Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues

Article *in* Journal of Revenue and Pricing Management · October 2002

DOI: 10.1057/palgrave.rpm.5170027

2 authors:

Larry R Weatherford
University of Wyoming

Stefan Pölt
Lufthansa

Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues

Larry R. Weatherford* **and Stefan Pölt**
Received (in revised form): 29th May, 2002

*College of Business, University of Wyoming, PO Box 3275, Laramie, WY 82071, USA
Tel: +1 307 766 3124; Fax: +1 307 766 3488; E-mail: lrw@uwyo.edu

Larry Weatherford *is Professor at the University of Wyoming. He holds a PhD from the University of Virginia. He has received several outstanding teaching awards and also has a best-selling textbook,* Decision Modeling with Microsoft Excel*, published in 2001. He has published 20 scholarly articles and presented 51 papers on five different continents to professional organisations. He has consulted with such major corporations as Walt Disney World, Hilton Hotels, American Airlines, Northwest Airlines, Lufthansa German Airlines, Swissair, Scandinavian Airlines, Air New Zealand, South African Airways, Unisys Corporation and Choice Hotels.*

Stefan Pölt *is a manager of revenue management tools at Lufthansa German Airlines. He holds a PhD in computer science from the University of Dortmund, Germany. He is recognised as an expert in forecasting and revenue management models and has given presentations at the last six AGIFORS Reservations and Yield Management Conferences.*

Journal of Revenue and Pricing Management, Vol. 1, No. 3, 2002, pp. 234–254

Abstract

KEYWORDS: yield management, inventory control, forecasting, computer simulations

Accurate forecasts of passenger demand are the heart of a successful revenue management system. The forecasts are usually based on historical booking data. These bookings do not reflect historical demand in all cases because booking requests can be rejected due to capacity constraints or booking control limits. This paper examines six different methods of unconstraining bookings to demand. Simulation analysis of many different scenarios of historical booking data are used with different percentages of the data being constrained, using simulated data, to show that the expectation maximisation and projection detruncation methods are the most robust and that, as the percentage of data constrained increases to 60–80 per cent, their estimate of the unconstrained mean increases by 20–80 per cent over the naïve unconstraining methods, which leads to less bias and more accurate forecasts. Finally, by means of actual booking data from a major US airline, it is shown that upgrading the unconstraining process can lead to revenue gains of 2–12 per cent.

INTRODUCTION

A challenge to all airlines that are active in the field of revenue management (RM) or yield management is to estimate what true demand would have been for their product, in light of the fact that all they really have a record of is the actual number of bookings. The difference between the two terms (bookings and demand) is simple — there may have been customers willing to travel on a given flight (demand) who were not able to book owing to capacity constraints or booking limits imposed on the various fare classes due the RM system's recommendation. In the field of statistics, this area is known as dealing with 'censored' data or statistics. In the airline industry, this process has commonly been called 'unconstraining' or 'detruncating'. The word 'unconstraining' will be used throughout the rest of the paper. Unfortunately, many airlines and RM vendors have implemented methods for dealing with these censored data that are largely ineffective. Ineffective means their estimates of unconstrained demand are generally too small. It is acknowledged, however, that because true demand is not observable, measures of effectiveness can be elusive when dealing with unconstraining.

The overall revenue benefits of an overall RM system have been well documented, with incremental revenue gains of 2–5 per cent being most commonly cited (Belobaba, 1992a; Smith *et al*., 1992). The focus in this paper will be to examine the various methods of unconstraining, to determine which are the most robust and effective techniques and then to quantify the revenue benefits of upgrading the unconstraining process. Other important areas of an RM system include: (1) dealing with no-show and cancellation patterns of customers through an overbooking module; (2) forecasting special events (eg World Cup soccer matches, Mardi Gras celebrations); and (3) trying to estimate the propensity of customers to upgrade from one booking class to a higher booking class. While these are important areas, they have been addressed previously in the literature.

The literature was surveyed to see whether the important topic of unconstraining had been previously addressed and not one scholarly paper had been written on the topic as it dealt with the airlines, while numerous presentations had been made to the RM Study Group of AGIFORS (Hopperstad *et al*., 1996; Pölt, 2000; Saleh, 1997; Skwarek, 1996; Weatherford, 2000; Zeni, 2001a) and one PhD thesis had been written on the topic (Zeni, 2001b). A couple of books were found from the field of statistics (Little and Rubin, 1987; McLachlan and Krishnan, 1997) that dealt with the general subject on a theoretical level. All of the AGIFORS presentations had the same theme — examine several different common unconstraining methods and compare and contrast their performance. One reference (Lee, 2000) that tried to quantify the revenue impact of improved unconstraining was also found. She found that revenues improve from 0.5 per cent to 1 per cent by switching from booking profile method to projection detruncation. This paper examines the performance obtained by the six different unconstraining methods cited in the six AGIFORS presentations and examines the impacts on the revenue performance obtained by these unconstraining methods. Actual bookings data from a major US airline company are used to determine the revenue impacts, as well as simulated data to determine the percentage improvement in estimation of the true demand. Most airlines not only store their actual bookings information, but also an indication from the reservation system as to whether a particular fare class was 'open' or 'closed' to bookings at specified checkpoints (also known as reading days, data collection points or snapshots) along the time interval from when bookings are first accepted on a given flight (eg 330 days out) to departure.

The six unconstraining methods explored are listed below.

(1) Use all data — ignore whether the bookings were 'open' or 'closed'. This is referred to as method Naïve #1 (N1).
(2) Use only 'open' observations — toss out all 'closed' ones. This is referred to as method Naïve #2 (N2).
(3) Replace 'closed' observations with the larger of (a) actual observation or (b) average of the 'open' observations. This method is known by multiple names — it is called the 'mixed' approach by Saleh (1997), the 'additive' approach by Pölt (2000) and 'imputed mean' approach by Zeni (2001a). This is referred to as method Naïve #3 (N3).
(4) Booking profile (BP) — this method compiles a picture over time (from when bookings are first accepted to departure) of how bookings build-up. The profile, however, is built only from historical 'open' data. This method is also known as the 'multiplicative' approach by Pölt (2000).
(5) Projection detruncation (PD) — a statistical approach that is explained in more detail in the section on 'Description of unconstraining methods'.
(6) Expectation maximisation (EM) — another statistical approach that is explained in more detail in the section 'Description of unconstraining methods'.

Findings are presented from a series of experiments in which these six different unconstraining methods are compared. The simulation data are limited to the study of how close/far a particular unconstraining method came from the true mean demand (which is never known in the case of actual data).

The paper continues in the next section with a review of the relevant previous literature. In the third section, the experimental design of the simulation is then reviewed, including how the simulated airline data used in the tests were generated and the methodology used to compare the estimation performance of the different unconstraining models. The fourth section gives greater detail on each of the unconstraining methods, along with numerical examples for each. The fifth section is devoted to a presentation of the simulation findings on the ability of the unconstraining methods to estimate true demand. The sixth section then describes the real airline booking data and the simulation used to quantify the revenue improvement due to using these different unconstraining methods in an actual RM system. The final section summarises the findings and makes some conclusions with respect to both the robustness of the unconstraining approaches and the relative magnitude of the revenue gains attributable to the new suggested unconstraining methods. Real-world implementation issues are also discussed.

PREVIOUS LITERATURE

Theoretical work in the statistical area of unconstraining started in the mid-1980s with the main approach being the EM algorithm (Little and Rubin, 1987; McLachlan and Krishnan, 1997). The EM algorithm is a very general iterative algorithm for maximum likelihood estimation in incomplete-data problems. Little and Rubin (1987) state that the EM algorithm formalises a relatively old *ad hoc* idea for handling missing data: (1) replace missing values by estimated values, (2) estimate parameters, (3) re-estimate the missing values assuming the new parameter estimates are correct, (4) re-estimate the parameters, and so forth, iterating until convergence. So, each iteration of the EM method consists of an E step (expectation step) and an M step (maximisation step). An additional advantage of the algorithm is that it can be shown to converge reliably. A disadvantage of EM is that its rate of convergence can be painfully slow if a lot of data are missing.

