Please cite this article as: Derhami, S., & Montreuil, B. (2021). Estimation of potential lost sales in retail networks of high-value substitutable products, *IISE Transactions*, doi: 10.1080/24725854.2021.1969484

Estimation of potential lost sales in retail networks of high-value substitutable products

Shahab Derhami[a,*], Benoit Montreuil[b,c]

[a]*School of Management, Binghamton University, State University of New York, NY 13902, USA*
[b]*School of Industrial and Systems Engineering, Georgia Institute of Technology, GA 30332, USA*
[c]*Physical Internet Center and Supply Chain Institute, Georgia Institute of Technology, GA 30332, USA*

Abstract

Sales data reveal only partial information about demand due to stockout-based substitutions and lost sales. We develop a data-driven algorithm to estimate stockout-based lost sales and product demands in a distribution network of high-value substitutable products such as cars, using only past sales and inventory log data and product substitution ratios. The model considers the particular customer and retailer behaviors frequently observed in high-value product markets, such as visiting multiple stores by customers for a better match and exploiting on-demand inventory transshipments by retailers to satisfy the demand for out-of-stock products. It identifies unavailable products for which a retailer could not fulfill demand and estimates the potential lost sales and the probability distribution of product demands for the potential lost sales using sales data in retailers with similar sales profiles while considering retailers' market sizes. We validate the results of our algorithm through field data collection, simulation, and a pilot project for a case of recreational vehicles. We also show the result of implementing our model to estimate lost sales across the large retail network of a leading vehicle manufacturer. Our case study shows sales data significantly underestimate the demand for most products.

Keywords: demand estimation, lost sales, product demand, stockout-based substitution, inventory transshipment, vehicle dealerships

1. Introduction

Most retailers and manufacturers regard stockouts and incomplete product availability as inevitable for reasons such as demand uncertainty, forecast errors, seasonality, high holding costs, broad product mixes, limited production capacity, long order-to-delivery times, and storage space limitations at retail outlets. Stockouts and incomplete product availability frequently occur in firms that distribute perishable and high-value products that are offered in high variety. Empirical studies show the customers are typically inclined to substitute a similar product for their desired one when it is not in stock (Fitzsimons, 2000; Gruen et al., 2002; van Woensel et al., 2007). This means that estimating customers' demands from sales transactions is biased if customers' substitution behavior and the availability of products at the time of sale are not considered (Anupindi et al., 1998; Jain et al., 2015). There are two sources of bias in this case. The first is the potential lost sales induced by customers deciding not to buy when their desired products are not available and none of the available products meet their expectations. This is a negative bias in demand estimation, which is referred to as *censored* or *spilled* demand. The second source of bias is the substitute sales that are induced by customers who purchase a satisfactory available alternative when their desired product is not in stock. This leads to an increase in the sales of available products and is termed *recaptured* demand or stockout-based substitution (Vulcano et al., 2012; Conlon & Mortimer, 2013).

Therefore, disregarding product unavailability leads to overestimating demand for in-stock products and underestimating it for out-of-stock products. In other words, a sales transaction reflects the real demand only if all products are available in stock at the time of the sale—a condition that rarely occurs in firms offering with a broad product mix over a large retail distribution network. Knowledge of products' demands rather than relying strictly on historical sales data is necessary for firms to make and assess supply planning, assortment planning, and inventory management decisions, considering both expected induced costs and revenues.

As will be detailed in the literature review in section 2, there is a stream of papers that have studied the demand estimation problem under customer substitution and incomplete product availability. These papers generally attempt to identify a customer's substitution probabilities under incomplete product availability for perishable products in grocery stores and vending machines where customers make decisions about their daily needs. The reported methodologies are not designed to appropriately

*Corresponding author
Email address: sderhami@binghamton.edu (Shahab Derhami)

estimate demand for high-value products such as cars, recreational vehicles, and trucks because, on the one hand, customers have distinctive behaviors and, on the other, retailers employ distinctive inventory management policies. The shopping behavior of customers in high-value product markets has three key characteristics that influence demand estimation:

1. A loyal customer may well be willing to visit multiple retail stores belonging to the same firm to find her desired product or a satisfactory match. The number of retailers visited by a customer differs across different types of customers (e.g., picky vs. easy-going) and depends on the distance that the customer has to travel. Hence, a lost sale for a retail store is not necessarily a lost sale for the retail firm as some unsatisfied customers at a given store may purchase a product in a subsequent visit to another store in the firm's network.
2. A customer may well know what her desired product is when visiting a retail outlet, and her preference choices (substitute products) can be estimated using products substitution fitness. The substitution fitness (or substitution rate) of a product reflects the desirability of the product as a substitute for a customer whose desired product is not available. It presents the probability that the customer substitute the product. Substitution fitness can be estimated from the characteristics, functionalities, or appearances of the products. For example, in the car industry, factors such as engine size, platform, model comfort and convenience, color, body type, and extra accessories are among the most influential factors in customer decisions. These factors can be used to estimate substitution rates between products and to develop a substitution matrix (van Rijnsoever et al., 2009; Derhami et al., 2021).
3. A customer may be willing to wait up to her *acceptable delivery time* to receive her product. This allows retailers to exploit network resources that can be accessible within the customer's acceptable waiting time. This time might be product dependent and vary across different groups of customers. For example, it may range up to several days for leisure products such as recreational vehicles (Derhami et al., 2021).

The inventory management policies retailers employ in high-value product markets have distinctive characteristics that impact demand estimation:

1. The combination of a high marginal profit and frequent occurrences of product unavailability justifies employing on-demand inventory transshipment between retail centers to satisfy demand for out-of-stock products (Zhao & Atkins, 2009; Shao et al., 2011; Rosales et al., 2013). That is, a retailer that cannot satisfy a demand for an out-of-stock product may request transshipment from other retail stores that have the product or a satisfactory substitute in stock and located in a distance that allows shipping within the customer's acceptable waiting time. Nevertheless, a sale is lost if the retailer cannot provide the customer with a satisfactory product within a satisfactory period. Hence, product unavailability-based substitution must be considered in this environment rather than only stockout-based substitution.
2. Due to sales-related processes such as registration and warranty, accurate information on product availability is recorded at each retail center. Exploiting the new generation of daily perpetual inventory review systems, exact times of stockout occurrences and sales transactions can now be known, which was not the case with the previous generation of periodic systems (Mersereau, 2015).

Motivated by a collaborative research project with a leading manufacturer of recreational vehicles, this paper studies the estimation of the potential stockout-based lost sales for high-value products across an interconnected distribution network in which retail centers transship inventory on-demand to satisfy demand for an out-of-stock product. Our proposed methodology uses sales data (sales transaction logs) and product availability at the time of sale to estimate potential lost sales due to stockout. It is vital to consider the above-outlined behaviors of customers and retail centers to derive an accurate estimation of demand. Our empirical study shows the sharing inventory in this market, established through on-demand inventory transshipments, contributes significantly to the sales and economics of both retailers and firms. This particularly distinguishes demand estimation in this market from others studied in the literature. Figure 1 presents the frequency of on-demand inventory transshipments in the retailer network of our industrial partner for one product line over one year. As the figure shows, the frequency of lateral transshipments is considerable, and this affects the product availability in the network. In this environment, an out-of-stock situation does not necessarily lead to the loss of a sale. The potential loss of a sale comes from the case of an "unavailable" product, which refers to an out-of-stock product that is unreachable thorough an inventory transshipment within the customer's acceptable waiting time.

We model customers and dealers decision making processes as follows. We assume that either the customer knows her desired product when she visits a retail outlet or a salesperson from the retail center guides her correctly towards choosing her desired product from the firm's product portfolio without being influenced by product availability. Her secondary choices (substitute products) can be estimated using the substitution fitness of similar products to her desired one. The customer may visit multiple retail centers of the firm to find her desired product or a satisfactory substitute—a product whose substitution fitness to her desired one is sufficiently high and is delivered to her within her acceptable waiting time. The length of the tour (i.e., the number of visited retail centers) is stochas-

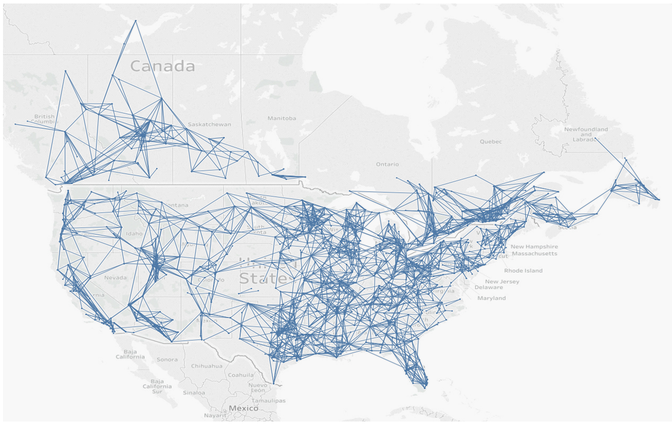


Figure 1: Inventory transshipments over one year among retailers in a distribution network of high-value products.

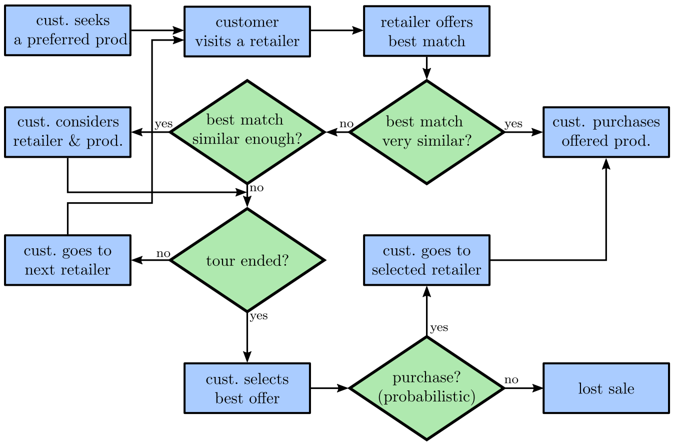


Figure 2: A customer's decision process.

tic and depends on the number of retailers in her search radius, travel distances, and her persistence in finding a better match to her demand.

The customer makes her decision to purchase or to continue the tour based on the closeness of the substitution fitness of the closest match to her desired product offered in a visit, given her waiting time constraint. She evaluates the offered product using two personal thresholds: the *immediate-purchase* and *considering-purchase* thresholds. The immediate-purchase threshold is the minimum substitution fitness to the customer's desired choice that invokes an immediate purchase decision and tour termination. The considering-purchase threshold is the minimum substitution fitness of the offered product to the customer's desired product that meets her minimum expectation but is not high enough to invoke an immediate purchase decision. In this case, customer keeps the retailer and the offered product in her *consideration list* and continues shopping. The considering-purchase threshold of a customer is always smaller than her immediate-purchase threshold. At the end of a tour, if a customer has not interrupted the tour with an immediate purchase decision, she considers purchasing a product from her consideration list. This decision is probabilistic, and the probability of making a purchase equals the substitution fitness (probability) of the best offered match to her desired product. The higher is the substitution fitness of a product, the higher is the chance of accepting the product as a substitute. If a customer does not purchase a product at the end of her tour, then she is considered a lost sale to the firm. This process is demonstrated in Figure 2.

The retailers attempt to satisfy a customer using in-stock products while considering possible eligible inventory transshipments from other retail centers or the firm's depots. An eligible transshipment is a transshipment that is delivered within the customer's acceptable waiting time. The delivery time is the sum of the transfer time between sending and receiving retail centers and the preparation time. The longer the delivery time is, the less desirable the time-specific product offer is to the customer. We model this by penalizing the substitution fitness value of the transshipped product to the customer's demand based on the delivery time (i.e., reducing the probability of accepting the the transshipped product as a substitute).

