

# A Demand Estimation Algorithm for Inventory Management Systems Using Censored Data

Alireza Nikseresht

Computer Science and Engineering Department  
Shiraz University  
Shiraz, Iran  
anikseresht@cse.shirazu.ac.ir

Koorush Ziarati

Computer Science and Engineering Department  
Shiraz University  
Shiraz, Iran  
ziarati@shirazu.ac.ir

**Abstract**—During the selling time horizon of a product category, a number of products may become unavailable sooner than others and the customers may substitute their desired product with another or leave the system without purchase. So, the recorded sales do not show the actual demand of each product. In this paper, a nonparametric algorithm to estimate true demand using censored data is proposed. A customer choice model is employed to model the demand and then a nonlinear least square method is used to estimate the demand model parameters without assuming any distribution on customer's arrival. A simple heuristic approach is applied to make the objective function convex, making the algorithm perform much faster and guaranteeing the convergence. Simulated dataset of different sizes are used to evaluate the proposed method. The results show a 23% improvement in root mean square error between estimated and simulated true demand, in contrast to alternate methods usually used in practice.

**Keywords**—demand; estimation; inventory; revenue; inventory; management; control; unconstraining; uncensoring

## I. INTRODUCTION

Modeling demand and estimating the true demand is crucial in every inventory management system. Inventory management systems rely on accurate demand estimating methods [1]. In inventory management systems the demands are usually observed in several periods during a selling horizon. Due to inventory constraints the demand is truncated in some periods, so only censored sales data are recorded. Using this kind of data as true demand causes underestimation or overestimation of true demands. If the companies overestimate the true demand, then they may stock more than the needs (overstock), so they eventually are forced to mark down the prices to sell those merchandises, especially when the goods are perishable or seasonal. If they underestimate the true demand, they may stock less than the needs (understock), in this situation they not only missing out on sales but also losing discount from their suppliers and customer loyalty, as product availability is an important factor of customer satisfaction [2]. In both situations the companies loose revenue. Ignoring the correlation of demand between related products leads to inefficient estimates of the true demand [3, 4]. Empirical studies show that an amount between 45% and 84%

of demand can be substituted [5, 6]. In [7], it is showed that the cost of ignoring demand dependency is 1% higher than ignoring censorship. To illustrate this problem, suppose that an online library offers 2 types of books for each ISBN: 1-new book, 2-used books and their quantities are 30 for new books and 15 for used books. In this context, assume that 20 customers arrive and want to buy used books. However, due to limited stock, just 15 customers will buy their preferable product, and the other 5 customers should do something else. In such a situation, the current information systems just registers the successful sales and does not record any information about unsuccessful requests, so the observed sales are a censored view of true demand. Therefore, in such cases, the sales data could not show the true demands. For example, here the sales data show 15 requests (the observed demand), but the true demand was 20 which is not visible. The observed sales of new books are 20, again this observed sales could not show the true demand because some customers bought new books as a substitute for used books, so the observed sales of new books are a mixture of true and substitute demand. As foregoing example shows, the substitution will increase sales in substitute products which are available. In the rest of the paper this phenomenon is referred as recaptured demand. Nevertheless, measuring true demands by using the available sales data is not an easy task [8]. If the accuracy of forecasting grows about 20%, the resulting revenue will improve about 1% [9]. In [10], it was found that if we have a negative bias in forecast, up to 3% of the potential profit may be lost. If the data remain censored and we use them as the true demands, the true demands will be underestimated, causing a spiral down effect over time [11-13]. As a result, it is rational to look forward to methods measuring the true demands from the available data and use these data for forecasting and managing inventory [14]. The process of finding true demand, which in literature is called unconstraining, fills the gap between the true and the observed demands, and its goal is to provide accurate data required for the forecasting module [15]. In this paper, we propose a method in order to unconstrain censored data or in other word estimate true demand using sales transaction data in a fast, accurate, and very simple manner. The proposed algorithm is an extension of RDFSO algorithm which is an iterative demand estimation algorithm considering stockout [16]. As RDFSO, our method just needs transactional sales

data and product availability, and does not assume any special distribution for customer arrival, thus, it is non-parametric and makes use of the advantages of not requiring any assumption on the form of the distribution function. The proposed method uses a simple heuristic approach to convexify the objective function, thus the algorithm is fast and always converge. By estimating substitute demand as well as true demand, the problem of demand double counting is avoided. The multinomial logit choice model is employed to model the demands, and the choice model parameters are estimated by applying an algorithm which tries to reduce the error between the observed and the estimated sales data in each iteration. Two types of datasets are used for evaluation: 1. A dataset with full observation of demand occurrences and 2. A dataset with censored observation of demands. Dataset 1 which contains unconstrained observation is served as a benchmark to evaluate our work.