Applied work in the area of unconstraining airline demand data began in the mid-1990s, when several researchers published very similar approaches. Skwarek (1996) and Hopperstad *et al.* (1996) reviewed four unconstraining methods (N2, N3, BP and PD) and found that among the four, booking profile and projection detruncation were the best as measured by showing a revenue increase of 2–3 per cent over using the N2 method. Saleh (1997) only reviewed the three naïve methods (N1, N2 and N3) and concluded that using a method like N2 can significantly understate the true demand and even a simple change such as moving to N3 from either N1 or N2 can be quite helpful. Pölt (2000) and Weatherford (2000) were the first to even look at the EM method in the airline context and also reviewed four additional approaches (N1, N2, N3 and BP). Weatherford concluded that N1 is better than N2 in most realistic cases, N3 is better than N1 and they both concluded that EM was the most robust method available for unconstraining, even if they measured its robustness in different ways (Weatherford measured how close the unconstraining method came to approximating the true mean; Pölt measured sampling bias and mean absolute error). Finally, Zeni (2001a) looked at all six methods (N1, N2, N3, BP, PD and EM) and concluded the same as Weatherford and Pölt — that N1 is better than N2, and that EM was the most robust method of all six as measured by error reduction. Unfortunately, none of these has been published to date in scholarly journals, nor did any of them look at all six unconstraining methods and quantify the revenue impact of the different methods.

Orkin (1998) looked at a very simple unconstraining method as applied to the hotel industry. Oppitz and Pölt (1998) as well as Lough (1997) investigated the problems of leg-based unconstraining in an origin and destination (O&D) world, where it is possible that at the same time some O&D paths (eg Denver–Chicago–Frankfurt) across a given leg (eg Denver–Chicago) might be available and others not (Los Angeles–Denver–Chicago). A related paper by Weatherford and Belobaba (2002) examines the impact on revenue from misestimation of the forecasts of demand for each fare class when fed into the optimisation engine. They concluded that forecasts that were off by 12.5–25 per cent on the low side (eg underestimating true demand due to poor unconstraining) can hurt revenues by 1–3 per cent on high-demand flights. Compared with the total revenue benefit cited earlier (2–5 per cent), one can see that unconstraining is a very important issue to be addressed by the airlines.

EXPERIMENTAL DESIGN AND SIMULATION METHODOLOGY

The first phase of the research will test all six of the unconstraining methods against three different simulated data sets (each with varying degrees of constrained observations). The purpose of this phase is to compare how close the different methods come to the true mean value, which can only be measured using simulated data. For the first data set, in step 1, 1,000 random observations representing demand will be generated (which will be used for 100 observations for each of ten checkpoints) from a normal distribution with mean = 20, standard deviation = 4. The assumption to use 100 observations is based on working with real airline data — they often have about 2-years' worth of historical departures available to make the next forecast. Since the data are seasonal by day of week, they only have one observation for each day of the week. So, they would have about 104 (2 yrs × 52 weeks/yr) Mondays' worth of observations available to make the next Monday's forecast.

In step 2, booking limits will be

randomly generated with successively lower mean values to create six different scenarios where between 0 per cent and 98 per cent of the actual observations generated in step 1 are constrained. For a simple example of how these two steps will work, suppose ten random observations were generated for demand [$\tilde{}N(20,4)$] as follows:

18 20 21 19 20 17 21 23 18 22

and then drew ten random booking limits from a normal distribution with mean = 20, standard deviation = 4 (ie ˜N(20,4)) as follows:

20 19* 17* 22 21 18 19* 21* 22 20*

An asterisk indicates a case where the booking limit constrains the demand. The following information (ie the actual bookings — the lower of the two random values, and whether the observation was affected by the booking limit or not) would then be fed to the six different unconstraining methods:

18 19* 17* 19 20 17 19* 21* 18 20*

and see what estimate of unconstrained demand each method would come up with. In this example, 50 per cent of the observations are constrained and the estimated mean values from the six different methods would be compared (numerical examples with each method are given in the next section). A summary table is presented in Table 1.

Next, using the same ten initial observations for demand, ten new random booking limits would be generated from a different normal distribution (eg with mean = 18, standard deviation = 4) and an even higher percentage of constrained observations would most likely be found. This process is repeated six times, generating six different percentages of constrained observations (ie 0 per cent, 20 per cent, 40 per cent, 60 per cent, 80 per cent, 98 per cent). The actual mean values for the booking limits used to generate these six scenarios for data set 1 were 35, 25, 22, 18.5, 15 and 7. It is important to note that 100 per cent of the observations cannot be allowed to be constrained because the statistical methods (EM, PD) rely on an iterative estimation and thus have no additional information (ie open observations) to help in such a case.

For the second simulated data set, in step 1 1,000 random observations for demand will be generated (again ten checkpoints with 100 observations each) from a normal distribution with mean = 20, standard deviation = 7. In step 2, booking limits with successively lower mean values will be randomly generated to create scenarios where between 0 and 98 per cent of the observations are constrained. The purpose of this data set is to show the effects of a larger variance (compared with data set 1). The actual mean values for the booking limits used to generate these six scenarios for data set 2 were 37, 26, 22, 18.5, 14 and 7.

For the third simulated data set, in step 1 1,000 random observations for demand will be generated (again ten checkpoints with 100 observations each) from a normal distribution with mean = 4, standard deviation = 2. In step 2, booking limits with successively lower mean values will be randomly generated to create scenarios

Table 1: Summary table for ten observations

True demand	18	20	21	19	20	17	21	23	18	22
Booking limit	20	19	17	22	21	18	19	21	22	20
Observed demand	18	19	17	19	20	17	19	21	18	20

Table 3: Updated demand using booking profile method

Observ. No.	*DCP 1*	*DCP 2*	*DCP 3*
1	18.00	18.59	18.77
2	18.40	17.00	18.00
3	18.40	19.00	19.19
4	19.00	19.63	20.00
5	20.00	20.66	20.86
6	17.00	17.56	17.73
7	18.40	18.00	18.67
8	18.40	20.00	19.00
9	18.00	18.59	20.00
10	18.40	21.00	19.00
Mean	18.40	19.00	19.12

DCP 3). The revised observations are then calculated as shown in Table 3.

The PD and EM methods are both iterative methods. Each iteration consists of two steps. The first step is to replace closed observations by estimated values and the second step is to (re-)estimate the mean and standard deviation of the underlying probability distribution. The difference between EM and PD occurs in the estimation of the closed observations based on the estimated parameters. EM uses a maximum likelihood estimation, while PD uses another heuristic which is scalable by a parameter (τ). Both methods start with an initial estimate of the mean (18.8) and the standard deviation (1.25) using all available data. For each closed observation, EM calculates a conditional expectation from the distribution normal (18.8, 1.25) under the condition that the estimate of the unconstrained value for the closed observation is at least as large as the actual value. The first iteration of EM in the above example gives:

EM (iter #1):
18 19.9 19.0 19 20 17 19.9 21.5 18 20.7

with a new mean of 19.3 and a new standard deviation of 1.3. After six iterations, the EM algorithm converges to

EM (iter #6):
18 20.3 19.6 19 20 17 20.3 21.7 18 20.9

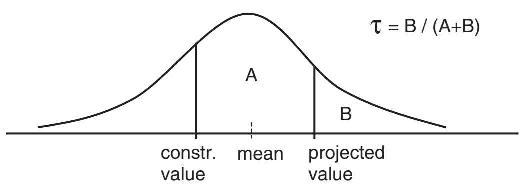
with a new mean of 19.5 and a new standard deviation of 1.4. Note that the estimate of the mean has been increased from 18.8 to 19.5 with this unconstraining method and, furthermore, that this estimate is higher than that obtained by any of the three naïve methods.

Projection detruncation uses a different heuristic for the estimation of the unconstrained value of the closed observations in each iteration. Instead of calculating a conditional expectation, it balances two things: (a) the 'weight' (or area) of the probability distribution between the original constrained value and the new estimate or projected value (area A in Figure 1) with (b) the weight (area) between the new estimate and infinity (area B in Figure 1).

The parameter τ allows PD to scale the aggressiveness of its unconstraining. With $\tau = 0.5$, PD really balances the two areas A and B and gives very similar results to the EM method (ie the only real difference is that PD takes the conditional median, while EM takes the conditional mean). With smaller values of τ, the closed observations are projected to higher (more aggressive) values and PD starts to diverge from the EM approach.

As an example of the PD method, using a more aggressive τ value (0.3), the first

Figure 1: Estimation process for constrained observations using the projection detruncation method



where between 0 and 98 per cent of the observations are constrained. The actual mean values for the booking limits used to generate these six scenarios for data set 3 were 12, 7.5, 5, 3.4, 2.1 and 0.7.