In this paper, we use past sales transactions and inventory log data to estimate the potential lost sales due to product unavailability across all stores in a network of retailers that distribute high-value substitutable products. Given the target industry considered in this research, the model uses an exogenous substitution matrix induced by product characteristics, appearances, and functionalities. Using only past sales and inventory log data, we first transform sales into the expected sales shares for all potentially desired products for a sales transaction for which demand would result in substituting to the sold products. This is used to estimate the customer arrival rate for each product. We then identify the unavailable products for which potential demand could not be fulfilled with either an exact match or a satisfactory substitute for all retailers within a replenishment period (e.g., a month). We account for the probabilistic lengths of the customer tours and the possibility of eligible inventory transshipments when computing products' unavailable periods. Finally, we estimate customer arrival rates in the period that a product is unavailable at a retailer using the expected sales shares of the product in a cluster of retail centers that have the same demand patterns, while the market sizes of the retailers are also considered in estimating the untapped demand.

We validate our model with a case of recreational products through three different approaches. We demonstrate that our estimated lost sales are statistically equivalent to the results of a field data collection across more than 100 dealerships and use simulation to assess the accuracy of our algorithm in estimating lost sales and product demands. The mean absolute error of our model in estimating annual lost sales was on average less than 6 vehicles per dealership (across more than a thousand dealerships) and less than 60 units per product across the United States and Canada. The results of a pilot project conducted by our partner company showed that the sales improvement achieved by temporarily maximizing product availability in part of the distribution network was very close to the estimated lost sales for that region. The numerical analyses revealed that sales data significantly underestimated demand for most products and retail centers and that preventing incomplete product availability by maintaining an appropriate set of inventory can result in millions of dollars of additional revenue.

The remainder of this paper is organized as follows. In section 2 we review relevant research papers and describe the gap in the current literature and the contributions of this paper. Then, we demonstrate the proposed algorithm in section 3 and describe its three main steps in detail. We discuss model validation and parameter tuning in section 4. We used three approached to validate the model: a field data collection, simulation, and a pilot project. Finally, we discuss numerical analyses for a case of recreational products in section 4.5.

2. Related research

Demand estimation under customer substitution has been extensively studied in the inventory management, marketing, revenue management, and economics literatures. As described in Kök & Fisher (2007), two types of customer substitution models are used in the literature: (1) utility-based substitution, where customers associate a utility with each product and choose the available product with the highest utility (van Ryzin & Mahajan, 1999; Transchel, 2017), and (2) exogenous substitution, where customers choose from a complete list of products, and if the selected product is not available, they may accept a substitute according to a given substitution probability (Nagarajan & Rajagopalan, 2008; Fisher & Vaidyanathan, 2014). Our model is a special case of the second category where the probability of customer substitution depends on the similarity of the offered product to the customer's desired product. In the following, we review demand estimation papers that are similar to ours in some respects.

Anupindi et al. (1998) is a seminal study on demand estimation under stockout and customer substitution. They studied both perpetual and periodic inventory systems under an independent Poisson customer arrival rate assumption and exogenous customer choice probabilities. They found the maximum likelihood estimate (MLE) of the demand parameters and substitution probabilities using an expectation-maximization (EM) algorithm by treating the stockout times as missing data. The number of parameters in their model is abundant, as it requires estimating arrival rates for all possible sets of products that a customer may encounter in a stockout situation. To overcome this issue, they restricted their choice model by imposing a "one-stage substitution" restriction for which a portion of the demand for an out-of-stock product is transferred to the second product, and if that product is also out-of-stock, then the demand is considered a lost sale. They applied their model to vending machines and showed that the resulting demand rates are different from the observed sales.

Musalem et al. (2010) used partial product availability information recorded in a periodic inventory review system to estimate customer demand under stockout. They used a Bayesian approach to simulate the evolution of the inventory and estimated the choice model and lost sales parameters using Markov chain Monte Carlo.

Vulcano et al. (2012) developed an EM algorithm to find the MLE of the demand parameters and the potential lost sales under stockout and customer substitution using a multinomial logit (MNL) choice model with a non-homogeneous Poisson model of customer arrivals. Their approach treats sales transactions as incomplete observations of the primary demand and applies the EM algorithm to the primary demand model to estimate parameters of the choice model. They tested their method on case studies in the airline and retail industries.

Conlon & Mortimer (2013) proposed an EM algorithm to estimate demand parameters under stockouts in a periodic review inventory system. As in Anupindi et al. (1998), they treated stockout times as missing data. Applying their approach to the data obtained from vending machines, they showed that disregarding product availability leads to substantial bias in demand estimation.

The mainstream of relevant research in the field of inventory management focuses on the estimation of demand for the assortment planning problem. See Section 1 of the online supplemental materials for a detailed review of such papers.

One of the most widely used inventory management policies in the retail networks of high-value products to satisfy demand for an out-of-stock product is lateral inventory transshipment, the impacts of which on demand estimation have not been adequately studied in the inventory and revenue management literature. The mainstream of research on inventory transshipment focuses primarily on inventory control policies (Grahovac & Chakravarty, 2001), optimal transshipment protocols (Wee & Dada, 2005), profit sharing policies (Huang & Sošić, 2010), and pricing policies (Dong & Rudi, 2004). See Paterson et al. (2011) for a detailed review of these papers.

Most of the papers on assortment planning under customer substitution estimate demand parameters using choice-based models (such as MNL and locational choice models). While the accuracy of these models depends exclusively on the parameters and type of the choice model considered, the independence of irrelevant alternatives (IIA) assumption made in these models is not valid for a dynamic customer substitution case in which the likelihood of substituting a product changes based on the available products offered to the customer.

A few papers such as Anupindi et al. (1998); Musalem et al. (2010); Vulcano et al. (2012); Conlon & Mortimer

(2013) have studied demand estimation from sales and product availability data, as we propose, but they differ from our work in methodology and problem scope. These papers mainly use EM algorithms to obtain the MLEs of the demand parameters and substitution probabilities, assuming predefined probability distribution functions such as Poisson for customer arrivals. Some of these papers assumed either static substitution, where customer preferences depend on the offered assortment in stores rather than the store inventories, or a limited number of substitution stages (e.g., one-stage substitution, where only substituting with the best substitute is considered, and demand is regarded as a lost sale if the best substitute is not available) (Anupindi et al., 1998). Our model considers dynamic substitution and proposes a new approach to estimate demand using sales data in other branches of the retail network considering retailers' market sizes.

Furthermore, the existing papers on demand estimation are mostly focused on a basic retail case where a customer with a choice set decides which product to purchase or leave the store without a purchase. Unlike our model, they do not consider customers visiting multiple stores to find a desirable match and retail centers performing inventory transshipment to satisfy the demand for out-of-stock products, situations that frequently occur in today's competitive markets, especially in high-value products markets such as vehicles. Our paper contributes to the demand estimation literature in four key ways:

1. We model dynamic substitution whereby customers select an available product from their preferred choices when their desired model is not available.
2. Our model considers specific customer behaviors and inventory management systems that are prevalent in high-value product markets but have not been adequately addressed in previous studies: we model (a) a customer visiting multiple retailers to find a better match and (b) a retailer exploiting on-demand inventory transshipments from other retailers, a depot, or a manufacturing plant to satisfy the demand for an out-of-stock product.
3. Our proposed approach uses a new data-driven approach to estimate the customer arrival rates and is not restricted to any given distribution of customer arrivals.
4. Our model estimates potential lost sales for retailers by accounting for their market sizes.

Without loss of generality, the model proposed in this paper can be used for markets where products have significant values, customers are prone to substitution, and retailers exploit multi-source transshipment. Examples of such markets include automobiles, trucks, recreational vehicles, jewelries, appliances, and customer electronics.

3. Estimation of the potential stockout substitutions and lost sales

Our data-driven approach uses past sales transactions and inventory log data to estimate potential lost sales induced by product unavailability. As synthesized in Figure 3, it consists of three steps:

1. Finding all potentially desired products for each sales transaction and calculating expected sales shares for these products to account for potential out-of-stock substitutions.
2. Finding unfulfillable products for which demand cannot be satisfied by either an exact match or a substitute and the time period that a product has been unfulfillable.
3. Estimating customer arrival rates for unfulfillable products at each retailer using the expected sales shares of the products.

An unfulfillable product is an unavailable product whose satisfactory substitutes are also all unavailable too. Stockout-based lost sales occur during the time periods that a product is unfulfillable. Hence, customer arrival rates for unfulfillable products are required to estimate potential lost sales. We use the expected sales shares (calculated in step 1) to estimate customer arrivals. That is, for each unfulfillable product in a store, we use the expected sales shares of the product in stores with similar demand profiles that have had the product or its satisfactory substitutes in stock. Details of these steps are described in the following sections.

3.1. Transforming sales transactions to expected sales shares

In the absence of information on true demand, sales transactions do not necessarily reflect the products desired by customers, as the purchased products may have been the result of stockout substitutions. We estimate customer arrival rates using sales occurrences; however, to account for stockout-based substitutions, for each sales transaction, we identify unavailable products that when initially demanded would have then resulted in the customer instead purchasing the sold product. These unavailable products and the sold products comprise the potentially desired products for a sales transaction and receive expected sales shares based on their substitution fitness to the sold product. The expected sales shares then provide a basis to estimate customer arrival rates.

We assume exogenous substitution model. Different approaches can be used to estimate substitution fitness (probability) such as expert opinions, analytical models, or surveys. We calculate it based on the characteristics, functionalities, and appearances of the products. We calculate the substitution fitness of product p' to p by

$$S_{p'p} = \sum_{f \in \mathcal{F}} w_f a^f_{p'p}, \tag{1}$$

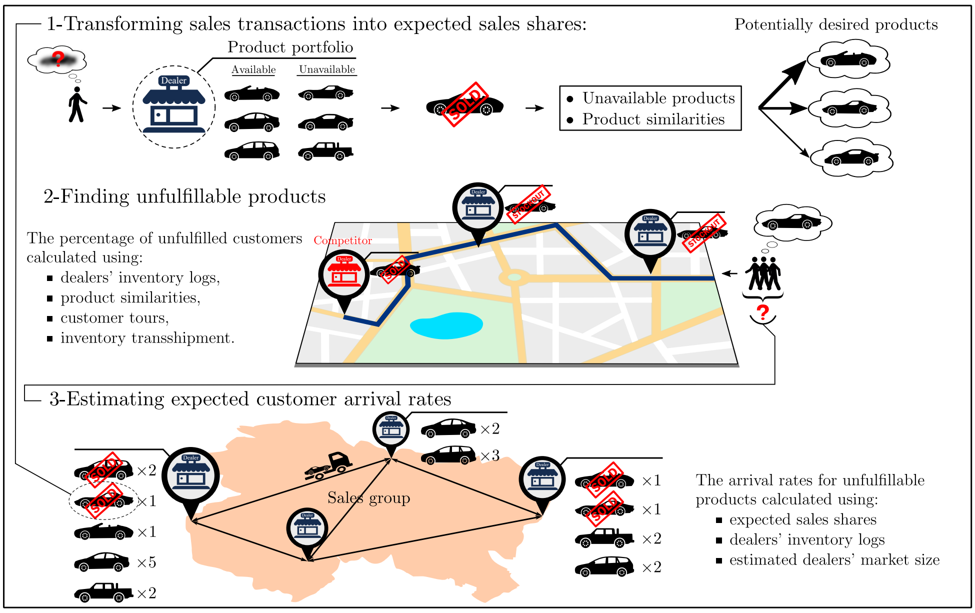


Figure 3: Overview of the approach as applied to a network of vehicle dealerships.

where $\mathcal{F}$ is the set of product features, $a^f_{p'p}$ $(0 \leq a^f_{p'p} \leq 1)$ is the substitution fitness (probability) of product p' to p with respect to feature f, w_f $(0 \leq w_f \leq 1)$ is the weight of feature f in assessing substitution desirability of products, and $\sum_{f \in \mathcal{F}} w_f = 1$. Parameter $a^f_{p'p}$ reflects the probability of substituting feature f of the desired product by that of product p'. The features we used to estimate substitution fitness ratios in our case study are engine, platform, number of seats, model year, and color. Note that $0 \leq S_{p'p} \leq 1$. Parameter $S_{p'p}$, then, reflects the probability of substituting product p by p'.