## II. LITERATURE REVIEW

One of the most popular approaches to estimate substituted demands is the multinomial logit model (MNL) and expectation maximization (EM) is often employed to find the customer choice model parameters. Authors in [17] applied a discrete choice model to model customer behavior and found the related parameters using EM method. In their proposed method, they assumed a distribution function for demand and then tried to find its parameters from the observed data, indicating that their methods perform well. Authors in [18, 19] utilized the idea of customer choice set to model the buying behavior of customers. They applied EM algorithm to unconstrain the censored data, not assuming a fixed arrival rate for customers as is regularly done. Instead, they estimated a demand function for each group of customers by analyzing observed data. They tested their method on real data, demonstrating that their proposed method just overestimates 5% and takes 10 to 15 iterations to converge, which makes it suitable for real-world application. Authors in [20] proposed an EM-based method estimating spilled and substitute demands. Their method only needs the observed sales data, product availability, and company market share. Their main idea is to consider the problem as primary demand or customer first choice. They supposed that each customer has a set of choices with a primary or first choice. Each customer could buy his/her first choice, and if the first choice is not available, the next choice is substituted or leaves the system without purchasing anything. Then they tried to estimate primary demands, substitute demands, and no purchase count, using EM. In revenue management setting, authors in [21] employed a Double Exponential Smoothing or Holt Winter-based method to find true demand using the observed sales. They compared their method with naïve methods, PD, and EM and showed that their method outperforms the others. Modeling customer choice is a high dimensional problem and it is difficult to dealing with, so to deal with its difficulty, many researchers prefer to assume a priori distribution and parametric model which they think is able to adequately capture choice model behavior [22]. The side effects of all parametric approaches are misspecification of the model and overestimation or underestimation of true demand. Authors in [23] studied an assortment planning model with substitutable products. They

considered three types of substitution: 1-stockout based 2-assortment-based 3-the customers select their favorite item to buy from what they see on shelves. It is also assumed that if a customer could not buy his/her favorite item, he/she will buy a substitute product or leave the store. By this assumption, the number of substitutions is limited and the problem is made tractable. Then, a procedure for estimating substitution parameters is proposed with a heuristic approach for solving the assortment planning problem. Authors in [24] applied a stochastic newsvendor model for perishable and non-perishable items that jointly optimizes assortment and associated order volumes considering storage constraints. Their method is able to find the efficient solution for ordinary problem sizes in retail using an optimal procedure and in case of large problem instances a heuristic is applied to find near optimal solutions in reasonable time. Using numerical tests, they showed that their heuristic produces near optimal solutions and outperforms heuristic of authors in [23]. They also revealed that considering substitution effects has a significant impact on the total revenue and solution structure. Authors in [25] proposed a method able to learn demand using continuous observation of sale occurrences in a Bayesian fashion. They have considered 3 types of observations: 1. Sales along with a binary variable which shows the product availability status. 2. Continuous observation of sales until the stockout happens. 3. Full observation of the demand. They showed that in case of continuous observation, a firm optimal ordering quantity is greater than observation type 1, then with numerical results they found that in continuous observation 76.1% of loss in expected profit eliminates in contrast to type 1 of observation. They assumed that the demand is independent, so they do not consider product substitution in their work. Our paper brings two main contributions to the literature. First, our work contributes to the literature by introducing an easy-to-implement non-parametric estimation algorithm for estimating multinomial logit choice model parameters in a reasonable computation time and accuracy. Second, we also have devised a heuristic to convexify the objective function which shortens the execution time and guarantees the convergence.

## III. PROBLEM DESCRIPTION

Demand estimation under the customer-driven substitution scheme is studied in this paper. In the customer-driven substitution scheme, when the first-choice product of the customer is not available on the shelf, the customer may purchase, with a certain probability, another product in the same category in lieu of his/her first-choice product. For example, market A offers a category which contains 4 items. Now suppose that a customer comes to buy a product in that category, and with respect to features such as price, weight, production date etc he/she prioritizes the items and chooses an item. In other words, each customer faces a set of choices, and then he/she arranges these choice sets based on their benefits and chooses to buy the first item of that sorted choice set. Suppose that customer B wants to buy a bottle of milk, then faces the choice offered by Retailer A (Table I). He arranges the items as C1-C2-C3-C4, meaning that the highest priority is buying the product C1, then C2, and so on. If a customer could not buy product C1, two conditions occur: the customer buys the other product (substitute), C2, or leaves the system without

purchase (denial). It is obvious that the number of sales of products C1 and C2 could not show the real demand. If these data are used as true demands and are fed into the forecasting module, a spiral down effect will occur in estimation, and the total revenue begins to decrease [11]. Retailers are interested in recognizing how customers prioritize the products offered and select an item to buy.