The experimental design and simulation methodology for the revenue testing phase of the research is described at the beginning of the section 'Revenue impacts of different unconstraining methods' below.

DESCRIPTION OF THE UNCONSTRAINING METHODS

This section describes the six different unconstraining methods in greater detail and gives a numerical example. The simple example of ten observations for bookings from the previous section is used to explain how each method works:

18 19* 17* 19 20 17 19* 21* 18 20*

(Remember that an asterisk indicates a constrained observation, where the booking limit was smaller than the demand.)

Method N1 replaces all closed observations with the mean of all observations (18.8). Method N2 replaces all closed observations with the mean of all open observations (18.4). Method N3 replaces all closed observations with the larger of the mean of all open observations and the actual (closed) individual observation. In summary, the three naïve methods end up with the following estimates of demand for each of the ten observations:

N1:
18 18.8 18.8 19 20 17 18.8 18.8 18 18.8
Average: 18.6
N2:
18 18.4 18.4 19 20 17 18.4 18.4 18 18.4
Average: 18.4
N3:
18 19 18.4 19 20 17 19 21 18 20
Average: 18.94

Because of the complicated nature of the booking profile method, it can only be described with an example over several data collection points (DCP), which are alternatively known as reading days or checkpoints. Table 2 shows the initial observations available.

Table 2: Initial observed demand across three DCPs

Observ. No.	*DCP 1*	*DCP 2*	*DCP 3*
1	18	20*	22*
2	19*	17	18
3	17*	19	21*
4	19	21*	20
5	20	19*	21*
6	17	20*	23*
7	19*	18	21*
8	21*	20	19
9	18	19*	20
10	20*	21	19
μ open obs.	18.4	19.0	19.2
Ratio		2.033	1.513

* indicates a case where the booking limit constrains the demand.

The booking profile method in DCP 1 replaces any closed observation with the mean of the open observations. In DCP 2, it replaces any closed observation with the revised observation in DCP 1, multiplied by a ratio described in the next sentence minus 1. The ratio is calculated as the sum of the mean of open observations in DCP 2 and the mean of open observations in DCP 1, divided by the mean of open observations in DCP 1 ('Ratio' in last row of table, column DCP 2). For DCP 3, it replaces any closed observation with the sum of the revised observation in DCP 1 and the revised observation in DCP 2, multiplied by a ratio described in the next sentence minus 1. The ratio is calculated as the sum of the mean of open observations in DCP 3, DCP 2 and DCP 1, divided by the sum of the mean of open observations in DCP2 and DCP 1 ('Ratio' in the last row of the table, column iteration of PD using the same data as above gives:

PD (iter #1):
18 20.2 19.5 19 20 17 20.2 21.6 18 20.8

with a new mean of 19.43 and a new standard deviation of 1.36. After 12 iterations, the PD algorithm converges to:

PD (iter #12):
18 21.5 21.1 19 20 17 21.5 22.5 18 21.9

with a new mean of 20.05 and a new standard deviation of 1.82.

SIMULATION RESULTS ON UNCONSTRAINING ESTIMATES

Recall that the purpose of this phase is to compare how close the different unconstraining methods come to the true mean value, which can only be measured using simulated data. For each of the three data sets, in the first step 1,000 random observations representing demand from a normal distribution will be generated. In the second step, booking limits with successively lower mean values will be randomly generated to create six different scenarios where between 0 per cent and 98 per cent of the actual observations are constrained. The actual bookings and an indicator of whether the observation was affected by the booking limit or not are then fed to the six different unconstraining methods to see what estimate of unconstrained demand they come up with.

Because both the PD and EM methods are statistically based, it is anticipated that they would outperform the naïve methods. The extent to which they did better was surprising. In Figure 2, the results for the first data set (recall this set has a true mean value of 20) indicate the substantial improvement that the PD (using $\tau = 0.5$) and EM methods generate over both the BP and the Naïve methods, especially as the percentage of data that are constrained increases to 40 per cent or higher. For example, with data that are 80 per cent constrained, EM returns an estimate of 18.2, while N1 gives an estimate of 14.6. For the second data set (recall this set has a true mean value of 20), similar benefits of PD and EM over both the BP and the Naïve methods are shown in Figure 3. For the third data set (recall this set has a true mean value of 4), similar benefits of

Figure 2: Estimates of unconstrained means from six unconstraining methods on data set 1

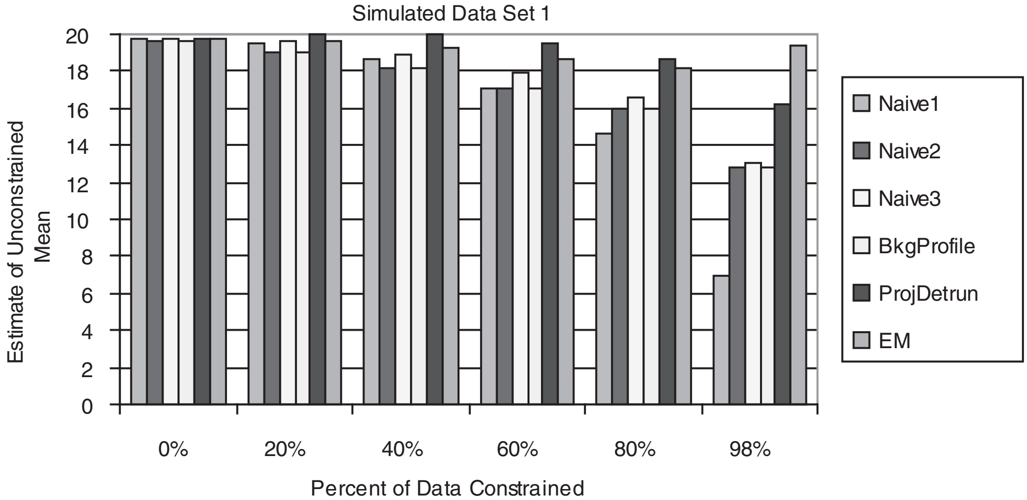
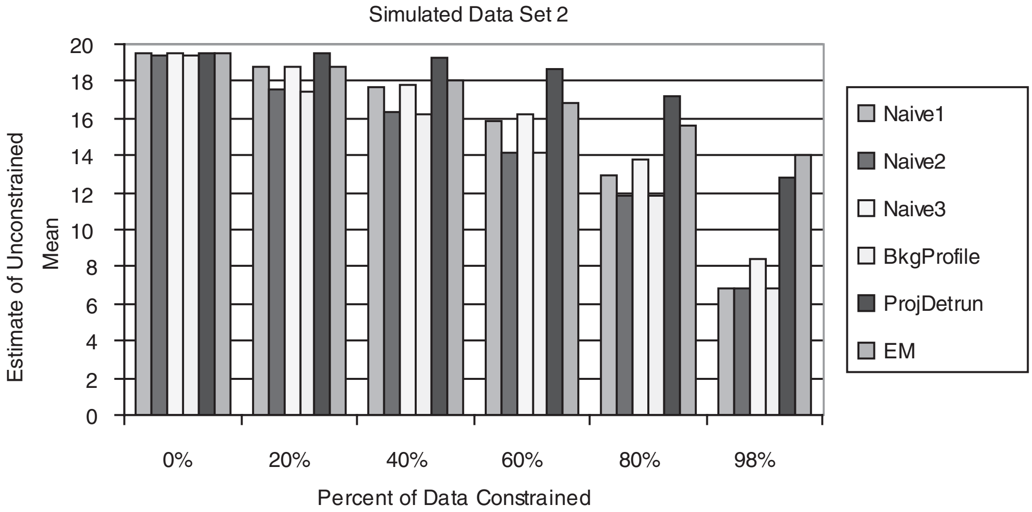


Figure 3: Estimates of unconstrained means from six unconstraining methods on data set 2



PD and EM over both the BP and the Naïve methods are shown in Figure 4. It was interesting how weakly the booking profile method performed, but in reality it is a lot like N2 in the sense that it only uses 'open' data to estimate the booking profile, with the added weakness that in effect it is using a 'multiplicative' approach to estimating (eg mean bookings at checkpoint 2 divided by mean bookings at checkpoint 1), which has been shown previously to have serious flaws (Weatherford, 1998).

