Step 4 is the procedure used for denoising the original signal. The wavelet coefficients correspond to the details of the signal. The method considers the small details to be noise and deletes or smoothes them out without substantially affecting the main features of the original signal. The two types of thresholding are hard and soft. Hard thresholding is the usual process of setting to zero the elements whose absolute values are lower than the threshold. Soft thresholding is an extension of hard thresholding, first setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards 0. In Step 4, soft thresholding gave us better results, and we used it in the simulations.

For the threshold value, we had several choices, among them, the universal threshold and the cross-validated threshold. We chose the universal threshold value, that is, $\lambda_{\text{UNIV}} = \sqrt{2 \cdot \ln(D)} \cdot \sigma$, with $\sigma^2$ the noise variance estimated from the coefficients’ standard deviation. The universal threshold is useful for obtaining a starting value when nothing is known of the signal condition.

We assumed a nonwhite noise in our signal (noise not having a continuous and uniform frequency spectrum over a specified frequency band). As a consequence, we had to rescale thresholds using a level-dependent (within the wavelet decomposition) estimation of the level noise (Chui 1992).

The denoised signal $f$ is of the form $[f_1, f_1, f_1, f_2, f_3, f_3, f_3, \ldots, f_1]$. We calculated the new bins by clustering together adjacent bins (in the initial histogram) of equal observation count in the denoised signal $f$.

**Updating the Fitted Distribution**

In addition to the previous notation, we used the following:

- $m_1, \ldots, m_t$ midpoints of the fitted bins.
- $p_1, \ldots, p_t$ probability vector associated with the bins.
- $y_1, \ldots, y_t$ probability vector for recent observations.
- $x_1, \ldots, x_t$ updated probability vector.
- $k$ number of new observations.

The number of bins ($t$) remains unchanged throughout the process. We constructed the probability vector $y_1, \ldots, y_t$ by counting how many new
observations fall into each bin and dividing this number by the total number of observations, \( k \).

We used the weighted least squares to compute \( x \) from \( p \) and \( y \). The optimization problem is

\[
\begin{align*}
\min & \quad \beta \cdot \sum_{i=1}^{t} (y_i - x_i)^2 + (1 - \beta) \sum_{i=1}^{t} (y_i - x_i)^2 \\
\text{s.t.} & \quad \sum_{i=1}^{t} x_i = 1, \\
& \quad x_i \geq 0 \quad \forall i = 1, \ldots, t,
\end{align*}
\]

where \( \beta \) is a weight between 0 and 1. The quadratic optimization model minimizes the weighted sum of the squared forecasted errors over all value intervals, and when used periodically (every two months in our case), it has the effect of tracking gradual changes in the probability distribution of the random variable.

The unique optimum of the convex quadratic problem (1) is

\[ x = \beta \cdot p + (1 - \beta) \cdot y. \]  

(2)

When the mean value of the random variable changes substantially, we refit the distribution before updating. Such situations could be encountered during high demand periods, such as Christmas, or when a lot of flights must be cancelled due to bad weather.

The method has two parameters that need to be carefully analyzed: \( \beta \), the smoothing factor, and \( k \), the number of new observations. We chose \( \beta = 0.8 \), based on the expert knowledge Sabre provided. Murty (2002) also recommended a value of 0.8 or 0.9 for \( \beta \): the influence of the second term in (2) should be small because the vector \( y \) is based on only a small number of observations.

Murty (2002) argued that one should use at least 50 observations to update the distribution. We used data for two months, which corresponds to approximately 60 observations.

The Overbooking Model

We sought to determine the impact of the estimate of the cargo show-up rate on the overbooking policy that airlines commonly use. Most airlines do not address the issue of multidimensionality when adapting passenger models to cargo. Usually, they run the overbooking model for weight and for volume separately.

We followed the same scheme and adapted the existing newsvendor-like overbooking model to the newly estimated (weight) show-up rate. Luo and Cakanyildirim (2005) described new approaches on cargo overbooking.

We used the following notation:

- \( \text{SR} \) discrete random variable for the show-up rate.
- \( f_{\text{SR}}(x) = P(\text{SR} = x) \) probability mass function of the show-up rate.
- \( v \) authorized capacity.
- \( \text{SU} = \text{SR} \cdot v \) random variable corresponding to the show-ups.
- \( f_{\text{SU}}(u) = P(\text{SU} = u) \) probability mass function of the show-ups.
- \( \text{SP} = \max[0, c - \text{SU}] \) random variable corresponding to spoilage.
- \( \text{OF} = \max[0, \text{SU} - c] \) random variable corresponding to off-loads.