TABLE I. PRODUCTS AND THEIR FEATURES OFFERED BY RETAILER A IN A CATEGORY

Items	Weight(liter)	Production date	Fat	Price \$
C1	1.5	2016-06-01	1%	1.2
C2	1	2016-06-01	1%	1
C3	1	2016-05-30	2%	1
C4	0.85	2016-06-01	3%	0.75

IV. DEMAND MODEL

We consider a firm that presents n products. Selling time horizon is discretized and, indexed by  $t = 1, \dots, T$ , where T is the number of periods or time frames between the time when a product is presented for sale and the time it expires. The periods could be different in length, for example period 1 may be a day, but period 2 may be 12 hours. A set of products ( $C_t$ ) is offered in period t:  $C_t = (c_{1t}, c_{2t}, \dots, c_{nt})$ .  $o_{jt}$  is the number of observed sales of product j in the interval t, and  $O_t$  is the observed sale vector in interval t:  $O_t = (o_{1t}, o_{2t}, \dots, o_{nt})$ . In our model each product  $c_i$  has a fixed preference weight  $v_i$  which customers sort the products based on the product preference weight, then choose the first one from the list if it was available. The preference weight of no purchase is  $v_0 = 1$ . We assumed that the preference vector is constant in each period for all customers. Using these preference weights we rewrite the MNL choice probability as in [26]:

$$P(j, t) = \frac{v_j}{\sum_{i \in c_t} v_i + v_0} \tag{1}$$

The probability of choosing product j which is not available in period t is:  $P(j, t) = 0$ . Based on no purchase preference weight or  $v_0 = 1$  the no purchase probability is defined as:

$$P(0, t) = \frac{v_0}{\sum_{i \in c_t} v_i + v_0} = \frac{1}{\sum_{i \in c_t} v_i + 1} \tag{2}$$

We are able to calculate the number of purchases in each period using the number of customers arriving in period t or  $A_t$ . The probable number of sales of product j in period t if all products are available (true demand) is as follows:

$$d_{jt} = P(j, t) \cdot A_t = \frac{v_j}{\sum_{i=1}^n v_i + v_0} A_t \tag{3}$$

Using the estimated number of sales, we could calculate total sale in each period t as:

$$E_t = \sum_{j \in c_t} d_{jt} = \sum_{j \in c_t} \frac{v_j}{\sum_{j \in c_t} v_j + v_0} A_t \tag{4}$$

By using (4), we can estimate the number of arrivals:

$$q_t \approx E_t \Rightarrow q_t \approx \sum_{j=0}^n d_{jt} = \frac{\sum v_j}{\sum v_j + v_0} A_t \Rightarrow A_t \approx q_t \frac{\sum v_j + v_0}{\sum v_j} \tag{5}$$

The number of customers who could not buy their first choice in period t is estimated by using:

$$s_t = \sum_{j \notin c_t} d_{jt} \tag{6}$$

The number of requests which choose class k as a substitute in period t, could be estimated by the means of multiplying  $S_t$  (the number of customers who could not buy their first choice) by the probability of choosing class k:

$$r_{kt} = s_t \cdot P(k, t) = s_t \cdot \frac{v_k}{\sum_{j \in c_t} v_j + v_0}, k \in c_t \tag{7}$$

The number of sales of product j in period t or  $o_{jt}$  consists of three components: 1- the number of customers interested in product j as their first choice (true demand or  $d_{jt}$ ), 2- the number of customers buying j because their first choice was not available (recapture or  $r_{jt}$ ), and 3- the number of customers unable to buy product j as their preferred product (spill or  $S_{jt}$ ) [27, 28]. Thus, the demand mass balance equation is as follows:

$$o_{jt} = d_{jt} + r_{jt} - S_{jt} \tag{8}$$

Equation (6) is used to compute spill and recapture. It is also used to estimate the observed sales. It should not be forgotten that at first we need the preference vector V to be able to compute spill and recapture. Therefore, at first we try to estimate the preference vector, and by using it along with foregoing formulas the number of true demands, recaptures, and spills is estimated.