Attention now turns to the matter of bias, defined here as the true mean of the sample before any booking limits are applied (ie demand) minus the estimated mean as derived by the individual unconstraining methods. Of course, the true mean cannot be observed in practice, but fortunately with simulated data the true value is known. By its definition then, a

Figure 4: Estimates of unconstrained means from six unconstraining methods on data set 3

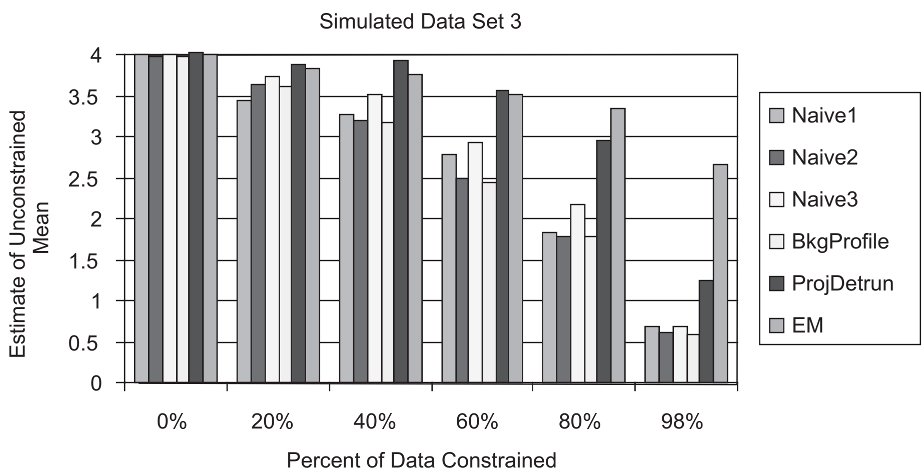
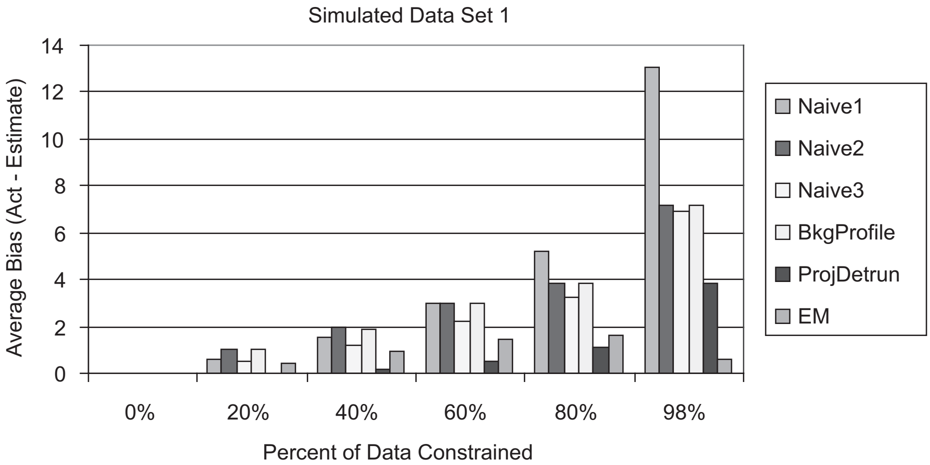


Figure 5: Average bias from six unconstraining methods on data set 1



positive bias means that the estimate was *less* than the true mean. Conversely, a negative bias would mean that the estimate was more than the true mean. The bias values reported are aggregated numbers (ie the average bias over all the ten different checkpoints and over all 100 random observations in each checkpoint). For simulated data set 1, Figure 5 shows that the bias is essentially zero when none of the data is constrained for all of the six methods, but as the percentage of data constrained grows, so does the underestimate of the true mean by each of the methods. But, as before, It can be seen that both PD and EM are definitely the 'best' of the six methods, meaning they have the smallest bias. Ideally, an unconstraining method that has zero bias under any condition of the raw data (eg high or low percentage

Figure 6: Average bias from six unconstraining methods on data set 2

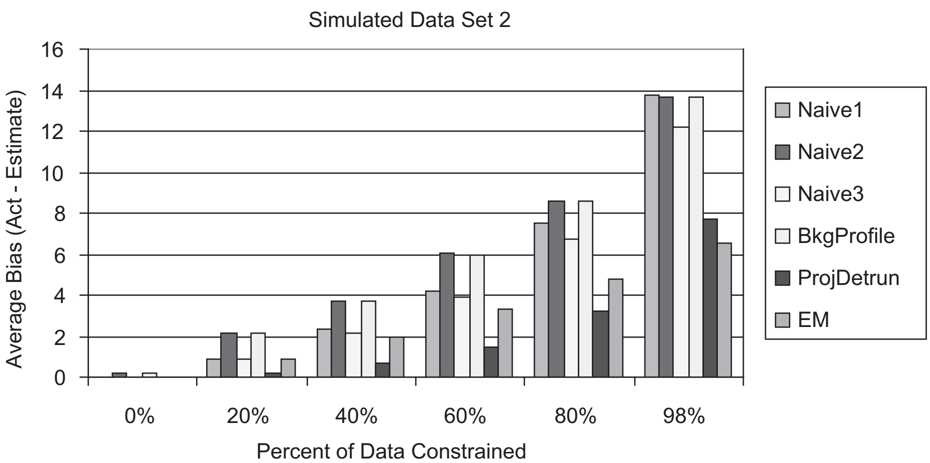
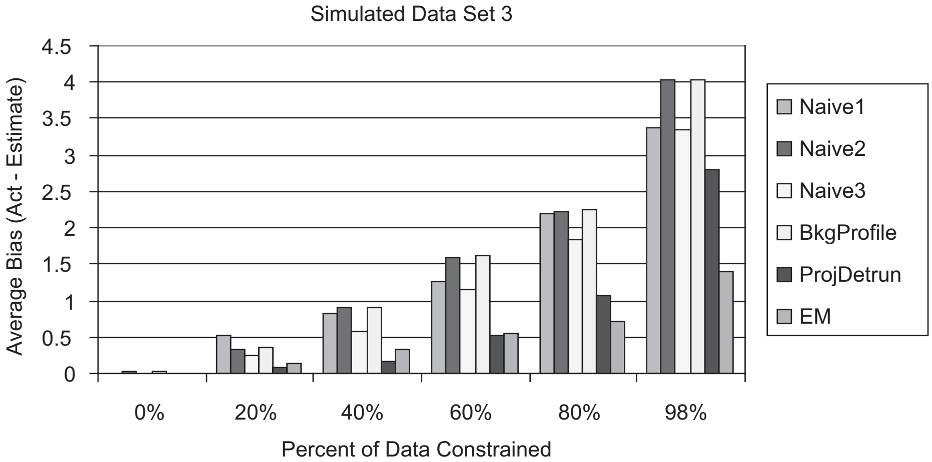


Figure 7: Average bias from six unconstraining methods on data set 3



constrained) could be found, but unfortunately that does not seem to be the case here, with even EM and PD exhibiting a positive bias in the 60+ per cent constrained data environment. The same basic trend emerges using simulated data sets 2 and 3, as shown in Figures 6 and 7.

