# Retail markdown price optimization and inventory allocation under demand parameter uncertainty

Andrew Vakhutinsky

Oracle Labs, Burlington, MA

email: andrew.vakhutinsky@oracle.com

**Abstract**

This paper discusses a prescriptive analytics approach to solving a joint markdown pricing and inventory allocation optimization problem under demand parameter uncertainty. We consider a retailer capable of price differentiation among multiple customer groups with different demand parameters that are supplied from multiple warehouses or fulfillment centers at different costs. In particular, we consider a situation when the retailer has a limited amount of inventory that must be sold by a certain exit date. Since in most practical situations the demand parameters cannot be estimated exactly, we propose an approach to optimize the expected value of the profit based on the given distribution of the demand parameters and analyze the properties of the solution. We also describe a predictive demand model to estimate the distribution of the demand parameters based on the historical sales data. Since the sales data usually include multiple similar products embedded into a hierarchical structure, we suggest an approach to the demand modeling that takes advantage of the merchandise and location hierarchies.

## 1 Introduction

The e-commerce channel of the retail industry is currently experiencing very fast growth. According to the U.S. Census Bureau US Census, in 2017 e-commerce retail sales in the United States grew by 15.5%, whereas the entire retail sector's revenue grew only by 4.3%. At the same time, profit margins have stayed very low. For example, as reported in (Rigby 2014), Amazon.com has averaged only 1.3% in operating margin over the past three years, compared to 6% to 10% for department/discount stores. Therefore, even a small revenue increase that does not involve additional operating cost translates into a significant profit margin increase for an e-commerce retailer.

A typical e-commerce retailer fulfills orders from multiple warehouses, or fulfillment centers (FC), that are generally geographically dispersed, and thus the cost of shipping an order to a customer can differ significantly depending on the FC. At the same time, when an item approaches its end of life, its inventory levels frequently vary greatly among the FCs. Therefore, maximizing total profit should involve both assigning FCs to customers in order to balance the load among the FCs as well as pricing the item based on each customer's price elasticity and cost of service.

On the other hand, more traditional brick-and-mortar retailers also supply their stores from multiple warehouses. As the stores are usually clustered into so-called price zones, it allows the retailer to execute a flexible pricing policy offering different prices in the price zones with different price sensitivity. In both cases, we have multiple supply centers and customer groups with a variety of demand parameters and service costs.

In this paper, we consider merchandise products with a limited life cycle, such as fashion or electronics. In most cases, these products are ordered by the vendor in a fixed quantity before the start of the sales season and are not replenished during the season. If the demand for a product at its current price is insufficient to sell the entire order amount, a markdown, or permanent price discount, is applied. A large body of literature considers different aspects of markdown and related price optimization problems due to demand uncertainty such as in (Harsha and Perakis 2010). A recent survey of these problems is in (Chen and Chen 2015). Some recent research papers focus on omnichannel retail (Harsha et al. 2019) and optimal inventory levels at FCs (Lei et al. 2018) as well as markdown with an assortment-dependent demand model, (Vakhutinsky et al. 2012) and (Smith and Achabal 1998). In (Lei et al. 2021), the authors propose an approach to solving the problem of joint pricing and fulfillment of multiple customers from multiple FCs. In most of these and other papers, the demand model parameters are assumed to be known and the demand uncertainty is modeled as fluctuations in demand realization. The case of mispricing due to the error in estimating the model parameters is explained by an example in (Gallego and van Ryzin 1997) as the market response to the price changes that may be different from the predicted by the model. In (Mai and Jaillet 2019) the authors solve the problem of price optimization under parameter uncertainty using a robust optimization approach.

In this paper we formulate the markdown price optimization as a stochastic programming problem, that is, to optimize the expected value of either revenue or profit. Based on our experience with the practical implementation of the markdown problem for several retailers, there are multiple business rules essentially aimed at safeguarding against mispricing due to the incorrectly estimated price elasticity of the demand. As under-pricing and inventory depletion early in the season results in especially large loss of revenue, a retailer may limit the percentage of the single markdown price drop, impose the limit on the earliest markdown and/or minimal time between the two consecutive markdowns. In addition to preventing steep discounts generated by an automated price recommendation engine, the latter two restrictions are aimed at improving the accuracy of the demand model before taking further actions. Incorporating the uncertainty of the demand parameters into the decision making process in many cases results in more shallow price discount recommendations, which allows for fewer applications of the business rules, thus improving the quality of the pricing recommendation. In a more general schema of markdown implementation, a retailer runs markdown optimization application on the periodic basis using the observed sales data. As the sales season progresses, the markdown optimization uses parameter estimates based on larger amount of the sales data and with less parameter uncertainty.