V. NPTDEA ALGORITHM

The nonparametric true demand estimation algorithm or NPTDEA is an iterative algorithm which starts with an initial value for preference vector and tries to make a better estimate for preference vector in each iteration by minimizing an error function which is the gap between estimated and observed sales in each period:

$$e = o_{jt} - (d_{jt} + r_{jt} - s_{jt}) \tag{9}$$

The objective function which should be minimized is:

$$\min_{v_j} (E = \frac{1}{2} \cdot \beta^2) \tag{10}$$

A. Analysis of Objective Function

To show the objective function is non-convex, assume a firm offer two products A and B in ten periods. At the first 6 periods both of the products were available, but later only product B was available. Based on foregoing equations the

error function in the first 6 periods which all products were available (so no spill and recapture exist) becomes:

$$\beta = \left( x[t,1] - v_1 * \frac{\sum_{i=1}^2 x[t,i]}{v_1 + v_2} \right)^2 + \left( x[t,2] - v_2 * \frac{\sum_{i=1}^2 x[t,i]}{v_1 + v_2} \right)^2 \tag{11}$$

Also the hessian matrix is not positive definite. The 3D plot of (11) is shown in Figure 1. In the last 4 periods which product A is not available, because of recapturing spilled demands, the error function is like:

$$\beta = (x[t,2] - (v_2 + 1) * \frac{\sum_{i=1}^2 x[t,i]}{v_1 + v_2} - v_1 * (v_2 + 1) * \frac{\sum_{i=1}^2 x[t,i]}{v_1 + v_2 + 1})^2 \tag{12}$$

The value of  $v_1$  is not updated as product A is not available, so we assume the value of  $v_1$  is set to its optimal value, then we plot the error function by fixing  $v_1$  at its optimal value. The plot is shown in Figure 2 and obviously is not convex and it is ill structured.

**B. Convexification Process**

Because of the fractional terms in the demand equation (3) the objective function is not convex. We made a simple heuristic to make that function convex. As we said before, the nonconvexity is because of the existence of the fractional term in the demand function, so if we replace the denominator ( $\sum v_j + v_0$ ) with a fixed value, the objective function becomes convex. This value could be extracted from market share. We have considered the market share the same as that in [29]:

$$M = \frac{\sum v_j}{\sum v_j + v_0} \xrightarrow{v_0=1} \sum v_j = M / (1 - M) \tag{13}$$

Figure 3 shows the error function after convexification. The initial value of preference weight for product j could be found as:

$$v_j = \left( \frac{M}{1 - M} \right) / \text{number of products} \tag{14}$$

**C. Pseudo Code of NPTDEA:**

1. Ct: Set of available products in period t.
2. V: Preference weight vector.
3. U: Utility of products vector.
4. At: The number of arrival in period t.
5. Ot: The vector of purchases in period t.
6. M: The market share (if it is available)
7. Initialize :
  - Estimating  $v_j$
  - For each product j in [1..n]

$$v_j = M / (1 - M) / n$$

8. For each period t in [1..T]
  - Compute  $A_t$  based on  $v_j$ s:

$$A_t = q_t \frac{\sum v_j + v_0}{\sum v_j} \quad A_t = q_t \frac{\sum v_j + v_0}{\sum v_j}$$

Compute no purchase:  $d_{ot} = A_t - q_t$

For each product j in [1..n]

$$d_{jt} = \frac{v_j}{\sum v_i + v_0} A_t$$

if ( $c_{jt}$  not available)  
 $V_{j\_new} = V_{j\_old}$   
 $S = S + d_j$

else

$$r_j = s \frac{v_j}{\sum_{j \in c_t} v_j + v_0}$$

$$V_{j\_new} = \text{mean} \left( \frac{o_{jt}}{(A_t / \sum v_j) + (s / \sum_{j \in c_t} v_j + v_0)} \right)$$

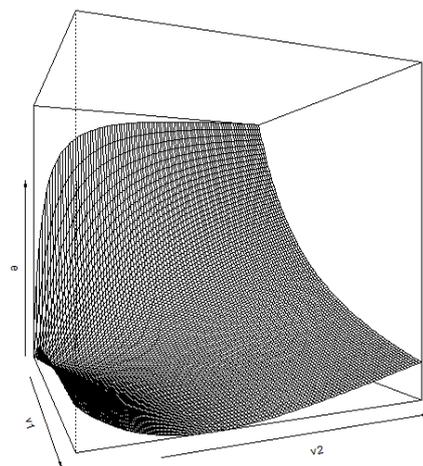


Fig. 1. The 3D plot of equation 11 (error function before convexification)

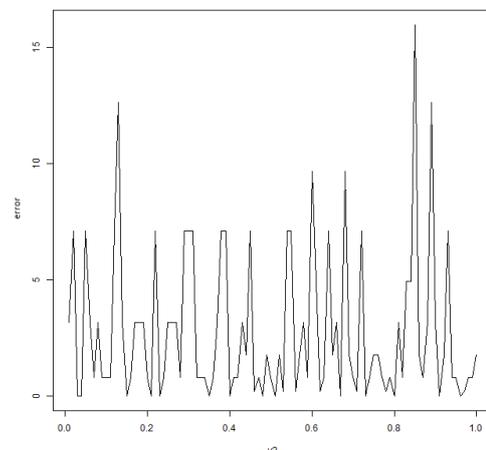


Fig. 2. Plot of (12) by fixing the  $v_1$  at its optimal level.