The next set of graphs show what percentage improvement in the estimate of the unconstrained mean is gained by an individual unconstraining method over the N2 method. Figures 8–10 show that the percentage increase can be 10–40 per cent or more (depending on the data set) with the PD and EM methods as the percentage of the data that are constrained starts hitting 60 per cent or higher. This is a very significant improvement, especially when one remembers, as previously cited (Weatherford and Belobaba, 2002), that the effect of underestimating demand by 12.5–25 per cent can hurt revenues by 1–3 per cent on high-demand flights. The results shown here of potentially underestimating by 40 per cent or more owing to poor unconstraining would seem to predict rather

Figure 8: Percentage increase of estimated unconstrained mean of five methods over N2: for data set 1

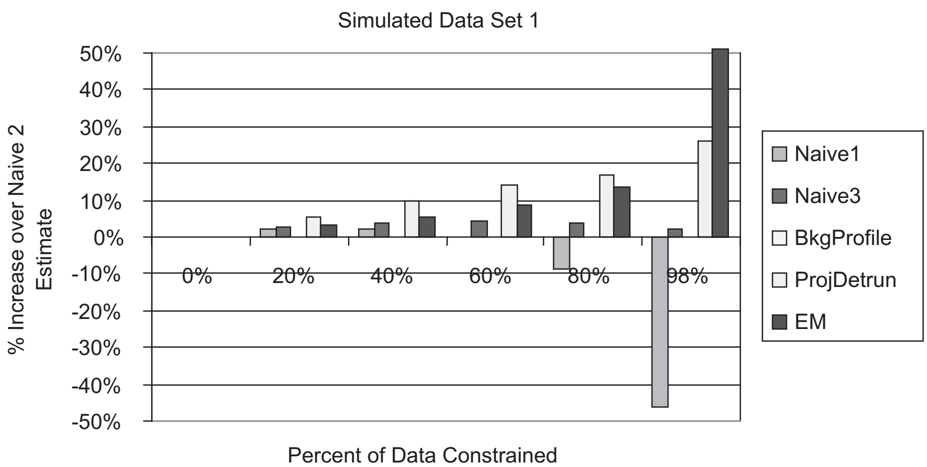
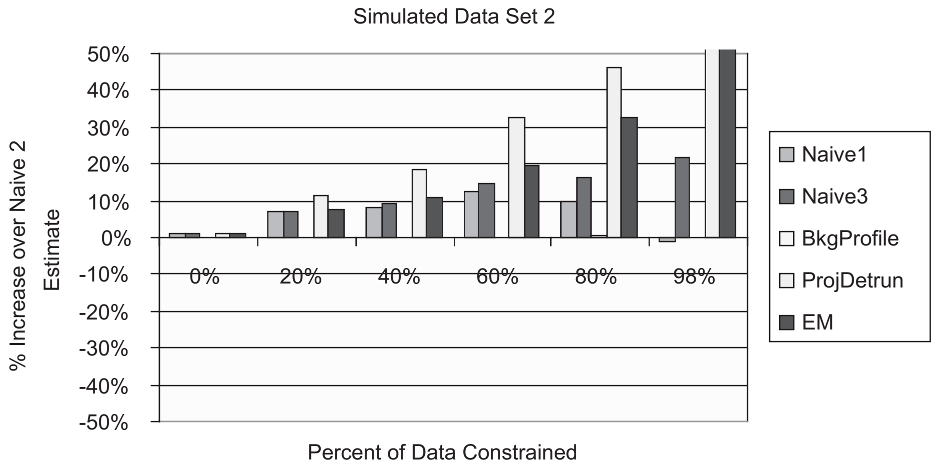


Figure 9: Percentage increase of estimated unconstrained mean of five methods over N2: for data set 2

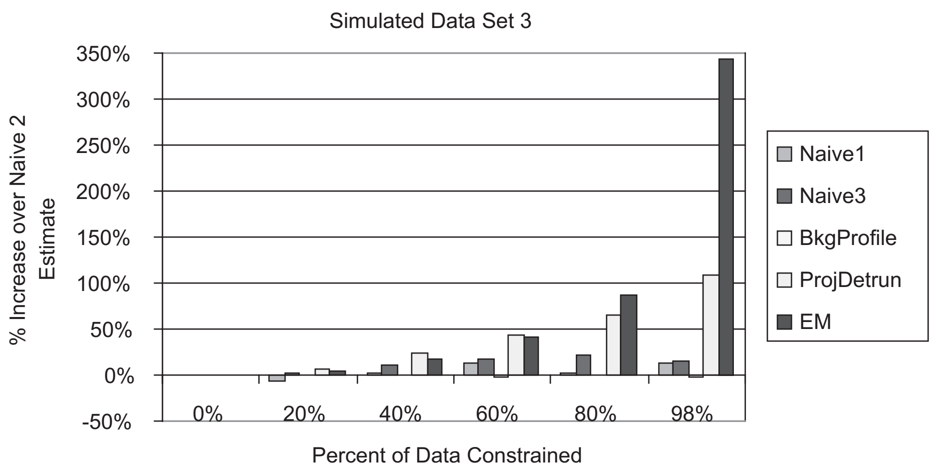


large impacts on revenue when the study on revenue is pursued in the next section. One interesting side note is that the N1 method starts underperforming N2 in data set 1 at the 80–98 per cent constrained data level.

More aggressive approaches on the two robust unconstraining methods (PD and EM)

In the discussion on how the PD method works, one of the parameters that had to be set ahead of time was the τ level. In all of the previous results, the value of $\tau = 0.5$ was used because it represented equal weights on the area between the constrained value and the estimate of unconstrained value *and* the area between the estimate of the unconstrained value and infinity. As shown in Figures 5–7, it works well for data sets that are largely unconstrained (0–40 per cent), but the performance deteriorates at higher percentage constrained data sets (60+ per cent).

Figure 10: Percentage increase of estimated unconstrained mean of five methods over N2: for data set 3



The question is: Can the τ level be adjusted to make it more aggressive in estimating the unconstrained mean? The answer is yes. Figure 11 shows the results of testing τ levels of 0.4 and 0.3, along with the base case of 0.5 on simulated data set 1. For a given percentage of the data that are constrained, decreasing the τ value does have the effect of increasing the estimate of the unconstrained mean. Note that as the percentage of data constrained gets up to the 80–98 per cent level, the estimate is now too high. Similar results can be seen for data set 3 in Figure 12.

In practice, some kind of maximum cap on the estimated unconstrained mean would need to be implemented, as well as an unconstraining system that would automatically choose decreasing values for τ (from 0.5) as the percentage of data constrained increases. But the very good news is that in the 60–98 per cent data constrained range, the level of τ can be adjusted to make the algorithm come up with a more aggressive estimate of the unconstrained mean.

REVENUE IMPACTS OF DIFFERENT UNCONSTRAINING METHODS

The next phase of the research was to take the more robust unconstraining methods (EM and PD with $\tau = 0.5$) and the naïve methods (N1, N2, N3, BP) and integrate them into an RM system such that historical bookings data were available and the unconstraining methods were then used to forecast 'demand'. These forecasts were then used to establish optimal booking limits (using EMSRb, which has been the industry standard for leg optimisation since it was introduced; Belobaba, 1992b) and the booking limits were used to interface against randomly generated arrivals demand to calculate the actual revenue obtained using the various unconstraining methods. To keep the playing field level in comparing the five different unconstraining methods, the forecasting method used was the average of past unconstrained observations — which makes sense given that random demand was being generated from a stable mean, and EMSRb was used as the optimisation method. Thus, the only difference between the five revenue generating

Figure 11: Effect on projection detruncation estimate by adjusting tau value on data set 1

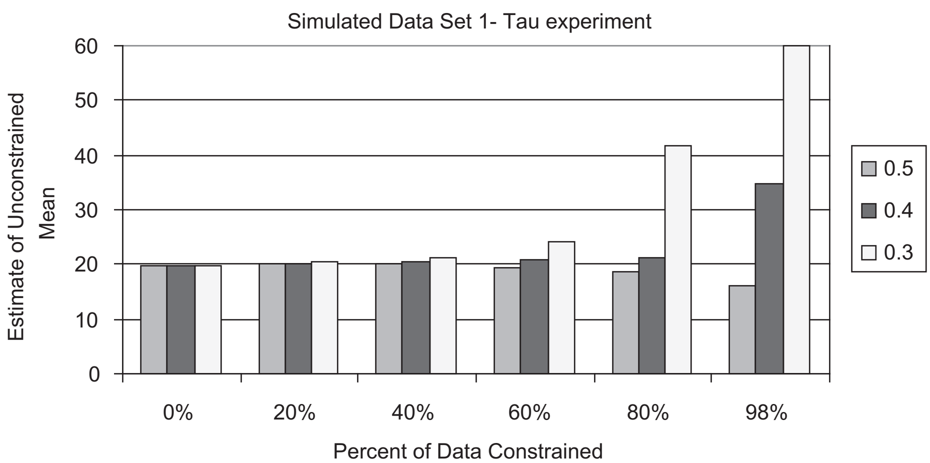
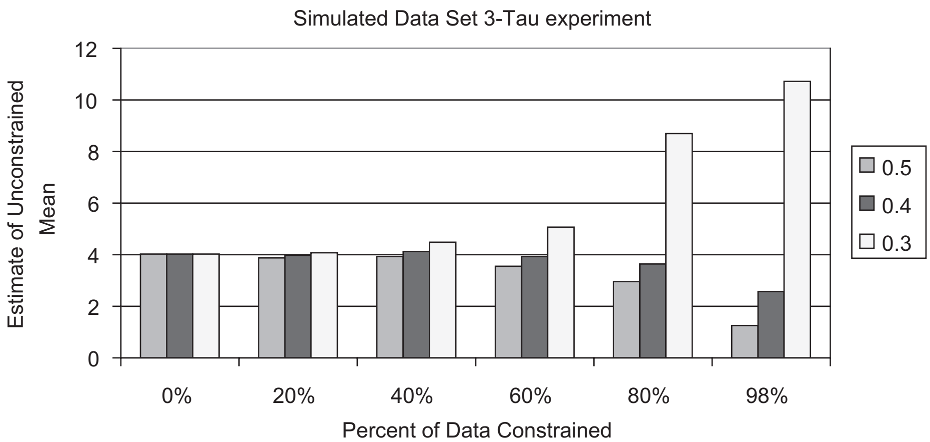


Figure 12: Effect on projection detruncation estimate by adjusting tau value on data set 3



approaches was the unconstraining methods themselves.