Consider a sales transaction at time t at retailer r in which product p was sold. We assign sales share to product p' for this transaction using the following expression:

$$s^t_{rp'} = \begin{cases} \frac{S_{p'p} - C_{p'}}{\sum_{p' \in \mathcal{P}_p}(S_{p'p} - C_{p'})} & \text{if } p' \in \mathcal{P}_p, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $C_{p'}$ is the considering-purchase threshold of a customer whose preferred product is product p', and $\mathcal{P}_p$ is the set of potentially desired products. Expression 2 splits sales transactions among potentially desired products for each transaction based on their substitution fitness to the sold product. $s^t_{rp'}$s will be used for estimating customer arrival rates. Only products that belong to $\mathcal{P}_p$ are eligible to receive sales shares from a sales transaction. Let $\mathcal{I}^t_r$ be the set of all in-stock products at retailer r at time t; then, set $\mathcal{P}_p$ contains product p and every product p' that meets the following conditions: (1) $p' \notin \mathcal{I}^t_r$, (2) $S_{p'p} > C_{p'}$, and (3) $S_{p''p'} < S_{pp'}$ $\forall p'' \in \mathcal{I}^t_r : p'' \neq p$. The set $\mathcal{P}_p$ contains all potentially desired products for a sales transaction that when demanded by a customer would ultimately lead to the customer purchasing the sold product. Condition (1) excludes all available products from the set because the customer would have purchased them instead of the sold product if she wanted them. Condition (2) ensures that only those products are considered for which the sold product is a satisfactory substitute. Condition (3) restricts $\mathcal{P}_p$ to only products for which the best match among in-stock products is the product sold. This is because a desired product that has a better available match than the product sold would result in the sale of a different product. See Section 2 of the online supplemental materials for an example describing how the expected sales shares are computed.

If set $\mathcal{P}_p$ is singleton (i.e., contains only p), then that means all potentially desired products have been available at the time of the sale and the sold product is considered the desired product for the customer. Therefore, product p receives the entire sales share. We do not consider the intrinsic demand of products when assigning the sales shares because the substitution decision is based on the substitution fitness (closeness of features and characteristics) of available alternatives to the desired one, not their popularity.

Assume that the retailer participates in an inventory transshipment policy. The longer it takes to deliver a transshipped product to a customer, the less desirable that product is to her. We model this by penalizing the substitution fitness of the transshipped products using a time decay factor e_{rd}, where d and r are sending and receiving centers, respectively. The values of e_{rd} are determined empirically based on the travel distance between the two centers (delivery times). To adjust the model to satisfy this condition, we need only to restrict $\mathcal{P}_p$ in (2) by a new condition: (4) $\max_{d \in \mathcal{E}_r : p' \in \mathcal{I}^t_d} \{e_{rd} S_{p'p}\} < C_i$, where $\mathcal{E}_r$ is the set of all centers that are available to retailer r for inventory transshipment. This condition excludes products that have been available through an inventory transshipment from set $\mathcal{P}_p$.

3.2. Finding unfulfillable products

A potential lost sale may occur only for an unfulfillable product, during the time that it is unfulfillable by a retailer. The time unit for tracking this period depends on the inventory review system and the frequency of sales and replenishment. The retailers in high-value product markets typically use a perpetual inventory review system, where the granularity of inventory records varies from hourly to daily. Our observations from a case study showed that tracking inventory levels on a daily basis provides sufficient accuracy to capture changes in the inventory level of a retailer in the recreational products industry.

Let p^t_{rp} be the probability that retailer r makes a successful sale to a hypothetical customer who demands product p on day t and $S^{t-\max}_{rp}$ be the substitution fitness of the best match available for product p. Then,

$$p^t_{rp} = \begin{cases} 1 & \text{if } S^{t-\max}_{rp} > M_p, \\ S^{t-\max}_{rp} & \text{if } C_p \le S^{t-\max}_{rp} < M_p, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where M_p is the immediate-purchase threshold of a customer whose desired product is p. It represents the minimum substitution fitness that invokes an immediate purchase decision by a customer. The probability of making a sale is one if the substitution fitness of the best available match to the customer's desired product is higher than her immediate-purchase threshold. Otherwise, it is $S^{t-\max}_{rp}$ if the substitution fitness is higher than her considering-purchase threshold. Considering only the retailer's inventory to satisfy the demand,

$$S^{t-\max}_{rp} = \max_{p' \in \mathcal{I}^t_r} \{S_{p'p}\}. \tag{4}$$

Let u^t_{rp} be the probability that the firm misses an opportunity to satisfy a customer who starts her tour at retailer r on day t and demands product p. Thus, $u^t_{rp} = 1 - p^t_{rp}$. $\sum_{t \in T} u^t_{rp}/\|T\|$ yields the percentage of unfulfillable demand for product p in period T (e.g., a month), and multiplying it by the customer arrival rate for product p yields the expected lost sales for the product.

The retailer may be able to provide the customer with a satisfactory match through an inventory transshipment. In this case, the best available matches through transshipment must be considered in addition to the retailer's stock. Hence,

$$S^{t-\max}_{rp} = \max \left\{ \max_{p' \in \mathcal{I}^t_r} \{S_{p'p}\}, \max_{d \in \mathcal{E}_r, p' \in \mathcal{I}^t_d} \{e_{rd} S_{p'p}\} \right\}. \tag{5}$$

A sale is lost for a retailer if neither an exact match nor a satisfactory substitute is available at the retailer. However, a customer who is willing to visit more than one store may find a satisfactory product at another store. Hence, a lost sale for a retailer is not necessarily a lost sale for the firm. Assume that v_j presents the fraction of customers who are willing to visit up to j stores to find their desired products, and n_v is the maximum number of stores that a picky customer visits. Then, $\sum_{j \in \{1 \dots n_v\}} v_j = 1$. Consider a picky customer whose desired product is p and visits up to n_v stores. The probability that the customer starts her tour at retailer r on day t and purchases a product by the end of her tour is $\max_{j \in \{1 \dots n_v\}} \{p^t_{d^r_j p}\}$, where d^r_j is the jth retailer that the customer visits.

Assume that only the kth store visited has a satisfactory product for a customer and she visits stores in ascending order of their distances to store r. Only the fraction of customers whose tour length (i.e., the list of visiting stores) is long enough to visit store k will reach the product. The remaining customers will be lost. Hence, the probability of making a successful sale depends on the best match each customer finds on her tour, $S^{t-\max}_{rp}$. If it meets her immediate-purchase threshold, then the success probability is one; otherwise, if it meets her considering-purchase threshold, the success probability is the substitution fitness of the best match to her desired product, zero otherwise. Hence, the expected portion of unfulfilled customers, which is the complement of the probability of successful sales, is:

$$u^t_{rp} = \sum_{j=1}^{n_v} v_j \left(\min_{k \in \{1 \dots j\}} \left\{ 1 - p^t_{d^r_k p} \right\} \right). \tag{6}$$

See Section 3 of the online supplemental materials for an example describing how the expected portion of unfulfilled customers are computed.

3.3. Estimating the potential lost sales

Demand (customer arrival) for unfulfillable products results in lost sales. To estimate customer arrival rates, we group retail centers based on their market profiles/sizes and use the expected sales shares of the products in retail centers of the same group that have had the product or its satisfactory substitutes in stock. Estimating customer arrivals over similar retailers rather then the entire distribution network provides more accurate estimates and captures regional demand and sales trends. A group of similar retailers may include a set of retail centers that have similar demand patterns and/or market profiles/sizes and/or are located in the same sales region. In this paper, we group retailers based on their market profiles (e.g., agricultural, sport, hunting), sizes (small, medium, and large dealers), and regions.

The average customer arrival rate for product p over period T is obtained by taking the average of s^t_{rp}s over all retailers who have had the product or its satisfactory substitutes in stock. It is given by

$$\left(\frac{1}{\|\mathcal{R}_p\|}\right)\sum_{r\in\mathcal{R}_p}\sum_{t\in T}s_{rp}^{t}, \tag{7}$$

where $\mathcal{R}_p \subset \mathcal{R}$ is the set of all retailers that belong to the same sales group, $\mathcal{R}$, and have had product p or its satisfactory substitutes in stock at least for one day in T. A sales group consists of a set of homogeneous retailers that have similar sales profiles. We employed them to account for the heterogeneity of demand in different retailers. For example, in our case study, the agricultural sports vehicles had higher sales in some dealers. These dealers belong to the same sales group. If all retailers have similar sales profiles (i.e., demand shares of product categories do not significantly vary in retailers across different geographical regions), then there will be a unique sales group that includes all retailers. Period T should be defined long enough to capture statistically sufficient customer arrivals for all products but not too long to impair the precision and miss seasonality trends. Our testbed industry (recreational vehicle industry) experiences monthly demand seasonality; therefore, we set T to one month in our experiments and estimated demand and potential lost sales on a monthly basis.

