Article:

Kourentzes, Nikolaos, Li, Dong orcid.org/0000-0003-3883-8688 and Strauss, Arne (2017) Unconstraining Methods for Revenue Management Systems under Small Demand. Journal of Revenue and Pricing Management. ISSN 1476-6930

https://doi.org/10.1057/s41272-017-0117-x

Online Supplement to "Small Demand Unconstraining Methods for Revenue Management Systems"

Nikolaos Kourentzes, Dong Li, Arne K. Strauss

A Benchmark Unconstraining Methods

Averaging

Weatherford and Pölt (2002) propose a simple averaging method to unconstrain data that can be applied to small and intermittent demand as well. For a given historical booking curve, we divide the time horizon into 10 equal-sized periods and classify each as open if no restriction existed throughout the entire period, or otherwise as constrained. We calculate the average demand received over the open periods. For each closed period, we define estimated demand as the maximum of the observed demand in that period (which may happen if it was only partially constrained) and the average demand over the open periods. The resulting unconstrained total is the sum over all 10 estimations.

Random Walk

Both exponential smoothing methods above are based on the assumption that the observed booking curve, up to the period that the restriction is enforced, contains useful time dynamics that can be modelled. The random walk model, also known as the Naive, would operate on the assumption that all the information is contained in the very last period, i.e. when the restriction is imposed and hence the unconstrained demand is:

$$F_{T+h} = A_T. \tag{A.1}$$

The random walk model has the advantage that it has no parameters to estimate and hence can be used in any circumstances, irrespective of data availability or how many bookings have occurred prior to the restriction period. As such, it can be used as a powerful benchmark for any more complex unconstraining demand methods. We argue that any more complex methods should outperform the naive.

Teunter-Synthetos-Babai Method

Teunter et al. (2011) recognized that a limitation of Croston's method is that it reacts very slowly to information, only when a demand is observed, and therefore does not update its estimate when there are long periods of zero demand. Motivated by an inventory setting, Teunter et al. (2011) argued that for items with long periods of inactivity modelling obsolescence is important and proposed a new Croston-type method for intermittent data. The Teunter-Synthetos-Babai (TSB) method separates the time series into two components, the non-zero demand (z_t) and the probability of a demand event (p_t). The non-zero demand is modelled in the same way as for Croston's method. The probability of a demand event is a vector that is equal to 0 when no demand was observed and equal to one otherwise. This vector is then modelled with single exponential smoothing, resulting in a predicted probability of demand for the future periods. Note that the the demand size estimate updates only when a demand is observed, while the probability of demand updates every period. The final forecast is:

$$f_{T+h} = \hat{z}_T \hat{p}_T. \tag{A.2}$$

When there are long periods of zero demand $\hat{p}_t$ becomes lower, reflecting the higher probability of obsolescence.

Single Exponential Smoothing

Furthermore, SES is a simpler model compared to other demand prediction methods, having half as many parameters. Hence, it requires less data to optimize which is desirable when dealing with sparse booking arrivals. In this context SES is used to model the booking arrivals series and an expected rate of booking arrivals is produced. This is then cumulated in the same way as it was described for Croston's method to unconstrain the demand of the booking curve.

B Data Generation for Accuracy Study

Generation of Booking Curves

We consider 90 days per booking curve with daily Poisson arrival rates shown in Table B1. The rates are defined in a way such that overall total expected demand equals the mean demand figure in the first row. The percentage split of demand over the four time windows is the same in every demand scenario, namely 17%, 36%, 31% and 16%, starting with window 90-30 and ending with period 1-0, respectively. This percentage split has been derived from the car rental data set.

Table B1: Poisson arrival rates for a fixed car group/length-of-rental.

	Mean demand			
Days to pick-up	6	10	14	18
90-30	0.02	0.03	0.04	0.05
30-7	0.09	0.16	0.22	0.28
7-1	0.31	0.52	0.72	0.93
1-0	0.96	1.60	2.24	2.88

We generate a collection of 100 booking curves from this non-homogeneous Poisson distribution for a given mean demand scenario, representing the collection of all available booking histories for the same car group/station/length-of-rental of the same pick-up weekday, say, Monday, during the same season (assuming that there is seasonality over the year).