Numerous iterations (ie 1,100) of this complete RM system loop were carried out so that the effect of the five different unconstraining methods over time could also be observed (effectively simulating 1,100 weeks of history or over 20 years). The details of the actual airline data used and the simulation methodology are described next, followed by the actual revenue performance of the different unconstraining methods.

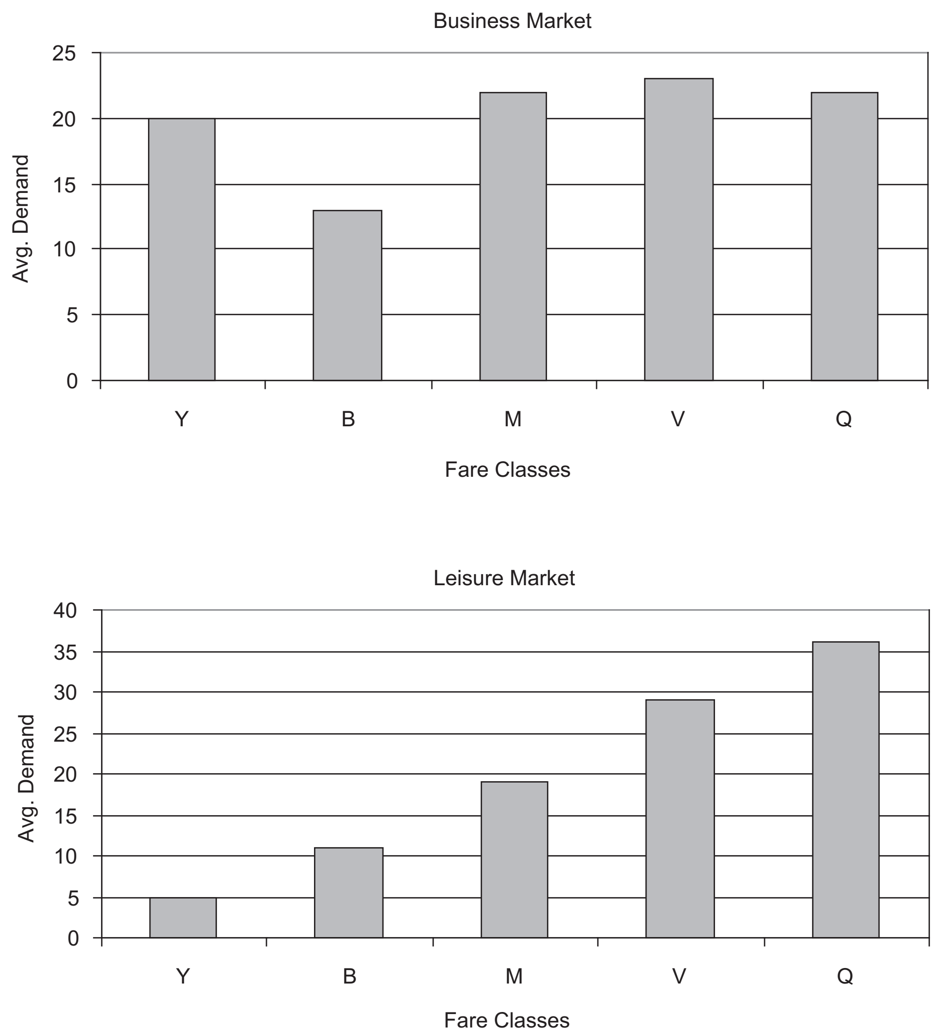
The revenue simulation model that employed in the revenue evaluation is similar to that described in Belobaba and Weatherford (1996). The major assumptions of the simulation process are: (1) demand for each booking class is separate and independent of other class demands; (2) demand for each class is stochastic and can be represented by a probability distribution (Poisson); and (3) the proportion of demand expected to arrive within a given booking period is known.

A Poisson distribution was used to model the number of customer arrivals within each booking period. The full range of bookings was taken by day up to 360 days out and the values for cumulative bookings at ten different checkpoints were selected out. Over the ten booking periods prior to departure (checkpoints), the mean arrival rates of each fare class replicate the tendency seen by airlines on most of their flight legs — that of low-fare passengers requesting seats before passengers willing to pay the highest fares. Requests for different fare classes are actually interspersed over the booking process, and thus a few higher fare passengers may actually book before all of the lower fare ones have booked. In addition, within a booking period, all the arrivals from the five different fare classes are scrambled together.

Two different data sets from a major US airline were examined — one more of a 'business' route and the other more of a 'leisure' route. To see the total mean demand (over the ten booking periods) for the five fare classes in each data set, see Figure 13.

The multiple booking periods (ten) also allow the simulation to replicate the practice by airlines of updating their demand forecasts and revising the optimum booking limits as bookings are accepted. In this revenue simulation, the booking limits are updated by re-solving the optimisation model at the beginning of the ten booking

Figure 13: Mean demand by fare class for the two different markets



periods prior to departure, given the remaining available flight capacity and forecasted demand still to come.

For each of the two data sets, three demand-to-capacity levels, or 'demand factors', were created ranging from 0.90 to 1.50. Under the simulated revenue management environment, these demand factors result in average load factors between 89 and 99 per cent, which would clearly be considered to be in the above-average to very high demand range (ie 'critical' flights) for most airline companies. It is important to emphasise from the outset that the simulation results presented in this paper are most relevant for such high-demand situations, in which the use of a revenue management system is recognised to have the greatest revenue impact. Visual C++ was used to code up the simulation and, for each of the scenarios (two data sets × three demand factors), 1,100 trials were run to ensure statistically significant revenue differences. In each case, the total revenue was calculated for each of the five unconstraining methods tested and then a percentage increase (decrease) over the N2 method was calculated. N2 was chosen as the 'base case' because it was the simplest method that still showed some reflection of the fact that the data could be constrained. (N1 is also very simple but ignores the fact that data could be constrained.)

For the 'business' market, Figure 14 shows that at a demand/capacity (D/C) ratio of 0.9, there is no significant revenue difference between the five methods. Basically, the reasoning for this is that if demand does not reach capacity very often, there are not very many constrained observations in the data, and therefore the method used to unconstrain is not as critical. As the D/C ratio started to increase to 1.2, it could be seen that N1 underperformed by nearly 1 per cent, BP underperformed by 10 per cent, while N3 outperformed by 1 per cent and PD and EM outperform N2 by 1.3 per cent. The reason BP does so poorly is that it only uses 'open' data to estimate the booking profile, with the added weakness that in effect it is using a 'multiplicative' approach to estimating the unconstrained demand when an observation is closed (eg mean 'open' bookings at checkpoint 2 divided by mean 'open' bookings at checkpoint 1). The 'multiplicative' idea has been shown previously, in a forecasting setting, to have serious performance flaws (Weatherford, 1998).

As the highest D/C ratio is approached, these revenue differences become even more pronounced. N1 underperforms by 7.5 per cent, BP underperforms by 24 per cent, while N3 outperforms by more than 2 per cent, and PD and EM outperform N2 by 2.9 per cent. Again, remember that even an average revenue benefit of 0.5 per cent (a mixture of flights that have 0 per cent benefit, other flights with a 1–1.3 per cent benefit, and still others with 2–2.9 per cent benefit) for a large airline such as Lufthansa German Airlines with annual revenues exceeding €15bn can mean substantial benefits of €75m.