Solving the markdown optimization problem requires a comprehensive predictive model, which

would account for various exogenous and endogenous effects typically confronting the retailer, such as a holiday-driven change in customer demand, presence of similar items resulting in demand transference, a.k.a. demand cannibalization, and pricing and promotion effects. Some existing studies in the area of revenue management describe similar models incorporating such effects, for example, (Caro and Gallien 2012) for the case of a fast-fashion retailer, (Ferreira et al. 2015) for the case of an online fashion retailer, and (Ito and Fujimaki 2017) for a grocery retailer. In more recent papers (Vakhutinsky et al. 2019) and (Cohen et al. 2020), the authors proposed the methods to estimate the demand model parameters of individual product items by incorporating the sales of other like items with similar demand characteristics such as price elasticity and promotion lift. Our approach extends this idea to estimate the parameter distribution as a posterior to the Bayesian prior using Markov Chain Monte Carlo (MCMC) simulation similar to the approach laid out in (Cho et al. 2020).

We formulate the markdown optimization model in section 2. First, in section 2.1, the problem is formulated as a basic markdown problem for a single customer group served from the same FC; then, in section 2.2 we provide a formulation for the model with multiple FCs supplying multiple customer groups at specific service costs and describe a min-cost network flow approach to solving the joint inventory allocation and markdown optimization problem. In both cases, we provide a closed-form solution and a gradient descent algorithm for a log-linear demand model and uniform distribution of the price elasticity parameter that allows for a straightforward practical implementation. The predictive modeling methods are described in section 3.

## 2 Joint inventory allocation and markdown optimization problem

In this section, we consider a markdown problem with multiple warehouses or Fulfillment Centers (FC) that supply goods to different groups of customers. The customers can be individual customers such as online shoppers or customer groups sharing the same demand parameters or, more generally, they could be stores or even groups of stores. We assume that demand model parameters are not known exactly but measured within certain limits. More precisely, we assume that each customer has their demand modeled as a function of the product price with parameters that follow a known joint distribution. The problem that we would like to solve is to find the best price to maximize the expected profit derived from selling the given amount of product inventory located at multiple FCs accounting for the costs of handling and shipping the product to the customer groups. We also assume that the sales period is bounded by a certain exit date, after which the product is salvaged at near-zero price. In addition, we assume that there is no replenishment of the product in the sales period at the FC level but the product is replenished at the customer group level when the demand arises. Thus, when a customer group represents a brick-and-mortar store, our model would ignore the situation when the on-hand inventory at the store is shipped but not sold thus incurring the shipping costs without revenue.

Regarding the expected demand as a function of price $\bar{d}(p)$ and expected revenue $\bar{R}(p) = p\bar{d}(p)$,

we make the following two assumptions:

1. $\bar{d}(p) > 0$ for all $p > 0$ and monotonically decreasing function of price $p$.
2. The expected revenue $\bar{R}(p) \to 0$ as $p \to 0$ or $p \to +\infty$.

The first assumption does not allow modeling the demand as piece-wise linear function and the second assumption excludes the constant elasticity power law demand model (e.g., in the form $d(p) = \alpha p^{-\beta}, \beta = const > 0$). The time and price parameters, as well as demand output, are considered as continuous variables, which does not limit the practical applicability of the model as the integration in the model can be straightforwardly replaced with a summation. In the discussion that follows, $p = p(t)$ is assumed to the be the function of time.

### 2.1 Basic markdown optimization problem: a single customer group served from a single FC

In order to demonstrate the concept of markdown optimization under parameter uncertainty, we consider a basic markdown optimization problem of a single customer or a homogeneous customer population served from the same FC.

Suppose there are $S$ units of inventory that must be sold within the time horizon $T$. There is no replenishment during this time period and all inventory unsold by exit time $T$ is salvaged at a negligibly small near-zero price. We formulate the problem as finding the optimal pricing policy, that is, the price of an item as a function of time, $p : t \to p(t), t \in [0, T)$. In general, the price policy function should satisfy some markdown constraints, the most important of which is to be non-increasing. Other examples of constraints are limited price drops during markdown events and limited intervals between the drops. We denote the set of the allowed price policies as $P$. We assume there is a demand function $d(t, p; \theta)$ of pricing policy $p \in P$ at the time $t$ that also depends on function parameter vector $\theta$. Since in practice it is impossible to measure the exact parameter values, we assume that $\theta$ is a realization of a random variable $\Theta$, which follows a distribution given by p.d.f $f(\theta)$ with the domain $dom(f)$. $\theta$ may include such components as price elasticity of demand, promotion lift, and sensitivity to the reference price.