Generation of Restrictions

To investigate the impact of different degrees of available unconstrained observations, we use the following approach proposed by Queenan et al. (2007): first, we assume that the true final demand is Poisson distributed with means as shown above, namely $\mu := 6, 10, 14$ or 18. Next, for each mean demand scenario, we determine a cutoff value that represents the demand level above that the cumulative Poisson probability sums up to 20%, 40%, 60%, 80% or 100%, respectively. For example, the cutoff value for a 20% restriction level for Poisson-distributed final demand with mean $\mu = 6$ would be 8, i.e. the inverse cumulative Poisson distribution evaluated at 0.8. For each generated booking curve whose final demand exceeds or equals the cutoff value, we subsequently sample a restriction start time from the empirical distribution of restriction start times in the actual data.

We assume that the restriction remains in place until the end of the booking horizon. This assumption is not a particularly strong one since most restrictions in the actual data indeed satisfied that assumption; this is not surprising given that restrictions typically were imposed only shortly prior to the pick-up date.

C Specification of the Simulated RM System

The setup of our RM system aims to replicate the key modules of the system in place at our collaboration partner. We describe the modules that comprise the RM system as shown in Fig 5 in the following.

Demand Generation

We consider demand for a single car group only. Each product represents a combination of pick-up date and length-of-rental. The true booking curve for each product is generated by an non-homogeneous Poisson process as described above. Accordingly, the arrival rates over the entire booking horizon resemble the actual booking curves observed in the airport rental station. Each booking day is divided into three booking segments and thus the 30 days booking horizon is discretized into 90 time periods. The percentage of demand arrivals in each booking period, or the booking curve, is summarised in Table C1. Note that they are different across LoR. Compared to Table B1, we have considered more granular booking processes in the simulation.

The weekly seasonality pattern is reflected by higher demand being generated for weekdays and less for weekends. The mean demand per day is calculated from the actual customer bookings for a particular car group, which are summarized in Tables C2. The distribution of demand for different length-of-rental products reflects the empirical distribution of the same normalized to cover up to 7 days rentals and is

Table C1: Percentage of bookings in each booking period.

Time periods to pick-up	LoR 1	2	3	4	5	6	7
90-63	0.002	0.003	0.003	0.004	0.005	0.005	0.005
62-42	0.004	0.005	0.005	0.006	0.007	0.007	0.007
41-21	0.010	0.009	0.010	0.011	0.011	0.011	0.009
20-6	0.020	0.021	0.022	0.021	0.021	0.018	0.017
5-3	0.037	0.035	0.030	0.025	0.023	0.026	0.031
2-0	0.075	0.068	0.057	0.049	0.041	0.044	0.056

reported in Table C3.

Table C2: Mean demand by day-of-week.

DoW	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Daily Demand	71	96	79	85	86	87	64

Table C3: Demand distribution over length-of-rental.

LoR	1	2	3	4	5	6	7
Percentage	0.25	0.21	0.18	0.14	0.10	0.06	0.07

For each product, demand arrivals within each booking period are generated randomly by a Poisson distribution with the arrival rate obtained for the specific day of week, length-of-rental and booking period. Within one time period there could be multiple arrivals for different products since we sample from a Poisson distribution for each product separately. In the situation of multiple arrivals, their order is determined by random permutation. Since we consider length-of-rentals of up to 7, we generate in total $7H$ booking curves in each simulation run, where H is the simulation's time horizon.

Unconstraining

In unconstraining a particular product, all the historical booking curves of the same day-of-week and length-of-rental are taken into account. Therefore the unconstraining accuracy improves with the simulation progressing as more history becomes available. An initial horizon of 60 days, which forms part of the warm-up phase, is processed without unconstraining so as to have sufficiently many historic booking curves.