The expected customer arrival rates are not the same for all retailers. Larger retailers receive more customers (assuming a market-driven network). Thus, the retailer's market sizes must be taken into account when customer arrival rates are computed. Let D_r^T be the total demand of retailer r in period T. We use it to calculate the weighted average customer arrival rate for each product with respect to the relative market sizes of the retailers. The potential lost sales for product p at retailer r over period T are then given by

$$l_{rp}^{T}=\left(\frac{\sum_{t\in T}u_{rp}^{t}}{\|T\|\|\mathcal{R}_p\|}\right)\sum_{d\in\mathcal{R}_p:d\neq r}\left(\frac{D_r^T}{D_d^T}\right)\sum_{t\in T}s_{dp}^{t}. \tag{8}$$

However, l_{rp}^T is required to calculate D_r^T:

$$D_r^T=\sum_{p\in\mathcal{P}}l_{rp}^{T}+\sum_{t\in T}\sum_{p\in\mathcal{P}}s_{rp}^{t}, \tag{9}$$

where $\mathcal{P}$ is the set of all products in the portfolio. Thus, l_{rp}^T and D_r^T are mutually dependent variables. We propose an iterative procedure to consecutively update l_{rp}^T and D_r^T until the sum of the potential lost sales in the sales group converges. The process proceeds as follows. First, we set $l_{rp}^T = 0$ for all $r \in \mathcal{R}$ and $p \in \mathcal{P}$ and calculate D_r^T. Then, l_{rp}^T is updated using the D_r^T calculated in the previous iteration. This process continues until the gap of $\sum_{r\in\mathcal{R}} D_r^T$ between two consecutive iterations is smaller than an acceptable error, ϵ. Then, the expected number of lost sales

Algorithm 1 Steps of the proposed algorithm

initialize parameters
for all day $t \in T$ **do**
  for all retailer $r \in \mathcal{R}$ **do**
    for all sales transaction in day t **do**
      calculate s_{rp}^t for all products $p \in \mathcal{P}$
  for all retailer $r \in \mathcal{R}$ **do**
    for all product $p \in \mathcal{P}$ **do**
      calculate u_{rp}^t
set $n = 0$
set $l_{rp}^T(n) = 0$ for all $r \in \mathcal{R}$ and $p \in \mathcal{P}$
calculate $D_r^T(n)$ for all $r \in \mathcal{R}$
repeat
  set $n = n + 1$
  for all retailer $r \in \mathcal{R}$ **do**
    for all product $p \in \mathcal{P}$ **do**
      calculate $l_{rp}^T(n)$
    calculate $D_r^T(n)$
until $\left|\sum_{r\in\mathcal{R}} D_r^T(n) - \sum_{r\in\mathcal{R}} D_r^T(n-1)\right| < \epsilon$
calculate L_r^T for all retailer $r \in \mathcal{R}$
return L_r^T and $Pr(x_r^T = p)$ for all retailer $r \in \mathcal{R}$

in retailer r is obtained by

$$L_r^T=\sum_{p\in\mathcal{P}}l_{rp}^{T}, \tag{10}$$

and the probability that retailer r has lost a sale for product p is obtained by

$$Pr\left(x_r^T=p\right)=\frac{l_{rp}^{T}}{\sum_{p\in\mathcal{P}}l_{rp}^{T}}. \tag{11}$$

Algorithm 1 presents the steps of the proposed model.

4. Model validation and parameter setup

Empirical validation of demand estimation models through recording lost sales and all customers' desired products might in practice be particularly difficult and expensive. Gathering these data from a large distribution network on a scale that provides statistically valid conclusions is costly and difficult to implement because it is time and labor-intensive and requires software and hardware equipment and employee training. For these reasons, papers that have studied industries for which demand data are easily captured and the number of products carried is limited, such as vending machines, have employed field data for model validation (Anupindi et al., 1998; Conlon & Mortimer, 2013); while most papers that have studied industries with high product variety have used simulation for parameter tuning and, to some extent, model validation (Vulcano et al., 2012; Jain et al., 2015; Mersereau, 2015; Wan et al., 2018).

We applied three approaches to validate the algorithm. First, we collected field data and surveyed more than 100 dealerships of our partner company to obtain their observations of lost sales that occurred within a six-month period and analyzed the results with our estimates for those dealerships in the studied period. Second, we performed a simulation study to analyze the accuracy of the model

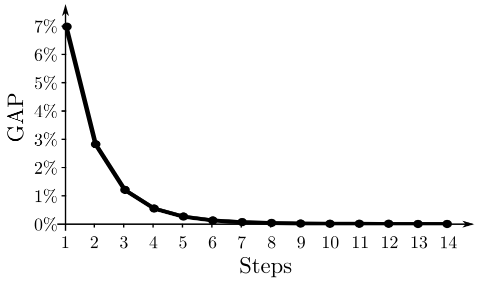
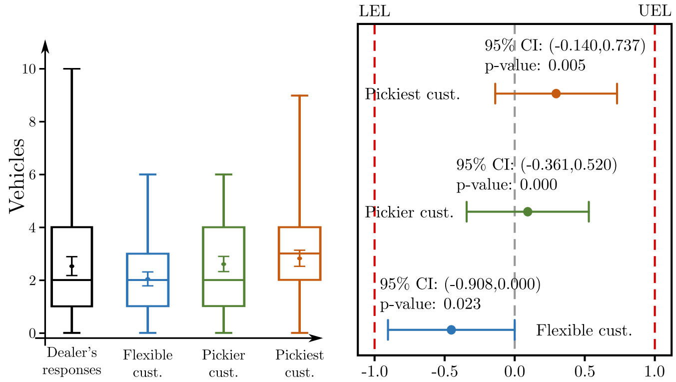


Figure 4: Convergence of the algorithm.



(a) Distribution of the potential lost sales

(b) Results of the paired sample equivalence test.

Figure 5: Comparing the results of the survey with three lost sales estimation scenarios

in estimating demand and lost sales at the product and dealer levels. Third, we analyzed the results of a pilot project that our partner company conducted to maximize product availability for one year in one of its sales regions to study sales improvements obtained by avoiding incomplete product availability.

Our partner company is a leading manufacturer of recreational vehicles. The company offers distinct product lines dedicated to specific types of recreational vehicles (e.g., quads, snowmobiles). The company has a large network of dealerships in North America. The dealerships are independently owned businesses, some of which occasionally exploit inventory transshipment to satisfy a customer's demand for an out-of-stock product. Dealerships place monthly orders and receive products in two months after placing their orders. We applied our approach to estimate potential lost sales for one of its product lines and used the company's data to validate our model.

4.1. Parameter tuning

The model contains four sets of parameters: e_{rd}, C_p, and M_p. These parameters control customer behaviors in accepting an inventory transshipment and making purchase decisions. We obtained the values of these parameters using a survey study and a scenario optimization approach that uses simulation and historical data. For more information about these analyses, see Section 4 of the online supplemental materials.

The algorithm was coded in Java and run on a computer equipped with an Intel Xeon processor E5-2630 V3 (2.4GHz) and 128 GB of RAM memory. The computational time for each sales group takes between 15 to 35 seconds. Figure 4 depicts the convergence of the algorithm. It presents the changes in the absolute percentage gap between the estimated demand in two consecutive iterations of the algorithm. As the graph shows, the gap between two consecutive steps shrinks fast and the algorithm converges rapidly. On average, the algorithm converged in 4.6 steps in our experimental study, setting $\epsilon = 0.01$.

4.2. Field data collection

In high-value product markets such as car dealerships, retailers' representatives interact directly with customers and have a fairly accurate sense of true demand being lost due to product unavailability. Since the information on true demand was not available in the distribution network of our partner company, we surveyed retailer representatives to obtain their observations about stockout lost sales to collect field data and compare these data with our estimates. We collected their observations through a questionnaire in which the dealers were asked "how many sales they believe they have lost in the last six months as a result of either not having a specific product in stock or not being able to get the product fast enough to meet the customer's need". In total, 114 dealerships participated in the survey. The collected data showed that they, on average, believed that they had lost 2.52±0.40 sales in the studied period (with $\alpha = 0.05$). We estimated the potential lost sales for participating dealers in the studied period. Figure 5 compares the dealers' responses with our model estimates. Our algorithm estimated the average potential lost sales to be 2.04±0.30, 2.6±0.32, and 2.82±0.34 vehicles for the flexible, pickier, and pickiest customer scenarios, respectively. The average of dealers' responses is close to the estimated lost sales for the pickier customer scenario and bounded by the flexible and pickiest customer scenarios.

We performed a paired sample equivalence test between dealers' responses, as the reference population, and all three scenarios, as test populations, to measure the equivalency of model estimates with the collected data. Figure 5b presents the results of the paired sample equivalence test between the model estimates for all three scenarios and dealer responses for equivalence limits of (-1,+1). The test rejects the null hypotheses (i.e., the difference between the means of the test and reference populations is greater than or equal to the upper equivalence limit or less than or equal to the lower equivalence limit) for all scenarios at the 5% significance level and confirms that the differences between the test and reference populations are within the equivalence limits. As Figure 5a demonstrates, the mean, confidence interval of the mean, and distribution of the pickier customer scenario are the closest to the field-collected data among the three scenarios. This is confirmed by the smaller p-value of the equivalency test for this scenario.

4.3. Simulation study

Simulation is widely used to validate demand estimation models because demand realization is known in a simulation model and can be used to measure the accuracy of a model in estimating the realized demand in the simulation. We used the simulation model described in section 4.1 to validate the accuracy of the model in estimating lost sales and true demand. The simulation model simulates more than a thousand of the company's dealerships in the United States and Canada while following the current business models that dealerships use to place monthly orders and process transshipments and the firm uses to process dealer orders and deliver products. The simulation period is one year and includes the entire product mix of the product line we studied in section 4.5. The demand estimation model was run every month in the simulation model. This setup, which is computationally equivalent to simulating one month of a single dealer with a single product 720,000 times (=12*1000*60), makes the simulation model computationally extensive. Hence the number of simulation replications should be chosen such that statistically valid results are obtained using reasonable computational efforts. Our experimental analysis showed that setting the number of replications to five results in a sufficiently narrow confidence interval of the mean across all simulation replications for computed metrics. We used a different realization of customer demand in each replication. We generated customer demand scenarios (i.e., different realizations of demand) for simulation using the estimated demand and lost sales obtained by the proposed algorithm.

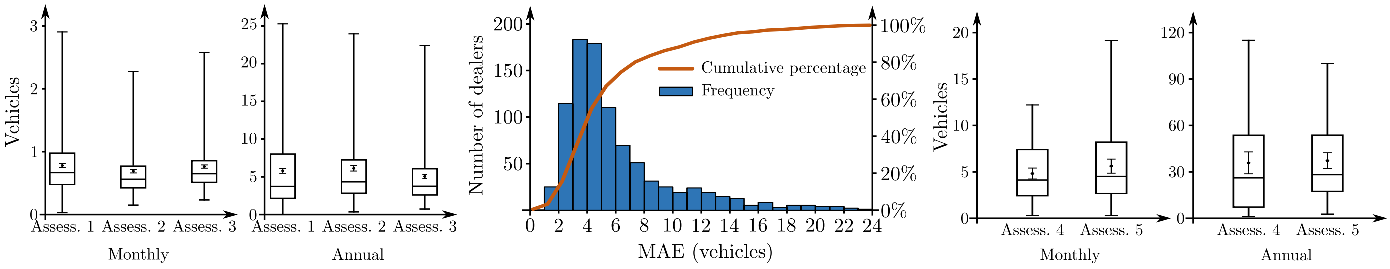
We used two approaches to assess the validity of the results of the model through simulation. In the first approach, we used simulation as a "black box" that receives customer arrivals and demand as inputs and then outputs sales and lost sales based on the customer and dealer behavior defined in this paper. The simulation model imitates customers and dealers' behaviors and their interactions as described in section 1 and maintains the same product availability in the network as it occurred in the real case using historical data (as explained in section 4.1). We used the estimated product demands and lost sales generated by the proposed algorithm to generate a realization of product demands in the simulation. Since product availability in the simulation was set to be identical to the real case, we expected the generated lost sales in the simulation be close to the estimated values that we fed into it. In other words, we expected the sales produced by simulation be close to the real case and the additional demand (estimated lost sales) we fed into the simulation model becomes lost sales (as we suspect occurred in the real case). The gaps between the generated sales and lost sales in the simulation and actual sales logs and the estimated lost sales reflect the accuracy of the algorithm.

Figure 6 shows the accuracy of the model in estimating the realized product demands and lost sales in simulation. The estimation error in this analysis can be used by dealerships for tactical planning such as budget, capacity, and staff planning. Figure 6a presents the mean absolute error (MAE) of the generated sales and lost sales (aggregated over all products) in the simulation across all dealerships over five simulation replications. Assessment 1 shows the distribution of mean absolute errors (MAEs) between sales made in the simulation and actual sales for all dealerships. Assessment 2 presents the distribution of MAEs between generated lost sales in the simulation and estimated lost sales for all dealerships. The average monthly error is less than one vehicle per dealership in both cases, and the error is less than one vehicle for more than 75% of dealerships. The annual MAEs are the errors between the sum of monthly estimates (estimates are calculated every month in the simulation) and actual values over one year. The average annual error is approximately six vehicles per dealership, and the error is less than ten vehicles over one year for more than 75% of dealerships. The marginal gap between the generated sales in the simulation and actual sales demonstrates the accuracy of the simulation demand scenarios, estimated by the proposed algorithm, in capturing the true customer demand and lost sales.