For the 'leisure' market, Figure 15 reveals some even more dramatic results, showing that the percentage revenue improvement seems to be highly dependent on the nature of the data set (mean

Figure 14: Revenue impacts of five unconstraining methods in business market

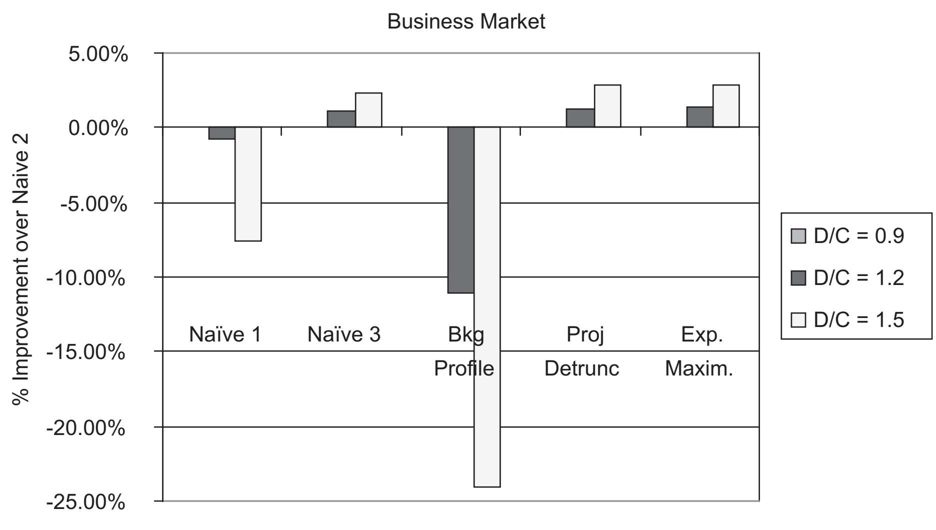
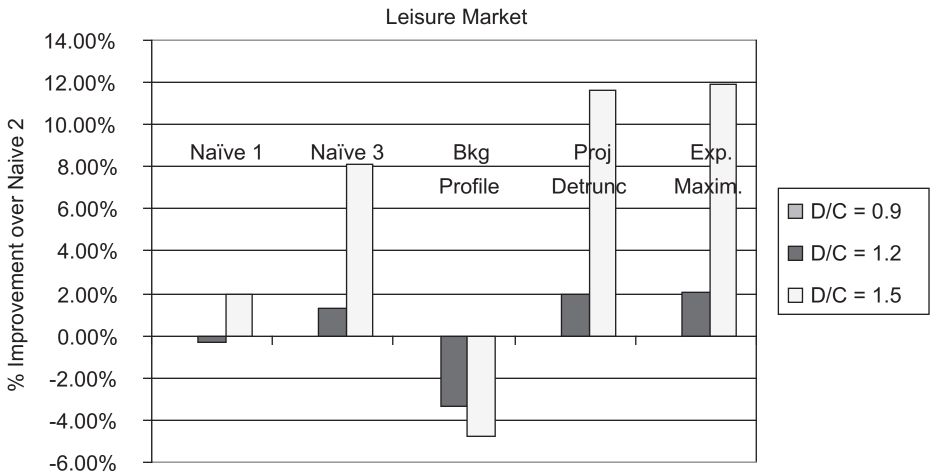


Figure 15: Revenue impacts of five unconstraining methods in leisure market



demand by fare class, average fares, how the demand comes in over the ten booking periods, etc.). At the D/C ratio of 0.9, there is still no significant revenue difference between the five methods. As the D/C ratio starts to increase to 1.2, it can be seen that N1 underperforms by 0.3 per cent, BP underperforms by 3 per cent, while N3 outperforms by 1.3 per cent, and PD and EM outperform N2 by 2 per cent. As the highest D/C ratio is approached, these revenue differences become even more exaggerated. BP underperforms by 5 per cent, N1 now outperforms by 2 per cent, while N3 outperforms by more than 8 per cent, PD outperforms by 11.6 per cent, and EM outperforms N2 by 11.9 per cent.

It was amazing to see such strong revenue differences, which highlighted even more the importance and urgency of upgrading the quality of the unconstraining methods used by many airlines today.

CONCLUSIONS AND IMPLEMENTATION ISSUES

In the evolution of revenue management theory and practice, more attention has been paid to the optimisation models (either leg or network based) used to allocate seats to different fare classes and overbooking models, rather than to forecasting models and specifically unconstraining models. This paper has focused on six different unconstraining methods that are either used in practice or known in the statistical world. It has explored the impacts of these unconstraining methods on both the actual estimates of unconstrained demand from a sample of realistic (though simulated), constrained booking data and the impact on revenue performance using a simulation, where the results from the unconstraining methods are used as the forecasts of future demand, which are then fed to the most commonly used decision rule for optimisation (EMSR). A booking simulation was applied that generated random customer arrivals to two different legs of actual airline demand data under a variety of demand-to-capacity factors to evaluate the revenue performance of the different unconstraining methods.

First an assessment was made of the actual numerical estimation of the unconstrained demand by the six methods. The simulation results showed that as the percentage of historical booking data that is constrained increases to 20 per cent or higher, the three naïve methods and the booking profile method really start to suffer. Only PD and EM perform adequately. Between these two more robust methods, in general, EM appears to be the more robust of the two methods, especially in situations where 90+ per cent of the data are constrained, although it can depend on the nature of the data.

Using the more robust methods for unconstraining can have a significant impact on the expected revenue performance, given that the airline currently uses the EMSR model for seat inventory control. In fact, revenue improvements of 1.3–2.9 per cent on 'business' legs with load factors between 96–97 per cent and revenue improvements of 3–12 per cent on 'leisure' legs with load factors between 97–98 per cent have been have demonstrated. (Note: although these load factors may seem high for practitioners, remember that no-shows and overbooking have not been dealt with in this study, so that the actual load factors in practice would come in smaller than these numbers.)

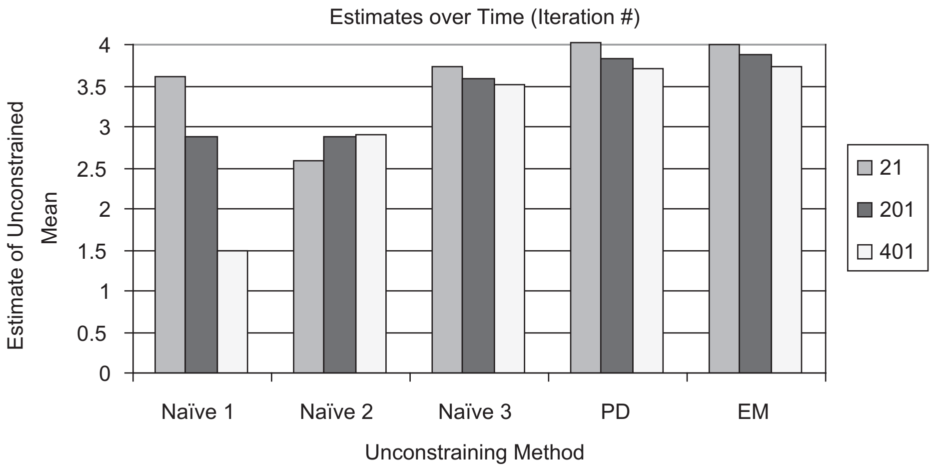
The results of these simulations should not be interpreted as conclusive with respect to the exact revenue impacts that can be expected for every airline company, as it was seen that differences in the composition of demand and booking patterns can cause substantially different impacts, even in the two examples tested (compare leisure market, DF = 0.9 with revenue increase of 0 per cent, with business market, DF = 1.2 with revenue increase of 1.3 per cent, with leisure market, DF = 1.5 with revenue increase of 11.9 per cent). The findings do demonstrate, however, the importance of using a good unconstraining method and that the potential revenue impact of upgrading the unconstraining process is at least as significant, if not more so, as the step in moving from leg-based optimisation to network- (or O&D) based optimisation or as the step in moving to better forecasting methods.

As for real-world implementation issues, six items are mentioned. (1) In trying to apply these leg-based unconstraining ideas to an O&D world, one runs into the difficulty that, whereas in the leg world, fare class availability is either open or closed (0/1), in the O&D world it is not so clean. It is possible to have partially open legs (some O&Ds are available and others are closed). (2) A company would probably want to apply a maximum 'cap' on the unconstraining results from the EM and PD methods as they can tend to overestimate the mean in highly constrained data situations. For example, if the original mean of the observations (both open and closed) is 13 and the estimate of the unconstrained mean is 50, it may seem prudent to check all the unconstrained results and 'cap' them at 2 × original mean (equals 26 in this example).