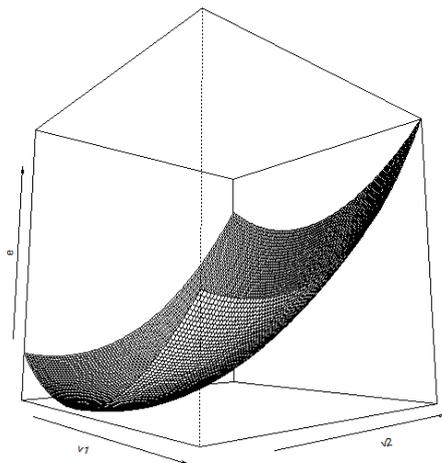


Fig. 3. 3D plot of error function after convexification.

VI. SIMULATION AND RESULTS

A. Simulation Process and Data

In this section the simulation process is described and the data used for the evaluating of the proposed algorithm. The simulation process is as follows:

- i. Select a customer choice model. This is a predefined model which some products and their preference weights. The market share and customer arrival pattern is also defined. This model may pick from real-world data or synthetic datasets. For example, we have a model with products c1 to c5 and the preference vector of  $V = (0.299, 0.199, 0.122, 0.092, 0.057)$ . The market share is 43% and the customer arrival in each period is simulated with a Poisson process with the mean of 80. The no purchase preference weight is  $v_0 = 1$  and the number of periods is 15.
- ii. Use the selected model to simulate the purchase of products in each period. By having the simulated number of arrivals in each period, we are able to simulate the number of sales for each product and the number of no purchase using (3).
- iii. Apply product availability matrix on simulated sales from step ii. The product availability info is in matrix format, with the number of columns equal to the number of periods and the number of rows equal to the number of products. For example  $AVL_{i,j} = 1$  means that the product  $i$  is available in period  $j$  and vice versa. This information could be available exogenously or filled randomly. By applying the product availability on simulated data, if the product  $i$  was not available on period  $j$  (ie.  $AVL_{i,j}=0$ ) the simulated sales of that product on foregoing period reset to zero (ie.  $d_{i,j} = 0$ ).
- iv. By executing NPTDEA algorithm on the generated dataset the preference vector is estimated.
- v. Using estimated preference vector we are able to find the true demand and to show the accuracy of proposed algorithm when the estimated true demand and simulated true demand are compared.

A sample of the simulated dataset is shown in Table II. It is shown that in the periods 1 to 3, all products are available and in the period 1 the number of sales for products C1 and C2 is 26 and 17, and so on. In the period 1, the number of arrivals to the system is 84, which is generated by a Poisson distribution function with mean=80. During that period, 58 customers out of 84 customers buy the products, while the other 26 leave the system without purchase. The number of arrivals and no purchase is not visible in real-world application. Thus, we separated these values in two lines entitled Hidden Data. Each dataset contains 100 instances of selling horizons, which each instance contains a category of items and the number of sales of each product in that category on every period. For example, in our 5X10 dataset which contains a category of 5 products and 10 intervals, we have 100 instances of sales. This means  $100 \times 10 \times 5$  intervals or 5000 sales data records.

TABLE II. A SAMPLE OF SIMULATED DATA FOR A SPECIFIC CATEGORY.