Figure 6c presents the accuracy of sales and estimated demand at the product level (aggregated over all dealerships). The estimation error in this analysis can be used by the original equipment manufacturer (OEM) for supply and production planning. While the accuracy of the model for the sales and lost sales of dealerships is important for retail planning, this analysis provides insights into the accuracy of the model from a supply chain planning perspective. Assessment 4 presents the MAEs between product sales in the simulation and actual sales. The average monthly and annual MAEs for product sales are less than six and 36 vehicles per product, respectively.

In the second model validation approach, we used the proposed model to estimate the realized customer demand in the simulation. In this approach, the simulation model simulates customers and dealers' behaviors and their interactions while following the business models that dealers use to place monthly orders (no historical data are directly used). We used only product availability and sales information from the simulation to estimate the realized demand and lost sales in the simulation model.

Assessments 3 and 5 in Figures 6a and 6c show the results of this test. Assessment 3 reports the MAEs between estimated lost sales and realized lost sales in the simulation per dealership. The average monthly and annual MAEs over all dealerships are less than one and six vehicles per dealership, respectively. Figure 6b presents the distribution of the annual MAEs in Assessment 3. As the graph reveals, the estimation error was less than seven vehicles for 80% of dealerships. Assessment 5 shows the accuracy of the model in estimating product demand (similar to Assessment 3). The average of monthly and annual MAEs



(a) MAEs in estimating dealerships' demand and lost sales (dealership level across all products).
(b) Distribution of annual MAEs for assessment 3.
(c) MAEs in estimating product demands (product level across all dealerships).

Figure 6: Accuracy of the model in estimating realized demand and lost sales in simulation across five replications of simulation.

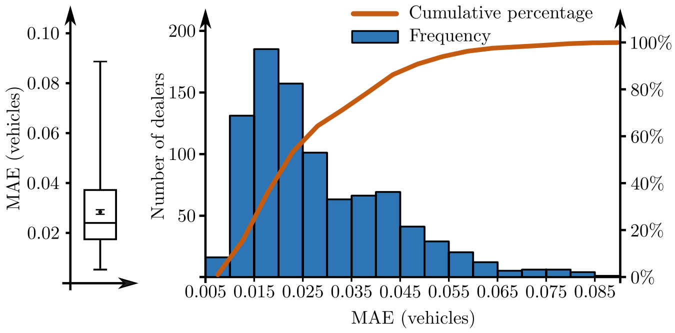


Figure 7: Assessment 6: Monthly MAEs of estimated product demands per dealership (dealer-product level).

for both assessments are less than six and 36 vehicles per product, respectively.

Figure 7 presents Assessment 6, which shows the distribution of MAEs of the estimated product demand at the dealer-product level. Assessment 6 shows the error in estimating the realized monthly product demand in simulation per dealership and per product. The dealerships placed stock replenishment orders every month. Thus, the estimation error in this analysis can be used by dealerships for stocking decisions. The monthly product demand estimation error was less than 0.05 vehicles for more than 90% of the dealerships.

As suggested by one of the anonymous referees, we used the mean absolute scaled error (MASE) to compare the precision of the model with a naïve method. The MASE is a measure of comparative accuracy for a forecast algorithm. It computes the ratio of MAEs of a more sophisticated model to those of a naïve forecast method. Hence, MASE less than one shows the proposed model outperforms the naïve forecast. The naïve forecast method uses the previous month's sales as the forecast for the next month. For more details about MASE, see Hyndman & Koehler (2006). Since our problem is not a forecast problem, we adjusted the original MASE method by using the monthly sales data instead of the original naïve forecast method. That is, we used the sales data as naïve estimates for product demand. This is more accurate than a typical naïve method because sales data contains partial information about product demand, but provides meaningful insights to practitioners because many companies use sales data instead of demand.

We performed regression analyses to study the effects of model type, dealer size, product availability, and product sales volume on MASE ratios aiming to find any combination of these factors that might result in the naïve method perform as well as the model. The details of the analyses are described in Section 5 of the online supplemental materials. While our statistical analysis did not detect any of the above factors or their combinations to be significant on the precision of the model, we found that the accuracy of the naïve method becomes closer to that of the model as the average sales of a product increases. The average MASE over all products was 0.86.

Computing MASE at the dealer-product level is not possible for all pairs of dealer-product because many dealers had zero sales for some products, and the division-by-zero error occurs. About 47% of pairs did not have a value due to this error. Excluding these pairs, the average MASE across all remaining dealer-product pairs was about 1.01, where about 50% of dealers had an average MASE of 0.97. This is mainly because the data at the dealer-product level is very sparse because of the nature of luxury product markets (i.e., small sale quantities but considerable revenue). Comparing the average MASE at the product level with the dealer-product level shows that while the performance of the model was close to the naïve method at the dealer-product level (distribution level), the model was capable of obtaining more accurate estimates relative to the naïve method at the product level. This shows that using the model to construct product demand estimates at the OEM level from the dealer-product level estimates by aggregating product demand at the dealerships improves the model's performance over the naïve method. Hence, although the model can be used for distribution planning, its usefulness increases for supply and production planning at the OEM level.

Figure 8 shows model estimates relative to sales and product demand for all products in North America over one year. Demand for most of the products was larger than the actual sales, except for four products for which it was smaller than sales (cases of a positive bias). The mean absolute percentage errors (MAPEs) of product demand

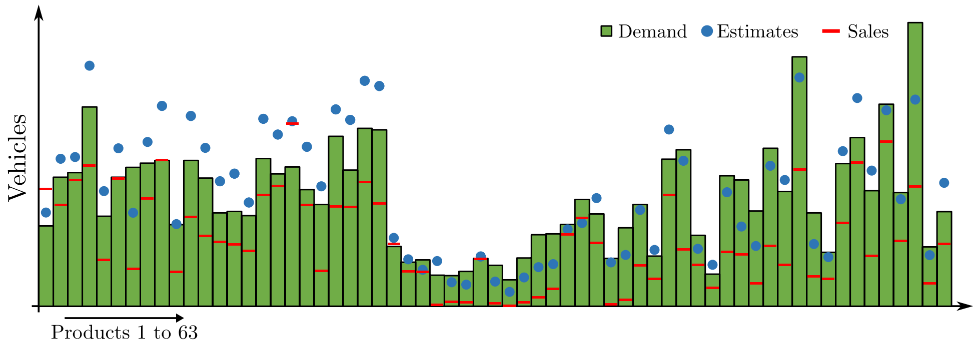


Figure 8: Simulation demand and sales vs. model estimates for all products over one year in North America.

estimates for the model and the naïve method are listed in Table 5 in Section 5 of the online supplemental materials. The average MAPE of the model and the naïve method across all products were 22.60% and 45.90%, respectively. The model underestimated demand for 27 products while the sales data underestimated demand for 59 products. While both demand underestimation and overestimation are costly and should be considered, the former costs potential gains form sales, which are usually costlier than inventory related costs imposed by demand overestimation. Demand underestimation by the proposed model led to 7.97% lost sales versus 40.06% lost sales induced by that of the naïve method.

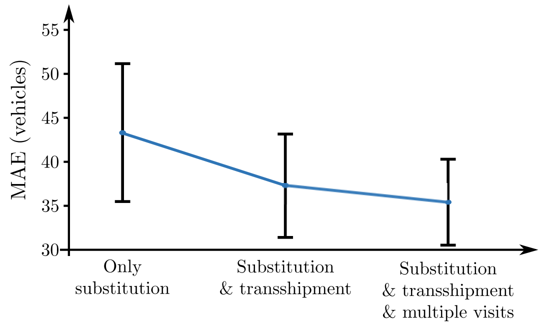


Figure 9: Precision of the model with respect to its main components: average of MAEs in estimating realized annual product demand in simulation and their confidence intervals across five replications of simulation, $\alpha = 5\%$.

We performed another experimental analysis to measure the relative gain in precision of the results with respect to the three main components of the model: substitution, transshipment, and customers visiting multiple stores. We built three different versions of our model for this purpose: a model that considers (1) only substitution, (2) substitution and inventory transshipment, and (3) substitution, transshipment, and visiting multiple stores by customers. Similar to Assessment 5 in Figure 6c, we used these models to estimate the realized product demands in the simulation model. We replicated this experiment five times for each case using a different realization of demand per simulation replication. We computed the MAE between the estimated annual product demands and realized product demands in the simulation. Figure 9 presents the averages of MAEs across five replications of the simulation and their 95% confidence intervals for the three scenarios. As the graph shows, the precision of the model increases as we move from a basic model that considers only substitution effect to a more realistic model considering all three components. Considering only substitution resulted in a 22% decrease in the precision of the model while adding transshipment effects improved the results by 14%. Among the three components, visiting multiple stores by customers had the lowest impact on the precision of the model mainly because the inventory profiles in nearby retailers were not diversified enough to satisfy a significant portion of unsatisfied customers visiting nearby stores.

4.4. A pilot project

Our partner company conducted a pilot project to practically assess the effects of increasing product availability. They chose one state in the United States to maximize product availability in the dealerships in this state for one year. They allowed the dealerships to maintain larger inventories with a wide variety of products and placed a strategic depot in a location that could serve all dealerships relatively quickly. The depot carried all products during the pilot and was used to transship products to dealerships on-demand to satisfy demands for out-of-stock products. The firm continuously monitored the depot's inventory and allowed dealers to place orders on a daily rather than monthly basis if necessary. The firm assigned orders from these dealerships with the highest priority and held their replenishment lead-time to under one week (as apposed to two months). For more information about the inventory manage system used in the pilot study, please see Derhami et al. (2021). Analyzing the sales in the pilot state revealed that sales increased to more than 30% above the average growth that the firm experienced trough the United States. This was in line with our estimate on the lost sales for this state, which equaled approximately 28%.

4.5. Experimental analysis

We used the model to estimate potential lost sales in the dealership network of our partner company across the

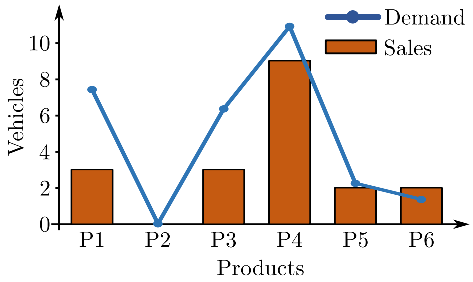


Figure 10: Actual sales vs. demand.

United States and Canada for a model-year of a product line that consists of more than 60 products. We grouped dealerships based on their market profiles (e.g., agricultural, sport, hunting), sizes (small, medium, and large dealers), and regions. We observed monthly seasonality trends in the sales data, therefore we set T to one month and run the algorithm on a monthly basis to account for the monthly seasonality. We selected one sales group to analyze and present the results. To protect against revealing sensitive information, we suppressed the axis labels of the charts that present aggregate data and randomly scrambled the sequence of months to disguise any sensitive trends. However, for the sake of comparison, the sequence of months was kept the same in all figures.