In general, the demand as a function of time depends on the prior pricing decisions. For example, it could be prior promotions changing the customer reference price point, or customer perception of the fair price, and also affecting the future demand through the "pantry loading" effect (see, e.g., (Cohen and Perakis 2018)). Another example is the inventory depletion effect when prior sales diminish the current inventory amount, which may negatively affect the current demand (see, e.g., (Vakhutinsky et al. 2012) and (Smith and Achabal 1998)).

Since inventory is limited, we can define the sell-off time $T_0$ for the given pricing policy as the time of the inventory exhaustion under the current pricing policy. It is determined as the solution to the following equation:

$$\int_0^{T_0} d(t, p; \theta)\, dt = S$$

Note that for some combinations of the demand models and pricing policies the sales may last infinitely long. In this case, $T_0 = \infty$. Define the end-of-sale time $T_1 = \min(T, T_0)$, that is, the time when the vendor runs out of the inventory or out of time whichever happens earlier. The total revenue under pricing policy $p$ can be expressed as

$$R(p, \theta) = \int_0^{T_1} p(t) d(t, p; \theta)\, dt \tag{1}$$

Then the expected revenue can be expressed as

$$\bar{R}(p) = \int_{dom(f)} R(p, \theta) f(\theta)\, d\theta, \tag{2}$$

and the markdown optimization problem is to find the optimal pricing policy $p^*$ that would maximize the expected revenue:

$$p^* = \arg\max_{p \in P} \bar{R}(p) \tag{3}$$

While in general, the solution to the above problem may not be computationally tractable when the dimension of the parameter vector $\theta$ is relatively high, it can be applied to practical cases when only one or two components of the demand function parameters are estimated with uncertainty. In many practical cases, it is price sensitivity or price elasticity parameter that is the most difficult to estimate for several reasons. For example, if there are no observed price changes of a product, the only estimation of the price elasticity can be obtained from observing similar products. Another example is when the discounts occur during the seasonal decline for the product demand. In this case, due to the so-called endogeneity effect, when the lower prices correlate with the lower demand for the product, the price sensitivity coefficient can be estimated at a much lower absolute value or even with the opposite sign. The usual practice, in this case, is also to consider the estimation of the broader product set or apply the range of endogeneity-mitigating instruments. In both of the examples, the estimation of the price elasticity will be within a certain range. In section 3 we propose a technique to obtaining the empirical distribution function of the coefficient estimates. In what follows in this section we demonstrate how this technique can be applied to a widely used special case of log-linear demand model.

### 2.1.1 Special case of log-linear demand model and uniform distribution of the price coefficient

In this section we consider one of the simplest demand models, the log-linear demand model, which is formulated as follows:

$$d(p, t) = s(t) \alpha e^{-\beta p} \tag{4}$$

where $s(t)$ is a seasonality factor independent of price. Without loss of generality, we can assume that $\int_0^T s(t)\, dt = 1$. Then it can be easily shown (see (Vakhutinsky et al. 2019) for the discrete-time case) that the optimal price policy for model (4) with exactly known parameters consists of a single

optimal price $p^* = \max(p_{opt}, p_{icp}) = \max(\frac{1}{\beta}, \frac{1}{\beta}\log\frac{\alpha T}{S})$ by finding the revenue-maximizing value $p_{opt}$ from the first-order optimality condition $\frac{\partial(pd(p,t))}{\partial p} = 0$ and the inventory-clearing price (ICP) value $p_{icp}$ from the total demand over period $T$, $\int_0^T d(p,t)\,dt = S$. Therefore, when considering the basic log-linear demand model, instead of function $p(t)$, we will use single variable $p$ omitting its functional dependency on $t$.

Note that the first-order optimality condition for the revenue-maximizing price is equivalent to the price elasticity of demand being equal to one:
$-\frac{\partial d(p,t)}{\partial p}\frac{p}{d(p,t)} = 1$. Also, the price is set to maximize the revenue when the inventory is high enough relative to the base demand $\alpha$ and the length of the sales period, or $\log\frac{S}{\alpha T} > -1$. In this case, some of the inventory is left unsold. The above is summarized in the following equation:

$$p^*_{det} = \begin{cases} \frac{1}{\beta}\log\frac{\alpha T}{S} & \text{if } \log\frac{\alpha T}{S} > 1; \\ \frac{1}{\beta} & \text{otherwise.} \end{cases} \tag{5}$$

Note that in the optimal solution to the deterministic markdown optimization problem with any demand function, the optimal pricing policy is such that the inventory is never completely sold out before the end of the sales period. Otherwise, the price can be raised and the same amount of the inventory can yield higher revenue.