Forecasting

Without unconstraining, the observed sales record is used for forecasting future demand; otherwise, we use the unconstrained demand estimates. We have chosen the moving average as forecasting method in the simulation study so as to eliminate further need for parameter optimization. Only the demand histories for the same day-of-week and length-of-rental are used in each forecast. The forecasting module simply uses the average over the corresponding demand estimates provided by the unconstraining module. If no unconstraining is used, it averages the constrained sales records.

Optimization

The RM system's optimization module uses the probabilistic non-linear program (PNLP) suggested for the car rental application by Schmidt (2009). Let $r_{s,l}$ denote the rental rate (price) for the product corresponding to pick up at day s and for length-of-rental l. The random variable $Y_{s,l}$ represents future demand for product (s, l) and $B_{s,l}$ is its booking limit. The fleet available at s is denoted by scalar A_s. The optimization horizon is $N = 30$ days. The PNLP at day t can be stated as follows:

$$\text{PNLP: max} \sum_{s=t}^{t+N} \sum_{l=1}^{7} r_{s,l} u_{s,l} \tag{C.1}$$

$$\text{s.t. } u_{s,l} = \mathbb{E}\left[\min\{B_{s,l}, Y_{s,l}\}\right], \qquad \forall\, s \in \{t, \ldots, t+N\}, \forall\, l \in \{1, \ldots, 7\}, \tag{C.2}$$

$$\sum_{\tau=t}^{s} \sum_{l=s-\tau+1}^{7} B_{\tau,l} \leq A_s, \qquad \forall\, s \in \{t, \ldots, t+N\}. \tag{C.3}$$

Since the demand variable $Y_{s,l}$ is discrete, we can calculate (C.2) by

$$u_{s,l} = \mathbb{E}[\min\{B_{s,l}, Y_{s,l}\}] = B_{s,l} - \sum_{y=0}^{B_{s,l}-1} F_{s,l}(y), \tag{C.4}$$

where $F_{s,l}$ is the cdf function for $Y_{s,l}$. It is obvious that $u_{s,l}$ is an increasing and concave function of $B_{s,l}$. In light of this property and the finite bound for $B_{s,l}$, equation (C.2) can be approximated by a set of piecewise linear functions. Specifically, it can be replaced by the following constraints.

$$u_{s,l} \leq \alpha^i_{s,l} B_{s,l} + \beta^i_{s,l}, \tag{C.5}$$

where $\alpha^i_{s,l}, \beta^i_{s,l}$ are the parameters for the $i^{th}(1 \leq i \leq I)$ linear function for product (s, l).

We next present how to determine these parameters. For a comprehensive account on this process refer to Talluri and van Ryzin (2006). For each product (s, l), sample $I + 1$ booking limit values in between 0 and A_s, denoted by $B^i_{s,l}$. Substitute each of them into equation (C.4) and denote the result by $u^i_{s,l}$. Essentially we have just calculated the expected demand to be accepted for $I + 1$ booking limit values. These $I + 1$ pairs of $(B^i_{s,l}, u^i_{s,l})$ determine I linear functions whose parameters are given by,

$$\alpha^i_{s,l} = \frac{u^{i+1}_{s,l} - u^i_{s,l}}{B^{i+1}_{s,l} - B^i_{s,l}}, \tag{C.6}$$

$$\beta^i_{s,l} = \frac{u^i_{s,l} B^{i+1}_{s,l} - u^{i+1}_{s,l} B^i_{s,l}}{B^{i+1}_{s,l} - B^i_{z,l}}. \tag{C.7}$$

For the perfect information scenario in which the true demand is known in advance, PNLP reduces to a deterministic linear program. The RM system of our partner company uses a PNLP-based optimization module.

Overall, the simulation uses a warm-up phase of 120 days so as to reduce the impact of the initial state of the system on the revenue performance of our unconstraining techniques since they rely on availability of sufficient historic data. Within that warm-up period the booking system uses the $FCFS$ policy to admit bookings until the capacity limit is reached.