We picked six products to present lost sales calculation details. Section 6 of the online supplemental materials shows details of this analysis. Analyzing the estimated lost sales of these products showed that product unavailability is not the only influential factor in the estimated lost sales. Products with similar stock availability in the retail network may induce different levels of lost sales due to the availability of satisfying substitutes, the popularity of the product, and the market size of the dealerships in which the product is not available.

Figure 10 displays the estimated demand vs. actual sales for the six selected products. The estimated demand for a product is the sum of its expected sales shares and potential lost sales. As the graph shows, the estimated demands are higher than the actual sales for all products except for P6. This means that using only sales data to estimate demand results in negative estimation bias for P1 to P5 and positive bias for P6. The estimated demand for P6 is less than the actual sales of this product because the product or its satisfactory substitutes were available at most dealerships, and as a result, relatively small lost sales were observed for this product. On the other hand, a portion of the actual sales of P6 was credited to other products as out-of-stock substitutions.

We tested our algorithm for the three (pickiest, pickier, flexible) customer scenarios. The aggregated sales and potential lost sales calculated for each scenario over a one-year period are depicted in Figure 11. As we expected, the potential lost sales incurred by the pickiest scenario are larger than the potential lost sales captured by the other two more flexible scenarios. This is because pickier customers do not as rapidly accept substitutes as flexible customers do, and therefore, the likelihood of satisfying the former with a substitute product decreases. As shown in Figure 11a, the potential lost sales are always higher for the scenario with the pickiest customers, followed by the pickier and flexible customer scenarios. This trend remained the same in all months of the year. Estimating the potential lost sales for scenarios with pickiest and more flexible customers provides lower and upper bounds on the average potential lost sales when accurate information on the pickiness of customers in accepting substitute product is not available.

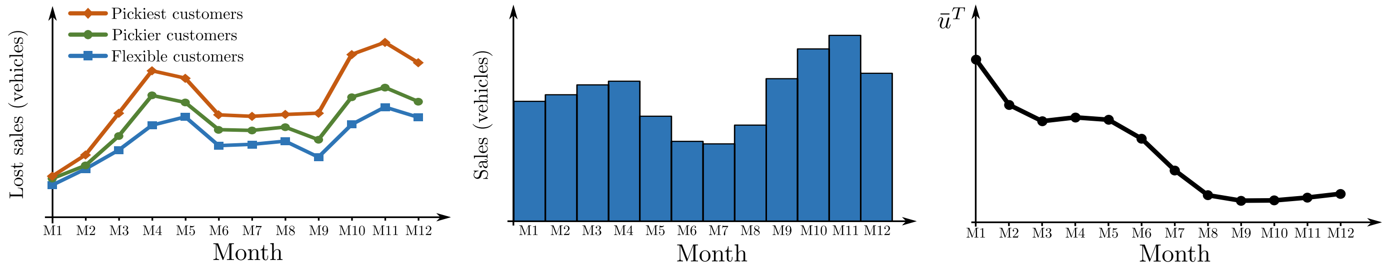
Figure 11 also shows that the model captures monthly seasonality effects and trends in the sales data. Comparing estimated monthly lost sales in Figure 11a with monthly sales in Figure 11b reveals that the estimated lost sales graphs follow sales seasonality trends. This is mainly because the model uses sales data to estimate customer arrival rates for lost sales calculations; therefore, underlying seasonality effects and trends in sales data are precisely reflected in demand estimation.

The estimated lost sales graphs in Figure 11a show that the total lost sales are higher in some months than in others (e.g., M4, M5, M10, and M11). Thus, the firm suffers more from incomplete product availability in the peak selling periods. Figure 11c presents the average product unavailability for all products and dealers in the selected sales group ($\bar{u}^T$ is the average u^t_{rp} across all dealers and products). In months M8 through M12, the average product unavailability (ignoring dealer sizes) remained nearly unchanged at its lowest value during the year, while the graphs of potential lost sales reached their highest values in month M11 and remained high in M12.

Figure 12 presents the scatterplot of the aggregate potential lost sales over a year in the United States. As the figure indicates, the magnitude of the potential lost sales varies across different regions (states). This was accurately captured by our model, as the studied market faces regional demands based on regional needs (e.g., high demand for agricultural recreational vehicles in a farm area vs. high demand for sport-tuned recreational vehicles in hunting/leisure areas). Capturing the estimated potential lost sales could boost firm revenue in the United States and Canada by more than 30%. This is equivalent to more than 120 million dollars for the product category here studied.

5. Conclusion

In this paper, we proposed a data-driven approach to estimate potential lost sales and stockout substitutions in a retail distribution network of high-value products. Our model considers particular customer behaviors and inventory management systems that are frequently observed in retail networks of high-value substitutable products. It models the probabilistic behavior of loyal customers in visiting multiple retail centers to find their desired product



(a) Estimated lost sales

(b) Sales

(c) Total product unavailability.

Figure 11: Sales, lost sales, and total product unavailability in a sales group over one year.

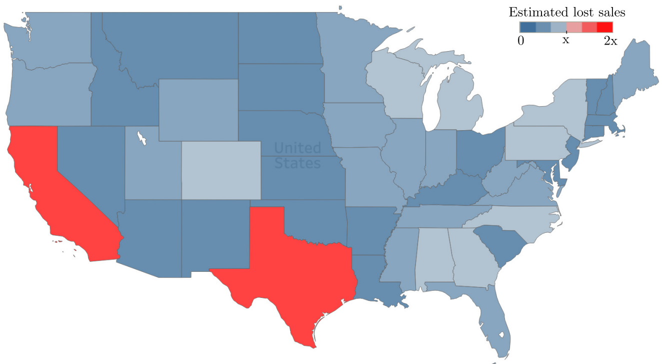


Figure 12: Distribution of the potential lost sales across the United Sates.

and considers lateral inventory transshipments that retailers use to satisfy demand for an out-of-stock product. The only data that the model requires are past aggregate sales data, retailers' inventory logs, and product substitution fitness.

Our approach has several attractive features. First, customer arrival rates are data driven. That is, sales and product availability data are used to estimate customer arrival rates under stockout substitutions. Hence, no assumption is made on the distribution or rate of the customer arrivals. Second, it estimates potential lost sales based on the market sizes of the retail centers using actual sales data. Thus, estimated demand for a large retail center located in a metropolitan area would not be the same as the demand for a small retail center in a suburban area. Third, unlike previous methods, our approach is not limited by the number of products, number of substitutes, choice models, or the number of products that can be simultaneously unavailable. Fourth, it models the probabilistic behavior of loyal customers in visiting multiple retail stores to find their desired product.

The results of implementing the model in a network of recreational vehicle dealerships demonstrate that the actual sales data considerably under- or overestimate demand for most products depending on the combination of factors such as product availability, sales, and retailer market size. Ignoring any of these factors when estimating the demand for a high-value product may undermine the resulting accuracy. However, our analysis on the aggregate potential lost sales showed that the impact of sales is higher than the other factors.

The data-driven approach presented in this paper provides a useful tool to managers, professionals, and researchers for making and assessing supply chain design, assortment planning, and inventory management decisions, considering both expected induced costs and revenues. Beyond recreational vehicle markets, the methodology can also be applied to similar markets such as cars, trucks, luxury goods, appliances, consumer electronics, and any other market where products have significant values and are prone to substitution and multi-source transshipment. Numerous extensions of the approach are possible, notably addressing variations of the core hypotheses.

Our model prevents double-counting returning customers in demand estimates in two ways. It does not count a lost sale for potential demand for an out-of-stock product that can be satisfied by transshipment, and it counts returning customers that are satisfied in the same month once in customer arrival calculations. Estimating unsatisfied customers who might return in subsequent months requires at least some partial information about such customers, which was not available in our case study. Usually, transshipment is set such that any customer who is willing to wait long enough can be satisfied, as was the case in our case study. However, in a case where transshipment or network product availability is restricted on a level that some loyal customers who are willing to wait may be left unsatisfied, the effects of returning such customers in the following months should be excluded from the estimated demand to prevent double-counting such customers in estimated demand. One way to address this is to split customer arrivals into new and returning customers.

Our case study demonstrated that there are significant potential lost sales and customer-compromising product substitutions occurring in large dealership networks. Even with almost steady product availability over multiple periods, a firm may incur massive lost sales during peak demand periods. A dynamic inventory management system targeting higher product availability, notably in high-demand seasons, across the entire retail network may enable a firm to reduce lost sales and product substitutions. This leads to promising avenues for further research, such as embedding the essence of our estimation approach into a smart dynamic dealer ordering approach.

Acknowledgments

The authors thank the anonymous referees and the editors for their helpful comments and suggestions on the earlier versions of this paper.

Notes on contributors

Shahab Derhami is an Assistant Professor of Business Analytics and Operations in the School of Management at Binghamton University. His research interests are in business analytics, machine learning, and supply chain management.

Benoit Montreuil is Professor and Coca-Cola Material Handling & Distribution Chair in the Stewart School of Industrial & Systems Engineering at Georgia Tech. He is Director of the Supply Chain & Logistics Institute, Director of the Physical Internet (PI) Center and PI Lab, and Co-Director of the SIReN Lab on Sentient Immersive Response Networks.

References

Anupindi, R., Dada, M., & Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, *17*, 406–423.

Conlon, C. T., & Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, *5*, 1–30.

Derhami, S., Montreuil, B., & Bau, G. (2021). Assessing product availability in omnichannel retail networks in the presence of on-demand inventory transshipment and product substitution. *Omega*, *102*, 102315.

Dong, L., & Rudi, N. (2004). Who benefits from transshipment? exogenous vs. endogenous wholesale prices. *Management Science*, *50*, 645–657.

Fisher, M., & Vaidyanathan, R. (2014). A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, *60*, 2401–2415.

Fitzsimons, G. J. (2000). Consumer response to stockouts. *Journal of Consumer Research*, *27*, 249–266.

Grahovac, J., & Chakravarty, A. (2001). Sharing and lateral transshipment of inventory in a supply chain with expensive low-demand items. *Management Science*, *47*, 579–594.

Gruen, T. W., Corsten, D. S., & Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Technical Report Grocery Manufacturers of America Washington, DC.

Huang, X., & Sošić, G. (2010). Transshipment of inventories: Dual allocations vs. transshipment prices. *Manufacturing & Service Operations Management*, *12*, 299–318.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Jain, A., Rudi, N., & Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, *63*, 134–150.

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, *55*, 1001–1021.

Mersereau, A. J. (2015). Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing & Service Operations Management*, *17*, 335–349.

Musalem, A., Olivares, M., Bradlow, E. T., Terwiesch, C., & Corsten, D. (2010). Structural estimation of the effect of out-of-stocks. *Management Science*, *56*, 1180–1197.

Nagarajan, M., & Rajagopalan, S. (2008). Inventory models for substitutable products: Optimal policies and heuristics. *Management Science*, *54*, 1453–1466.

Paterson, C., Kiesmüller, G., Teunter, R., & Glazebrook, K. (2011). Inventory models with lateral transshipments: A review. *European Journal of Operational Research*, *210*, 125–136.

van Rijnsoever, F., Farla, J., & Dijst, M. J. (2009). Consumer car preferences and information search channels. *Transportation Research Part D: Transport and Environment*, *14*, 334–342.

Rosales, C. R., Rao, U. S., & Rogers, D. F. (2013). Retailer transshipment versus central depot allocation for supply network design. *Decision Sciences*, *44*, 329–356.

van Ryzin, G., & Mahajan, S. (1999). On the relationship between inventory costs and variety benefits in retail assortments. *Management Science*, *45*, 1496–1509.