The airlines usually define the show-ups as the authorized capacity multiplied by the show-up rate. This definition is accurate if booking requests exceed the authorized capacity. When the booking requests are below the authorized capacity, the show-ups should be equal to the booking requests multiplied by the show-up rate, that is, the show-ups should be \( \text{SR} \cdot \text{v} \). Luo and Cakanyildirim (2005) showed that the two representations of the show-ups result in the same optimal solution for a one-dimensional model.

The known parameters are

- \( c \) physical capacity.
- \( c_s, c_o \) cost per unit spoilage and off-load.
- \( r \) admissible service level.
- \( v_l, v_u \) lower and upper bound on the authorized capacity \( v \).

Based on the definition of show-ups, we deduced its probability density function \( f_{\text{SU}} \):

\[
\begin{align*}
 f_{\text{SU}}(u) &= P(\text{SU} = u) = P(\text{SR} \cdot v = u) \\
 &= P(\text{SR} = u/v) = f_{\text{SR}}(u/v).
\end{align*}
\]
The expected spoilage can be calculated as

\[ E[SP] = E[\max\{0, c - SU\}] \]

\[ = \sum_{u=0}^{c} (c - u) \cdot f_{SU}(u) = \sum_{u=0}^{c} (c - u) \cdot f_{SR} \left( \frac{u}{v} \right) \]

\[ = \sum_{x=0}^{c/v} (c - x \cdot v) \cdot f_{SR}(x). \]

Similarly, we calculated the expected off-loads as

\[ E[OF] = E[\max\{0, SU - c\}] \]

\[ = \sum_{u=c}^{+\infty} (u - c) \cdot f_{SU}(u) = \sum_{u=c}^{+\infty} (u - c) \cdot f_{SR} \left( \frac{u}{v} \right) \]

\[ = \sum_{x=c/v}^{+\infty} (x \cdot v - c) \cdot f_{SR}(x). \]

We deduced the expression for the expected total cost as a function of the authorized capacity \( v \):

\[ E[TC] = E[C_{SP}] + E[C_{SP}] \]

\[ = c_s \cdot \sum_{x=0}^{c/v} (c - x \cdot v) \cdot f_{SR}(x) + c_o \cdot \sum_{x=c/v}^{+\infty} (x \cdot v - c) \cdot f_{SR}(x). \]

We aimed to minimize the expected total cost as a function of the authorized capacity \( v \) under service level and upper and lower bound constraints for \( v \). The overbooking optimization problem is

\[ \min \ c_s \cdot \sum_{x=0}^{c/v} (c - x \cdot v) \cdot f_{SR}(x) \]

\[ + c_o \cdot \sum_{x=c/v}^{+\infty} (x \cdot v - c) \cdot f_{SR}(x) \]

\[ \text{s.t.} \sum_{x=c/v}^{+\infty} (x \cdot v - c) \cdot f_{SR}(x) \cdot \left( \frac{v}{v} \cdot E[SR] \right) \leq r, \]

\[ v_1 \leq v \leq v_u. \]

The optimization problem should be solved for each reading day separately. The optimal overbooking level with respect to the given problem is \( O^{opt} = v^{opt} / c \cdot 100 \), where \( v^{opt} \) represents the optimal solution to (3).

**Acknowledgments**

We thank the referees for their helpful and constructive comments, which greatly improved the paper. We also thank Brani Vidakovic and Laurie Garrow for fruitful discussions and guidance. Pinar Keskinocak was supported by NSF Career Award DMI-0093844.

**References**


Luo, S., M. Çakanyildirim. 2005. Overbooking models for air cargo management. School of Management, University of Texas at Dallas, Richardson, TX.

Luo, S., M. Çakanyildirim, R. Kasilingam. 2005. Two-dimensional cargo overbooking models. Technical report, School of Management, University of Texas at Dallas, Richardson, TX.


product suites offered by Sabre Airline Solutions, targets to improve the efficiency and profitability of the cargo operation. Sabre® CargoMax™ Revenue Manager focuses on managing demand and capacity along with prices to maximize total network revenue. In May 2003 we engaged on a joint project with Georgia Institute of Technology, School of Industrial and Systems Engineering, for improving the overbooking model of Sabre® CargoMax™ Revenue Manager. Show-up rate estimation is one of the key elements of the overbooking model and hence, has a direct impact on the profits, the utilization of capacity, and customer service. The better the estimation, the more accurate the overbooking levels, and consequently the higher the profits.

“Up until now, we have been using a Normal estimator for cargo show-up rates. However, we observed higher than desirable offload and spoilage rates, and decided to investigate the impact of show-up rate estimation on these outcomes. Together with our collaborators from Georgia Tech, we researched the alternative show-up rate estimation methods, and found out that a discrete estimator is more appropriate for approximating the real cargo show-up rate compared to the Normal estimator. In the light of these findings, we plan to incorporate the discrete estimation in our overbooking process. I believe that the results of this research will have a tremendous impact on the airline industry due to the high number of airlines we serve as our customers.”