Pro ducts	C1	C2	C3	C4	C5	Hidden Data	No Purchases	#Arrivals
1	13	11	1	4	5		42	76
2	15	6	8	9	0		42	80
3	13	7	3	4	3		43	73
4	NA	9	5	7	5		67	93
5	NA	10	8	7	2		56	83
6	NA	NA	4	3	2		74	83
7	NA	NA	9	9	4		71	93
8	NA	NA	NA	6	1		67	74
9	NA	NA	NA	7	9		74	90
10	NA	NA	NA	NA	3	81	84	

NA means the product is not available in that period

B. Results

As a result of applying the NPTDEA algorithm on simulated dataset, preference vector  $V$  is estimated. Table III shows a sample of estimated preference vector in contrast to the true value of preference vector which is used to generate the simulated dataset. There are two favorable attributes for an estimator: accuracy and precision. Accuracy is lack of bias and precision is small variance. If an estimator is unbiased, then we just look at its variance. If it is biased we sometimes look at the mean squared error. As the Table III shows, our estimator has a negative bias on estimated values and this is because of changing the objective function to a convex form and the new convex function is always below the true values. Table III shows the mean square error and variance of proposed algorithm for estimating the values of preference vector. As Table III shows the values of MSE and Variance are almost equal and this is because of the mean of the estimated preference vector is almost equal to the true value of preference vector. Figure 4 exhibits the box plot of estimated preference values for 100 algorithm executions over simulated datasets. The line in the middle of the box is the median. The box itself represents the middle 50% of the data. The box edges are the 25th and 75th percentiles. There is some disk visible below or above of some boxes which shows that in some cases the

simulated sales were far from the regular values, so as a result the estimated preference of that simulated dataset is far from the median. Figure 5 shows the cumulative sum of simulated demand, estimated demand, observed sales and estimated sales over 15 periods for products c1 to c4.

TABLE III. MEAN, VARIANCE AND MEAN SQUARE ERROR OF DIFFERENCE BETWEEN ESTIMATED AND TRUE VALUES OF PREFERENCES.

Products		C1	C2	C3	C4	C5
Preferences		V1	V2	V3	V4	V5
TRUE	Value	0.299	0.199	0.122	0.092	0.057
Estimated	Value	0.286	0.197	0.121	0.091	0.056
Mean		-0.0136	-0.0029	-0.0004	-0.0015	-0.0003
Variance		0.0012	0.0007	0.0004	0.0002	0.0001
MSE		0.0014	0.0007	0.0004	0.0002	0.0001

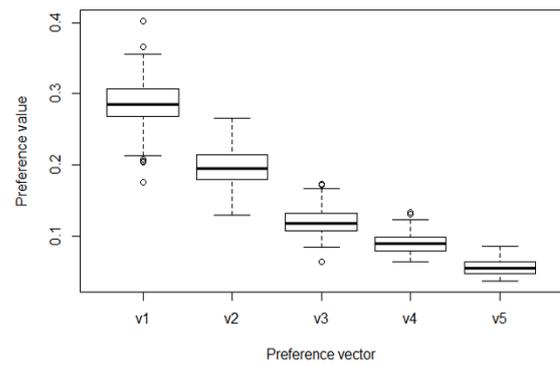


Fig. 4. The Boxplot of preference values.

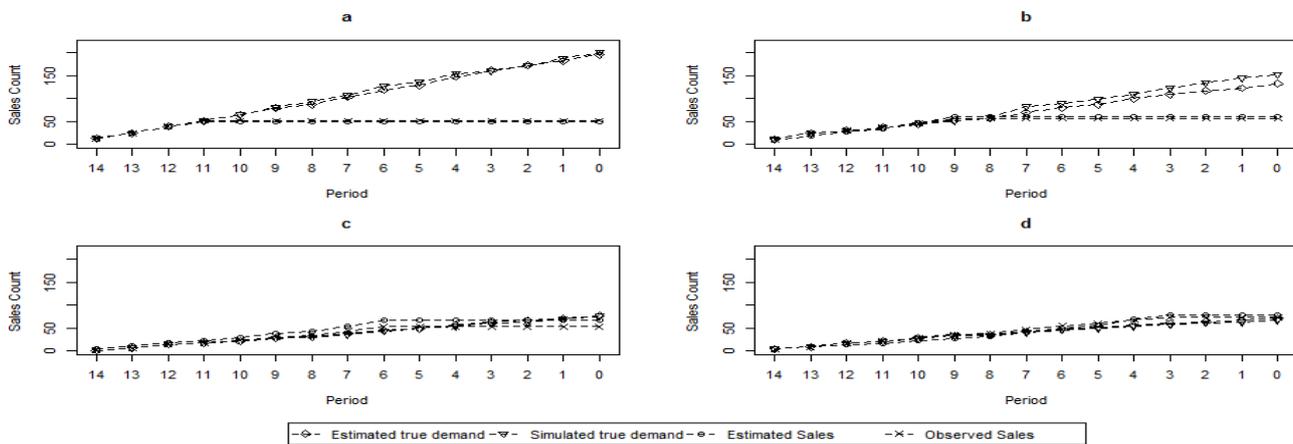


Fig. 5. The cumulative sum of simulated and estimated demand, observed sales and estimated sales over 15 periods for products c1(a), c2(b), c3(c) and c4(d)