Shao, J., Krishnan, H., & McCormick, S. T. (2011). Incentives for transshipment in a supply chain with decentralized retailers. *Manufacturing & Service Operations Management*, *13*, 361–372.

Transchel, S. (2017). Inventory management under price-based and stockout-based substitution. *European Journal of Operational Research*, *262*, 996–1008.

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

Wan, M., Huang, Y., Zhao, L., Deng, T., & Fransoo, J. C. (2018). Demand estimation under multi-store multi-product substitution in high density traditional retail. *European Journal of Operational Research*, *266*, 99–111.

Wee, K. E., & Dada, M. (2005). Optimal policies for transshipping inventory in a retail network. *Management Science*, *51*, 1519–1533.

van Woensel, T., van Donselaar, K., Broekmeulen, R., & Fransoo, J. (2007). Consumer responses to shelf out-of-stocks of perishable products. *International Journal of Physical Distribution & Logistics Management*, *37*, 704–718.

Zhao, X., & Atkins, D. (2009). Transshipment between competing retailers. *IIE Transactions*, *41*, 665–676.

Estimation of potential lost sales in retail networks of high-value substitutable products
(Supplemental Materials)

Shahab Derhami[a], Benoit Montreuil[b,c]

[a]*School of Management, Binghamton University, State University of New York, NY, USA*
[b]*School of Industrial and Systems Engineering, Georgia Institute of Technology, GA, USA*
[c]*Physical Internet Center and Supply Chain Institute, Georgia Institute of Technology, GA, USA*

This electronic companion includes supplemental materials for the article, "Estimation of potential lost sales in retail networks of high-value substitutable products," which has been accepted for publication in *IISE Transactions*. This document is organized as follows:

Section 1: Related research in inventory management *(see Section 2 in the paper)*

Section 2: An example to compute expected sales shares *(see Section 3.1 in the paper)*

Section 3: An example to compute expected portion of unfulfilled customers *(see Section 3.2 in the paper)*

Section 4: Details of parameter tuning *(see Section 4.1 in the paper)*

Section 5: Details of the MASE analysis *(see Section 4.3 in the paper)*

Section 6: Details of the experimental analysis for a sample of products *(see Section 4.5 in the paper)*

1. Related research in inventory management

The mainstream of relevant research in the field of inventory management focuses on the estimation of demand and substitution parameters for the assortment planning problem (Honhon et al., 2010; Goyal et al., 2016). See Shin et al. (2015) and Kök et al. (2015) for comprehensive reviews of these papers. Smith & Agrawal (2000) studied the assortment planning problem under stockout substitution using an exogenous probabilistic model of customer substitution while considering only a single substitution attempt. They showed that static substitution provides bounds on the demand for a product in a dynamic substitution model. Netessine & Rudi (2003) studied optimal inventory policy under stockout substitution for both a centralized inventory management system, in which products are managed by a central decision maker whose objective is to maximize the expected aggregate profit, and a decentralized inventory management system, in which independent decision makers maximize the expected profit generated by their products.

Kök & Fisher (2007) estimated demand and substitution rates for the case of stores operating with a full assortment and high service level (i.e., no stockout substitution, only assortment-based substitution). They used the resulting demand and substitution rates along with the inventory transaction data to find the MLE of the demand and substitution parameters under stockout-based substitution with an EM algorithm. They proposed an iterative heuristic to find the optimal assortment planning under one-level stockout-based substitution, shelf space, discrete maximum inventory levels, and delivery lead time constraints. They presented the results of their model in the supermarket industry.

Jain et al. (2015) studied the impact of sales transactions and stockout timings on the estimation of demand for assortment planning. They studied the problem as a parsimonious multi-period newsvendor problem in which lost sales are unobserved at the times of stockout and knowledge about demand is updated after each period in a Bayesian fashion. They showed that the optimal order quantity when only stockout times are observed is larger than that when complete demand is observed. Their numerical simulations showed that the expected loss in profit decreases significantly when the stockout times rather than stockout events in a period are observed.

Wan et al. (2018) studied the effects of considering customer substitution at both the store and product levels. They used a Bayesian Markov chain Monte Carlo algorithm to estimate parameters for two customer choice models, nested logit and exogenous substitution models, in which customers can substitute at either the store level or the product level. They showed that considering such substitution behaviors increases the initial inventory and expected profit when the profit margin is low.

2. An example to compute expected sales shares

Example 1. Consider a case where the product portfolio consists of only five products (P1,...,P5) with the substitution fitness matrix and customer purchase thresholds presented in Table 1. Assume that retailer 1 has recorded a sales transaction at day 0 for product 1 when it had only products 1 and 3 in stock (i.e., $I_1^0 = \{P1, P3\}$). The potentially demanded products for this sales transaction, $\mathcal{P}_1$, should meet the three conditions mentioned above. $P1 \in \mathcal{P}_1$ because the customer's preferred choice may have been the sold product, but $P3 \notin \mathcal{P}_1$ because it does not meet condition (1) (it was in stock at the time of the sale and therefore, the customer would have purchased it if her preferred choice was P3). P2, P4, and P5 all meet condition (1); however, $P4 \notin \mathcal{P}_1$ because $S_{14} < C_4$ and it does not meet condition (2), implying P1 is not a satisfactory substitute for P4. P5 meets condition (2) but it does not meet condition (3), because $S_{35} > S_{15}$, implying the in-stock P3 is a more preferred substitute for P5 and therefore, the sold product could not be P1 if the preferred choice of the customer had been P5. P2 meets all the three conditions. Hence, $\mathcal{P}_1 = \{P1, P2\}$, and the portfolio receives the following expected sales share for this sales transaction:

$$\begin{aligned}
s_{11}^0 &= \frac{(1-0.85)}{(1-0.85)+(0.9-0.80)} = 0.6, \\
s_{12}^0 &= \frac{(0.9-0.80)}{(1-0.85)+(0.9-0.80)} = 0.4, \\
s_{13}^0 &= s_{14}^0 = s_{15}^0 = 0.
\end{aligned}$$

3. An example to compute expected portion of unfulfilled customers

Example 2. Consider the product portfolio in Example 1. Assume that customers are willing to visit up to three retailers and $\boldsymbol{v} = (0.5, 0.3, 0.2)$. Consider retailer 1 (R1) and two other stores located near it (R2 and R3). Assume that R2 is closer to R1 and the customers of R1 who are willing to visit multiple stores, visit R2 next and then R3. That is, $d_1^1 = R1$, $d_2^1 = R2$, $d_3^1 = R2$. Assume that stores have the following inventories at day 9: $I_1^9 = \{P3\}$, $I_2^9 = \{P2, P4\}$, and

Table 1: Substitution fitness matrix and customer thresholds for Example 1

	P1	P2	P3	P4	P5
P1	1.00	0.90	0.80	0.70	0.80
P2	0.90	1.00	0.70	0.60	0.80
P3	0.80	0.80	1.00	0.60	0.90
P4	0.65	0.65	0.85	1.00	0.70
P5	0.70	0.75	0.80	0.80	1.00
C_i	0.85	0.80	0.80	0.75	0.70
M_i	0.95	0.92	0.92	0.90	0.85

$I_3^9 = \{P1, P5\}$. The probability that a customer whose preferred choice is P1 is satisfied by R1 at day 9, p_{11}^9, is zero because $S_{31} < C_1$ (i.e., P3 is not a satisfactory substitute for P1). If the customer is willing to visit another store, then $p_{21}^9 = max(S_{21}, S_{31}, S_{41}) = 0.9$. For customers that also visit the third store, $p_{31}^9 = max(S_{11}, S_{21}, S_{31}, S_{41}, S_{51})$, which equals one. Taking $\boldsymbol{v}$ into account, the probability of failure is obtained by:

$$\begin{aligned} u_{11}^9 &= 0.5(1-0) + 0.3(1-0.9) + 0.2(1-1) \\ &= 0.53 \end{aligned}$$

4. Details of parameter tuning

Parameters e_{rd}, C_p, and M_p model customer behaviors in accepting an inventory transshipment and making a purchase decision. They can be measured using customer behavior models, field studies, and/or surveys. However, conducting a survey or field study might in practice be highly expensive depending on the size of the market and distribution network. We propose an alternative scenario optimization approach to measure these parameters using simulation and historical data. We developed a digital twin of the entire retail distribution network of the company that comprehensively simulates customers and dealerships' behaviors and their interactions as described in section 1 of the paper (introduction). We used historical data to simulate the entire distribution network for a past model year exactly as it occurred (i.e., the same product availability at all dealerships, and the same customer arrivals). We tuned the customers' behaviors in accepting a substitution or transshipment by testing different values for the respective parameters in an effort to imitate actual customer behavior and obtain simulation results that are a close match to

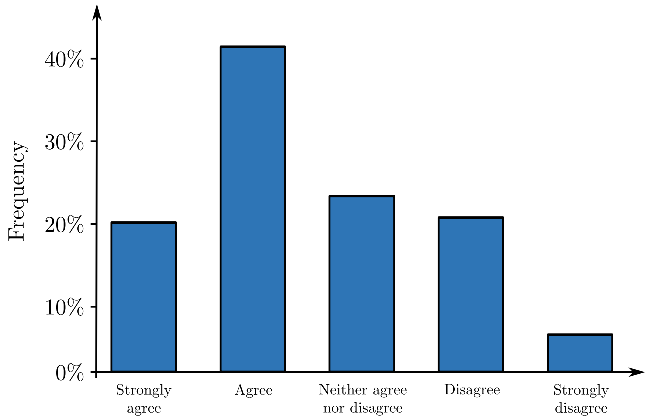


Figure 1: Survey question: customers seeking entry-level (less-expensive) vehicles are more flexible in accepting a substitute than are those seeking high-end (expensive) products.

the historical data. We used inventory and product shipment data (the products dealers received monthly) over one year as inputs to the simulation model to emulate product availability in the network and generated customers using sales data. We evaluated different scenarios of parameters based on the gap between generated sales and inventory transshipments in simulation and the historical sales and inventory transshipment data at the aggregate, product, and dealer levels.

Our empirical study shows that the flexibility of customers in accepting a substitute depends on the price of their desired product. Customers seeking high-end (expensive) products are less flexible in accepting a substitute than are those seeking entry-level (less-expensive) products. This was confirmed by the results of a survey we conducted over more than 100 dealerships; 77% of the dealerships agreed with this statement from a neutral to strong level (see Figure 1 for the details of the survey). In line with this observation, we defined C_p and M_p based on the prices of the desired products. To simplify the scenario optimization, we defined lower and upper bounds for these parameters, assigned them to the least and most expensive products, respectively, and used product prices to interpolate C_p and M_p for all products. Hence, only the lower and upper bounds of these parameters change in each simulated scenario, and C_ps and M_ps are calculated thereafter.

By running more than 50 scenarios of the parameters, we found that setting e_{rd} to 1, 0.99, 0.95, 0.90, 0.85, and 0.80 for same-day, one-day, ..., and five-day or longer deliveries best matches customer behavior in accepting an offered inventory transshipment. For C_p and M_p, we selected three scenarios to obtain the average lost sales along with lower and upper bounds. We termed them the flexible, pickier, and pickiest customer scenarios.The pickier customer scenario best imitates

Table 2: Lower and upper bounds of the thresholds for different customer scenarios.

Customer type	C_{min}	C_{max}	M_{min}	M_{max}
Flexible customers	0.80	0.90	0.90	1.00
Pickier customers	0.82	0.92	0.92	1.00
Pickiest customers	0.85	0.95	0.95	1.00

customer behavior in accepting a substitute among all tested scenarios and reflects the average lost sales. The two other scenarios provide lower and upper bounds on the average lost sales. The lower and upper bounds on C_p and M_p are shown in Table 2 for the three scenarios.

