hEART 2015 - 4th Symposium of the European Association for Research in Transportation
September 9-11, 2015, Technical University of Denmark (DTU)

A new approach to airline and train demand modeling using registered transactions

Sh. Sharif Azadeh [*] M.Y. Maknoon [*] P. Marcotte [**] M. Bierlaire [*]

[*]Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
[**]University of Montréal

March, 2015

Abstract

The purpose of Revenue Management (RM) is to enhance the profitability of a firm through the optimal management of its inventory. In the service industry (such as, airlines, railways, hotels), this can be achieved by controlling the availability of products, in order to redirect customers to products with high profit margins. Any such strategy is highly dependent on historical demand forecasts, and must cope with the lack of information resulting from censored demand. In this case, the observed demand does not match the true behavior of the customers, and may yield unreliable estimates.

In fact, as soon as a product is no longer available, true demand of airlines or trains may differ from registered bookings, as a result, inducing a negative bias in the estimation results in an artificial increase in demand for substitute products. According to (Weatherford and Belobaba, 2002) ignoring the data censorship phenomenon can lead to demand underestimation ranging from 12.5% to 25%, and negatively affect revenue by 1% to 3%, a significant amount for major rail or airline operators.

Most commonly used uncensoring methods in this context usually try to compensate the missing information by using parametric techniques such as Expectation-Maximization (EM). They are usually costly to be implemented when there is a stochastic process assumption for customer arrival rates at different booking intervals. These customers come from different segments based on their characteristics. In reality, one product could be more appealing for a customer with certain characteristics comparing to other ones from other segments.

In this research, we use aggregated data to predict demand of different products for different customer types. The only information at hand is the number of registered bookings associated with each fare class during different reservation intervals before each departure. Parametric assumptions on demand distribution complicates the process of utility estimation especially when product demands are correlated. Here, we introduce a non-parametric

and distribution-free estimation procedure that, based upon historical bookings, takes explicitly into account the set of available products. Although different approaches have been proposed to model demand under availability constraints there are still issues that need to be addressed: First, demand across fare products is not independent. Dealing with dependency yields a complex parameter estimation process that has been considered and tested by small instances (Stefanescu, 2009). However, while dealing with large volume of aggregate real data, these types of parametric approaches fail to estimate demand properly because of computation burden caused by correlations between different products demand. Second, as the proportion of censored demand in historical data grows, the accuracy of the standard estimation methods decreases (see Vulcano et al., 2012; Talluri and Van Ryzin, 2004; Haensel and Koole, 2011). Finally, not all customers follow the same pattern to make their decisions. This assumption complicates optimization models to estimate demand.

All these issues have motivated us to develop a distribution free demand estimation procedure. In fact, we aim at capturing seasonal effects as well as choice probabilities of different customers with different decision making patterns given registered transactions and the set of available products at a given time.

Hence, we divide the market into distinct customer segments. These segments are defined based on different characteristics of the passengers. The behavior of customers within each segment is homogeneous and they are most likely to respond similarly to a given marketing strategy (e.g. pricing, inventory control). The decision variables of our mathematical model are listed as follows:

- We avoid the parametric assumptions about customers' rate of arrival and parametric expressions of product utilities. As a result, utilities are treated as stand-alone variables inside the mathematical model. Product utilities differ for different customer segments.
- Demand seasonal effect is captured by introducing a potential demand for each day and each customer segment. In fact, the potential demand of a time interval for a given customer segment (and known Origin-Destination) belongs to one of the predefined number of clusters based on the expected demand flow.
- Sets of binary variables are introduced to assign potential demand of a given nest to one of predefined number of clusters as well as using binary variables to assign product utilities to their relevant customer segment (or nest).

We introduce a mixed integer nonlinear mathematical model whose objective function is to minimize the prediction error (the difference between registered transactions and the predicted demand). There are several sources of non-convexities in this mathematical representation: 1) Fractional expression of Multinomial logit (MNL). 2) Product of two decision variables in the objective function (i.e., daily potential demand and choice probabilities) 3) Class membership binary variables. Previously, in a different paper by the authors, similar

Class	Days	Products	MSE	Time	Class (%)	Nest (%)	Sim
2	7	8	8.36	986	0	4.2	49.31
	14	12	42.14	1109	0	2.8	56.77
	21	16	71.69	426	0	4.2	124.9
3	7	8	16.55	1598	4.8	0	34.91
	14	16	30.04	1616	2.4	2.1	27.31
	28	12	53.97	1584	0	0	95.6
4	14	16	17.8	1454	0	6.3	21.61
	28	16	66.99	855	0.9	6.3	87.73

problem had been tackled when the product utilities were assumed to be uniform across all customer segments ((Azadeh, Hosseinalifam and Savard, 2014), (Azadeh, Marcotte and Savard, 2014) and (Sharif Azadeh et al., 2015)). A global optimization approach and interval arithmetic were used inside a branch and bound resolution framework to solve the problem. However, in this paper, as we have an additional assumption about customer segments, the global optimization approach fail to respond. As a result, we use a tailored Adaptive Large Neighborhood (ALNS) Search technique to solve the problem. Two sets of operators are defined: destroy and insertion. The probability of choosing operators varies according to their performance. We use simulated annealing as stopping criterion. We have tested the resolution approach on different sizes of perturbed synthetic data. For our test case, we assume that we have 2, 3 or 4 clusters to which potential demand of each day is assigned. We consider two customer segments and we test our model for different number of departure days (7,14, 21, 28) and products (or fare classes; 8, 12 and 16).

The preliminary results presented above show that the proposed approach succeeds to correctly estimate the underlying demand for different sizes of data. Our evaluation criteria include: 1) Prediction error (the value of the objective function) 2) Classification errors including (assigning a departure day to its proper cluster based on the demand flow and assigning a product to the proper customer segment). We have also chosen an out of sample example to demonstrate the performance of the model (i.e., "sim" column in the table that shows the prediction error). In the next steps, we would like to test the approach for larger samples and compare the outcomes with other methods.

Keywords

Demand Forecasting, Discrete choice models, Optimization, ALNS

References

Azadeh, S. S., Hosseinalifam, M. and Savard, G. (2014). The impact of customer behavior models on revenue management systems, *Computational Management Science* **12**(1): 99–109.

Azadeh, S. S., Marcotte, P. and Savard, G. (2014). A taxonomy of demand uncensoring methods in revenue management, *Journal of Revenue & Pricing Management* **13**(6): 440–456.

Haensel, A. and Koole, G. (2011). Estimating unconstrained demand rate functions using customer choice sets, *Journal of Revenue & Pricing Management* **10**(5): 438–454.

Sharif Azadeh, S., Marcotte, P. and Savard, G. (2015). A non-parametric approach to demand forecasting in revenue management, *Computers and Operations Research* **In press, accepted manuscript**.

Stefanescu, C. (2009). Multivariate customer demand: modeling and estimation from censored sales, *Available at SSRN 1334353* .

Talluri, K. and Van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior, *Management Science* **50**(1): 15–33.

Vulcano, G., Van Ryzin, G. and Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data, *Operations Research* **60**(2): 313–334.

Weatherford, L. and Belobaba, P. (2002). Revenue impacts of fare input and demand forecast accuracy in airline yield management, *Journal of the Operational Research Society* pp. 811–821.