It is obvious that estimated sales are always close, but a bit upper than observed sales. One of the important aspects of the simulation is that even hidden data not observable in the real world are generated and thus the accuracy of the estimation process can be evaluated better. To gain a better insight of the results and of the accuracy of the proposed method we have compared the results against two other methods which are usually applied in practice: 1-expectation maximization or EM and 2-Projection Detruncation or PD methods [15, 30, 31]. To compare the results we have applied the three above methods on a dataset for 1000 times, then the mean of root mean square error for each method was calculated. The result of the comparison is depicted in Table IV. The results show that our proposed method outperforms EM and PD and has about 21% to 23% improvement in root mean square error compared to the other two methods.

TABLE IV. ROOT MEAN SQUARE ERROR BETWEEN ESTIMATED TRUE DEMAND AND SIMULATED TRUE DEMAND IN CONTRAST TO EM AND PD.

Methods	C1	C2	C3	C4
NPTDEA RMSE	3.19	4.05	1.53	0.68
EM RMSE	3.28	4.15	2.35	2.5
PD RMSE	3.28	4.21	2.14	2.46

TABLE V. EXECUTION TIME FOR DIFFERENT DATASETS.

Dataset	Execution Time for 100 instance(sec)
5X15	2.01
5X30	3.72
10X30	4.70
10X60	9.12
20X60	12.77

Different size datasets are used to show the efficiency and speed of the proposed algorithm. At first we evaluated it with a rather small dataset of 5 products and 15 periods. In this case the algorithm lasts 2.01 seconds to estimate 100 instances. The second dataset is composed of observed sales from a category with 5 products and 30 time intervals. The third data set is in a category having 10 products sold in 30 intervals. Our last and bigger dataset consisted of 20 products and 60 time intervals. For the last dataset the algorithm takes about 12.77 seconds to solve 100 instances, which is very good for this size of dataset in contrast to current demand estimation methods. Table V shows the execution time for all five datasets. R revision 3.2.2 was used to implement the algorithm on a computer with an Intel core2 Quad Q8300 and 4 GB of internal memory. Code execution time is measured by proc.time() in R.

## VII. CONCLUSION

This paper presented a new approach for the estimation of true demand. The demand is modeled using the customer choice model. Transactional sales data, product availability and market share are the only data used to estimate the choice model parameters, which are customer preferences in a set of substitutable products. Spilled and recaptured demands are estimated alongside with true demand. The problem of double counting demands in available products is resolved by considering spill and recapture. Most current methods assume an a priori distribution to estimate customer arrival in each period, so they suffer from misspecification of the distribution function and its parameters. Unlike these methods, the proposed method is nonparametric and estimates customer arrival in a simple manner. Numerical experiments confirm that proposed method performs well in terms of speed and estimation accuracy. Based on the simulation results, our method improved root mean square error between simulated and estimated demands by 23%, in contrast to other methods such as PD and EM. For a rather large dataset, it estimates the true demand, spill and recapture rather fast. We believe that the proposed method could be a good replacement for current demand estimation methods in practice.