5. Details of the MASE analysis

We computed the mean absolute scaled error (MASE) for three cases. In case one, we computed the ratios for dealer-product level (i.e., we computed monthly absolute errors (AE) for each pair of dealer-product and obtained MAEs across the year for the pair). In cases two and three, we computed the ratios for dealer level (i.e., computed AEs per dealer per month and MAEs across the year for each dealer) and product level (i.e., computed AEs per product per month and MAEs across the year for each product), respectively. Then, we performed regression analyses to study the effects of model type, dealer size, product availability, and product sales volume on MASE ratios aiming to find any combination of these factors that might result in the naïve method performing as good as the model.

To study the effects of dealer size, we clustered dealers into small, medium, and large groups based on their annual sales. The studied products belong to 7 different model types, which were introduced to the regression model as categorical variables. We computed product availability in dealerships on a daily basis. That is, for each pair of dealer-product, we counted the total number of days in a month that the dealer had the product or its satisfactory substitute in stock. We divided the results by the total number of days in the month and computed the average product availability in the year. The result of this process was a two-dimensional matrix showing product availability for a pair of dealer-product.

The regression model on the dealer-product level did not show any statistically significant relationship between MASE ratios and the studied factors. The same result was obtained for the dealership level analysis. However, the product level analysis showed a statistically significant

Table 3: The results of the regression analysis.

Variables	coefficient	p-value
Prod. avg. sales	0.0182	0.007
Prod. avail.	-0.3174	0.185
Prod. avg. sales × prod. avail.	-0.0015	0.875
Model type 2 (binary var.)	-0.1163	0.588
Model type 3 (binary var.)	0.2174	0.143
Model type 4 (binary var.)	0.0968	0.599
Model type 5 (binary var.)	-0.1101	0.667
Model type 6 (binary var.)	0.1255	0.546
Model type 7 (binary var.)	-0.0017	0.992

Table 4: Precision of the model and the naïve method on groups of products with different annual sales.

	$\bar{S}_p \leq Q_1$	$Q_1 < \bar{S}_p \leq Q_2$	$Q_2 < \bar{S}_p \leq Q_3$	$Q_3 < \bar{S}_p$
Average MASE	0.71	0.80	1.02	0.94
Model MAPE (%)	19.76	15.33	22.43	21.69
Naïve method MAPE (%)	34.91	49.91	29.20	25.26

$\bar{S}_p$: products average monthly sales.
Q_i: the ith quartile of products average monthly sales.

relationship between the MASE ratios and the average monthly sales of products. Table 3 shows the details of the regression model for this analysis. The average monthly product sales was the only variable with a significant p-value. Its coefficient is positive, indicating that an increase in the average monthly sales of products is associated to an increase in MASE. That means the naïve method performs close to the model for products with relatively large average monthly sales.

Table 4 compares the precision of the model with that of the naïve method for different groups of products based on their average monthly sales. We divided products into four groups using first, second, and third quartiles of the average monthly sales. As the regression analysis proposed, the MASE increased as the average sales increased. The average MASE reached its highest value for products whose average sales were between the second and third quartiles. The average MASE was 1.02 for this group, indicating that the precision of the naïve method was very close to the proposed model. For all other groups, the model precision was higher than the naïve method (MASE < 1 shows the model outperformed the naïve method).

Analyzing the mean absolute percentage error (MAPE) among the four groups showed that the precision of the model on the first two groups was relatively smaller than the last two groups, but no

Table 5: MAPE (%) of the model and the naïve method in estimating product demands

Prod.	Model	Naïve	**Prod.**	Model	Naïve	**Prod.**	Model	Naïve
1	16.59	45.71	**22**	36.87	27.25	**43**	12.48	45.04
2	14.32	21.66	**23**	26.67	30.11	**44**	20.23	24.48
3	11.79	5.44	**24**	24.87	41.65	**45**	7.03	63.81
4	20.61	29.56	**25**	14.64	4.49	**46**	18.85	41.62
5	28.12	48.51	**26**	6.57	20.87	**47**	28.94	42.86
6	22.46	0.89	**27**	21.42	26.45	**48**	12.58	58.36
7	32.69	73.28	**28**	45.46	96.30	**49**	36.96	58.79
8	14.87	24.60	**29**	21.99	85.00	**50**	36.65	75.90
9	37.46	0.26	**30**	38.11	89.01	**51**	11.16	61.99
10	0.62	58.22	**31**	4.26	1.60	**52**	9.93	64.00
11	30.55	38.85	**32**	40.09	93.46	**53**	8.34	45.25
12	23.61	45.07	**33**	45.65	100.00	**54**	33.26	68.03
13	33.84	31.15	**34**	40.20	92.06	**55**	8.86	49.65
14	39.78	35.08	**35**	45.28	87.70	**56**	8.73	41.55
15	14.46	39.24	**36**	41.70	76.19	**57**	23.37	14.97
16	26.87	24.61	**37**	6.06	12.15	**58**	17.43	56.29
17	29.76	9.25	**38**	22.12	17.56	**59**	3.03	18.56
18	32.91	31.04	**39**	17.14	31.54	**60**	5.83	42.76
19	36.58	13.44	**40**	8.52	96.00	**61**	27.14	57.76
20	17.91	65.41	**41**	34.68	91.71	**62**	13.72	61.94
21	15.85	41.44	**42**	5.36	60.15	**63**	30.48	34.01

trend was detected on the MAPE of the model based on the product sales. However, the precision of the naïve method increased from one group to another as the average sales increased. Improvement in the performance of the naïve method and also the slight performance deterioration of the model in the last two groups contributed to the increase in MASE ratios for groups of products with larger sales. However, the considerable improvement rate of the naïve method suggests that the former factor contributed more to this phenomenon.

The MAPEs of product demand estimates for the model and the naïve method are listed in Table 5.

6. Details of the experimental analysis for a sample of products

Table 6 compares estimated lost sales and relevant variables for the six products we picked to present detailed lost sales calculations. Parameter $\bar{u}_p^T$ is the average product unavailability and calculated by taking the average of u_{rp}^t for all dealers in a sales group over all $t \in T$ (here a

Table 6: Analysis of potential lost sales for six products.

Product	$\bar{u}_i^T$	$\hat{u}_i^T$	Total sales	Total sales shares	Total lost sales
P1	0.445	0.538	3	1.636	5.784
P2	0.444	0.538	0	0.011	0.008
P3	0.438	0.512	3	3.000	3.350
P4	0.120	0.152	9	9.000	1.899
P5	0.326	0.482	2	1.292	0.955
P6	0.315	0.375	2	1.097	0.265

month). A value close to one indicates poor product availability. Parameter $\hat{u}_p^T$ is the weighted average product unavailability and calculated by using dealer market sizes as the weights:

$$\hat{u}_p^T = \frac{\sum_{r \in R} D_r^T \sum_{t \in T} u_{rp}^t}{\|T\| \sum_{r \in R} D_r^T}. \tag{1}$$

This parameter reflects product unavailability in dealerships with respect to their market sizes. Values closer to one indicate that the product was unavailable primarily at larger dealerships. This is not part of the lost sales calculations and is presented in the table only to clarify the subsequent analyses. The total sales shares in the table are computed by summing s_{rp}^t over all dealerships in the sales group for the respective products. It reflects customer arrival for each product.

The $\bar{u}_p^T$ variables are almost equal for P1, P2, and P3, but the potential lost sales for these products differ. P1 has higher lost sales than P2 because the customer arrival rate for P1 (total sales share) was much higher than that of P2. However P3, which has a higher sales share than P1, is estimated to have fewer lost sales. This is because, as shown by $\hat{u}_p^T$, the market sizes of dealers for which product P1 was unavailable were larger than those of P3. As a result, poor product availability at larger dealerships generated more lost sales.

P2 has no recorded sales in the month considered but received a 0.011 unit sales share. This represents sales shares it received from other products sold for which P2 was a potential desired product. P4 had the highest sales and sales share among all products, but it had relatively moderate potential lost sales because the product or its satisfactory substitutes were available during the month at most dealerships. This is shown by the relatively small $\bar{u}_p^T$. However, it exhibited higher potential lost sales than P5 despite that P5 was not as available at dealerships. This is because P4 had higher customer arrivals than P5; hence, it was a highly demanded product, and its unavailability led to relatively larger lost sales. P6 has an almost equal $\bar{u}_p^T$ and sales share to

Table 7: Potential lost sales per dealership for P3.

Dealers	$\bar{u}^T_{rP3}$	Dealer market size		Lost sales for P3
		Total sales	Total lost sales	
D1	0.000	17	0	0.000
D2	0.867	8	5	1.148
D3	1.000	0	1	0.098
D4	1.000	4	2	0.424

P5 but is estimated to have experienced relatively much fewer potential lost sales than P5 due to poor product availability of P5 at large-sized dealerships.

Table 7 compares the potential lost sales of P3 in four dealerships in the same sale group and month following Table 6. The variable $\bar{u}^T_{rP3}$ is the average product unavailability of P3 over one month. Dealer market size (or total demand) is the sum of sales and potential lost sales at a dealership. As the table presents, D1 had P3 or its satisfactory substitutes in stock for the entire month and did not have any potential lost sales for this product. On the other hand, both D3 and D4 could not satisfy any demand for P3 in the entire month and received potential lost sales for P3. However, D4 had higher lost sales due to its large market size. D2 had slightly better availability than D3 and D4 for P3 but received more potential lost sales because its market size was larger than those of both D3 and D4.

References

Goyal, V., Levi, R., & Segev, D. (2016). Near-optimal algorithms for the assortment planning problem under dynamic substitution and stochastic demand. *Operations Research*, *64*, 219–235. doi:10.1287/opre.2015.1450.

Honhon, D., Gaur, V., & Seshadri, S. (2010). Assortment planning and inventory decisions under stockout-based substitution. *Operations Research*, *58*, 1364–1379. doi:10.1287/opre.1090.0805.

Jain, A., Rudi, N., & Wang, T. (2015). Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research*, *63*, 134–150. doi:10.1287/opre.2014.1326.

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, *55*, 1001–1021. doi:10.1287/opre.1070.0409.

Kök, A. G., Fisher, M. L., & Vaidyanathan, R. (2015). Assortment planning: Review of literature and industry practice. In N. Agrawal, & S. A. Smith (Eds.), *Retail Supply Chain Management: Quantitative Models and Empirical Studies* (pp. 175–236). Boston, MA: Springer US. doi:10.1007/978-1-4899-7562-1_8.

Netessine, S., & Rudi, N. (2003). Centralized and competitive inventory models with demand substitution. *Operations Research*, *51*, 329–335. doi:10.1287/opre.51.2.329.12788.

Shin, H., Park, S., Lee, E., & Benton, W. (2015). A classification of the literature on the planning of substitutable products. *European Journal of Operational Research*, *246*, 686–699. doi:10.1016/j.ejor.2015.04.013.

Smith, S. A., & Agrawal, N. (2000). Management of multi-item retail inventory systems with demand substitution. *Operations Research*, *48*, 50–64. doi:10.1287/opre.48.1.50.12443.

Wan, M., Huang, Y., Zhao, L., Deng, T., & Fransoo, J. C. (2018). Demand estimation under multi-store multi-product substitution in high density traditional retail. *European Journal of Operational Research*, *266*, 99–111. doi:10.1016/j.ejor.2017.09.014.