New implementation issues include: (3) for the EM method, some parameters need to be set as well — the maximum number of iterations allowed and what the criteria are for determining that the algorithm has converged. (4) Even for a leg-based application, there may be some inaccurate data if the indicator as to whether a booking class was open or not says that it was open on two consecutive checkpoints (eg day 330 and day 300), when in fact it might have been closed to bookings for a few days in-between the checkpoints. (5) Even though the simulation assumed that there was no correlation between demands in different fare classes, it is not hard to imagine that this may not be true. For example, say a customer wants to buy a deeply discounted ticket (Q class) on a given flight leg but is told that there are no more seats for sale at that price. S/he may be willing to trade up and buy a ticket in the next higher fare class (V class) if it is available.

Finally, a very interesting feedback loop

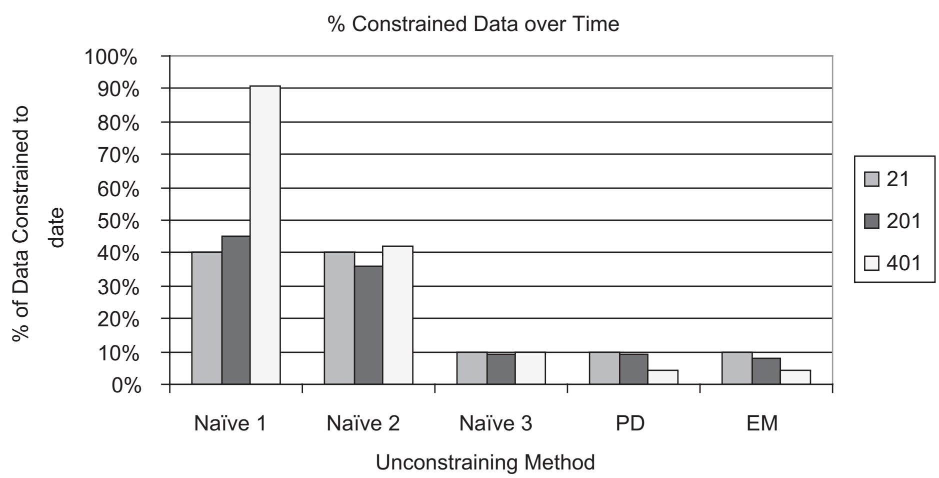
Figure 16: Estimates of unconstrained mean over time by five unconstraining methods



emerges when looking at what happens over time to different fare classes if they tend to be constrained and how well they are unconstrained. Out of the many data samples, the highest fare class (Y) was randomly picked from the 'business' leg, in the next-to-last booking period (nine out of ten) under the D/C ratio of 1.5 so that a lot of constrained values could be seen. Then the 1,100 iterations (that represent simulating 1,100 weeks of this flight leg departing) were randomly sampled, to look at the results for iterations 21, 201 and 401. The results are shown in Figures 16 and 17, and they tell an interesting story.

The true mean for this fare class/booking period combination is 3.6 arrivals. As shown in Figure 16, unconstraining method N2 underestimates the true demand (2.6 vs 3.6) early on in iteration 21 (owing to its being a not-so-robust method), and then it can never seem to catch up later on (iterations 201, 401 never get above an estimate of 3.0). N1 also

Figure 17: Percentage of data constrained over time by five unconstraining methods



seems to be stuck in the same trap. The feedback loop is a rather negative one: underestimate demand initially, forecast less demand for future flights, open up the Y fare class bucket less in the future, which leads to even more constrained values in the future, and even further underestimating. In sharp contrast, both the EM and PD method actually overestimate demand (4.0 vs 3.6) early on in iteration 21, which turns out to lead to a very nice feedback loop. Here, the pattern seems to be: slightly overestimate demand initially, forecast more demand for future flights, open up the Y fare class bucket more in the future, which leads to fewer constrained values in the future, and more accurate estimating due to revising the estimate downwards when demand does not materialise (note that at iteration 401, both EM and PD estimate the demand at 3.6, compared with the true mean of 3.6). Figure 17 shows the percentage of the observations that are constrained for the same data (fare class/booking period combination) and the same time intervals.

The conclusion is that airlines should definitely invest in more aggressive unconstraining methods (EM, PD) compared with the traditional approaches (N1, N2, N3). The simulation results showed that as the percentage of historical data that is constrained increases to 20 per cent or higher, the three naïve methods and the BP method really start to suffer. Only PD and EM perform adequately. Between these two more robust methods, in general, EM appears to be the more robust of the two methods, especially in situations where 90+ per cent of the data are constrained. Revenue improvements of 1.3–2.9 per cent on 'business' legs with load factors between 96–97 per cent and revenue improvements of 3–12 per cent on 'leisure' legs with load factors between 97–98 per cent have been demonstrated. The findings demonstrate the importance of using a good unconstraining method and that the potential revenue impact of upgrading the unconstraining process is at least as significant, if not more so, as the step in moving from leg-based optimisation to network (or O&D) based optimisation or as the step in moving to better forecasting methods.

References

Belobaba, P. P. (1992a) 'The revenue enhancement potential of airline yield management systems', in Behrendt, R. and Bertsch, L. (eds) *ASTAIR Proceedings*, Advanced Software Technology for Air Transport, London.

Belobaba, P. P. (1992b) 'Optimal vs. heuristic methods for nested seat allocation', in L'Heureux, E. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Brussels, Belgium, 28–53.

Belobaba, P. P. and Weatherford, L. R. (1996) 'Comparing decision rules that incorporate customer diversion in perishable asset revenue management situations', *Decision Sciences*, **27,** 2, 343–363.

Hopperstad, C., Belobaba, P. P. and Skwarek, D. K. (1996) 'Passenger O/D simulator update', in Cholak, U. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Zurich, Switzerland.

Lee, S. (2000) 'Forecasting and Detruncation in Network D. Report to PODS Summit XII', Unpublished report in possession of authors.

Little, R. J. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*, John Wiley, New York.

Lough, G. (1997) 'Unconstraining in a bid–price environment', *Proceedings of PROS 5th Annual User Conference*, Houston, TX.

McLachlan, G. J. and Krishnan, T. (1997) *The EM Algorithm and Extensions*, John Wiley, New York.

Oppitz, U. and Pölt, S. (1998) 'Unconstraining and bid pricing', in Hinson, J. (ed.) *Proceedings of 37th AGIFORS Annual Symposium*, AGIFORS, Bali, Indonesia.

Orkin, E. B. (1998) 'Wishful thinking and rocket science: the essential matter of calculating unconstrained demand for revenue management', *Cornell Hotel and Restaurant Administration Quarterly*, **39**, 4, 15–19.

Pölt, S. (2000) 'From bookings to demand: the process of unconstraining', in Busutill, L. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, New York.

Saleh, R. (1997) 'Estimating lost demand with imperfect availability indicators', in Cholak, U. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Montreal, Canada.

Skwarek, D. K. (1996) 'Revenue and traffic impacts of alternative detruncation methods', in Cholak, U. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Zurich, Switzerland.

Smith, B. C., Leimkuhler, J. F. and Darrow, R. M. (1992) 'Yield management at American Airlines', *Interfaces*, **22**, 1, 8–31.

Weatherford, L. R. (1998) 'Forecasting issues in revenue management', in Cholak, U. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Melbourne, Australia.

Weatherford, L. R. (2000) 'Unconstraining methods', in Busutill, L. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, New York.

Weatherford, L. R. and Belobaba, P. B. (2002) 'Revenue impacts of fare input and demand forecast accuracy in airline yield management', *Journal of the Operational Research Society*, **53**, forthcoming.

Zeni, R. H. (2001a) 'Improving forecast accuracy by unconstraining censored demand data', in Ratliff, R. (ed.) *Proceedings of AGIFORS Reservations and Yield Management Study Group*, AGIFORS, Bangkok, Thailand.

Zeni, R. H. (2001b) 'Improved Forecast Accuracy in Revenue management by Unconstaining Demand Estimates from Censored Data', PhD thesis, Rutgers University, Newark, NJ.