## REFERENCES

- [1] N. Agrawal, S. A. Smith, "Estimating negative binomial demand for retail inventory management with unobservable lost sales", *Naval Research Logistics*, Vol. 43, No. 6, pp. 839–861, 1996
- [2] C. Eltze, S. Goergens, M. Loury, "Grocery store operations: Which improvements matter most?", *Akzente*, Vol. 1, No. 1, pp. 74–81, 2013
- [3] C. Stefanescu, "Multivariate Demand: Modeling and Estimation from Censored Sales", Available at: <http://ssrn.com/abstract=1334353>, 2009
- [4] B. Tan, S. Karabati, "Retail inventory management with stock-out based dynamic demand substitution", *International Journal of Production Economics*, Vol. 145, No. 1, pp. 78–87, 2013
- [5] J. Aastrup, H. Kotzab, "Analyzing out-of-stock in independent grocery stores: an empirical study", *International Journal of Retail & Distribution Management*, Vol. 37, No. 9, pp. 765–789, 2009
- [6] G. Xin, P. R. Messinger, J. Li, "Influence of soldout products on consumer choice", *Journal of Retailing*, Vol. 85, No. 3, pp. 274–287, 2009
- [7] S. Karmarkar, D. Goutam, B. Tathagata, "Revenue impacts of demand unconstraining and accounting for dependency", *Journal of Revenue and Pricing Management*, Vol. 10, No. 4, pp. 367–381, 2011
- [8] R. R. Wickham, Evaluation of forecasting techniques for short-term demand of air transportation, MSc Thesis, Massachusetts Institute Technology, 1995
- [9] K. T. Talluri, G. J. van Ryzin, I. Z. Karaesmen, G. J. Vulcano, "Revenue management: Models and methods", 2008 Winter Simulation Conference, pp. 145–156, 2008
- [10] L. R. Weatherford, "A review of optimization modeling assumptions in revenue management situations", AGIFORS Reservations and Yield Management Study Group, 1997
- [11] W. L. Cooper, T. Homem-de-Mello, A. J. Kleywegt, "Models of the Spiral-Down Effect in Revenue Management", *Operations Research*, Vol. 54, No. 5, pp. 968–987, 2006
- [12] R. Saleh, "Estimating lost demand with imperfect availability indicators", AGIFORS Reservations and Yield Management Study Group, 1997
- [13] R. J. A. Little, D. B. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, 2002
- [14] A. Nikseresht, K. Ziarati, "Review on the Newest Revenue Management Demand Forecasting Methods", *International Conference on Management, Economics and Industrial Engineering*, Vol. 1, No. 1, 2015
- [15] P. Guo, B. Xiao, J. Li, "Unconstraining methods in revenue management systems: Research overview and prospects", *Advances in Operations Research*, Vol. 2012, Article ID 270910, 2012
- [16] A. Nikseresht, K. Ziarati, "Estimating True Demand in Airline's Revenue Management Systems using Observed Sales", *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 7, pp. 361–369, 2017
- [17] K. T. Talluri, G. J. Van Ryzin, *The Theory and Practice of Revenue Management*, Springer Science & Business Media, 2005
- [18] A. Haensel, G. Koole, "Estimating unconstrained demand rate functions using customer choice sets", *Journal of Revenue and Pricing Management*, Vol. 10, No. 5, pp. 438–454, 2011
- [19] A. Haensel, G. Koole, J. Erdman, "Estimating unconstrained customer choice set demand: A case study on airline reservation data", *Journal of Choice Modelling*, Vol. 4, No. 3, pp. 75–87, 2011
- [20] G. Vulcano, G. van Ryzin, R. Ratliff, "Estimating Primary Demand for Substitutable Products from Sales Transaction Data", *Operations Research*, Vol. 60, No. 2, pp. 313–334, 2012
- [21] C. C. Queenan, M. Ferguson, J. Higbie, R. Kapoor, "A Comparison of Unconstraining Methods to Improve Revenue Management Systems", *Production and Operations Management*, Vol. 16, No. 404, pp. 729–746, 2007
- [22] V. F. Farias, S. Jagabathula, D. Shah, "A Nonparametric Approach to Modeling Choice with Limited Data", *Management Science*, Vol. 59, No. 2, pp. 305–322, 2009
- [23] A. G. Kok, M. L. Fisher, "Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application", *Operations Research*, Vol. 55, No. 6, pp. 1001–1021, 2007
- [24] A. Hubner, H. Kuhn, S. Kuhn, "An efficient algorithm for capacitated assortment planning with stochastic demand and substitution", *European Journal of Operational Research*, Vol. 250, No. 2, pp. 505–520, 2016
- [25] A. Jain, N. Rudi, T. Wang, "Demand Estimation and Ordering Under Censoring: Stock-Out Timing Is (Almost) All You Need," *Operations Research*, Vol. 63, No. 1, pp. 134–150, 2015
- [26] L. G. Cooper, M. Nakanishi, *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness*, Kluwer Academic Publishers, 1988
- [27] S.-E. Andersson, "Passenger choice analysis for seat capacity control: A pilot project in Scandinavian Airlines", *International Transactions in Operational Research*, Vol. 5, No. 6, pp. 471–486, 1998
- [28] S. Ja, S. Rao, B. V. Chandler, "Passenger recapture estimation in airline RM", AGIFORS 41st Annual Symposium, 2001
- [29] T. S. Gruca, D. Sudharshan, "Equilibrium Characteristics of Multinomial Logit Market Share Models", *Journal of Marketing Research*, Vol. 28, No. 4, pp. 480–482, 1991
- [30] R. H. Zeni, *Improved Forecast Accuracy in Airline Revenue Management by Unconstraining Demand Estimates from Censored Data*, Universal-Publishers, 2001
- [31] Tarek Abdallah, G. Vulcano, "Demand Estimation under the Multinomial Logit Model from Sales Transaction", working paper, Available at: [https://www.researchgate.net/profile/Gustavo\\_Vulcano](https://www.researchgate.net/profile/Gustavo_Vulcano), 2016