When the price coefficient $\beta$ is not known exactly, it is easy to see that underpricing the item and running out of the inventory before the end of the season when the value of $\beta$ is overestimated results in greater revenue loss than overpricing the product when the value of $\beta$ is underestimated by the same amount. We will illustrate by considering the following case of a specific distribution.

Assume that the price coefficient $\beta$ is a realization of the random variable with uniform distribution $\beta \sim U(\beta_1, \beta_2)$. For a fixed price $p$ denote by $\beta_0$ the value of the parameter $\beta$ at which the inventory is cleared within the sales period $T$:

$$\beta_0 = \frac{1}{p}\log\frac{\alpha T}{S} \tag{6}$$

For values of $\beta$ in the $[\beta_1, \beta_0]$ interval, the inventory is sold out yielding the revenue $pS$. For $\beta$ in the $[\beta_0, \beta_2]$ interval, since the inventory is not depleted, the sales per unit of time are $\alpha T e^{-\beta p}$. If $\beta_0 < \beta_1$, the inventory is never depleted; if $\beta_0 > \beta_2$, the inventory is always sold out. Combining the above, the revenue defined in (1) can be expressed as:

$$R(p,\beta) = p \cdot \begin{cases} S & \text{if } \beta \in [\beta_1, \beta_0]; \\ \alpha T e^{-\beta p} & \text{if } \beta \in [\beta_0, \beta_2]. \end{cases} \tag{7}$$

By substituting expression for the revenue from (7) and p.d.f. for the uniform distribution $U(\beta_1, \beta_2)$, $f(\beta) = \begin{cases} \frac{1}{\beta_2 - \beta_1} & \text{if } \beta \in [\beta_1, \beta_2]; \\ 0 & \text{otherwise} \end{cases}$, into the general expression (2) for the expected revenue,

we obtain

$$\bar{R}(p) = \begin{cases} pS & \text{if } p < p_1; \\ \frac{1}{\Delta\beta}(S\log\frac{\alpha T}{S} + S(1-\beta_1 p) - \alpha T e^{-\beta_2 p}) & \text{if } p \in [p_1, p_2]; \\ \frac{\alpha T}{\Delta\beta}(e^{-\beta_1 p} - e^{-\beta_2 p}) & \text{if } p > p_2. \end{cases} \tag{8}$$

where we use notation $\Delta\beta = \beta_2 - \beta_1$, $p_1 = \frac{1}{\beta_2}\log\frac{\alpha T}{S}$ and $p_2 = \frac{1}{\beta_1}\log\frac{\alpha T}{S}$.

Notice that when the value of $\beta$ is known exactly, that is, $\beta_1 = \beta_2 = \beta$, the last line in the equation (8) can be computed as

$$\lim_{\Delta\beta\to 0}\frac{\alpha T(e^{-\beta_1 p} - e^{-\beta_2 p})}{\Delta\beta} = -\alpha T\frac{\partial e^{-\beta p}}{\partial\beta} = \alpha T p e^{-\beta p},$$

which coincides with the expression for the revenue for the revenue-maximizing price as stated earlier.

In order to find the revenue-maximizing price $p^* = arg\max_p \bar{R}(p)$, we differentiate the function for $\bar{R}(p)$ from equation (8) over $p$ and obtain:

$$\bar{R}'(p) = \frac{\partial\bar{R}(p)}{\partial p} = \begin{cases} S & \text{if } p < p_1; \\ \frac{1}{\Delta\beta}(\beta_2\alpha T e^{-\beta_2 p} - \beta_1 S) & \text{if } p \in [p_1, p_2]; \\ \frac{\alpha T}{\Delta\beta}(\beta_2 e^{-\beta_2 p} - \beta_1 e^{-\beta_1 p}) & \text{if } p > p_2. \end{cases} \tag{9}$$

It is easy to see that $\bar{R}'(p_1) = \frac{1}{\Delta\beta}(\beta_2 - \beta_1)S > 0$ and $\bar{R}''(p_1) = -\frac{1}{\Delta\beta}\beta_2^2\alpha T e^{-\beta_2 p} < 0$. Therefore, the first order condition $\bar{R}'(p) = 0$ delivers an optimal solution

$$p^* = \frac{1}{\beta_2}\log\frac{\beta_2}{\beta_1}\frac{\alpha T}{S} \tag{10}$$

when $p^* < p_2$, which is

$$\log\frac{\alpha T}{S} > \beta_1\frac{\log\beta_2 - \log\beta_1}{\beta_2 - \beta_1} \tag{11}$$

The price defined in (10) can be thought of as the modified ICP defined in the first line of (5). Interestingly, when the price coefficient is uncertain, the modified ICP is greater than the ICP in the deterministic case for any $\beta \in [\beta_1, \beta_2)$.

It is also interesting to compare conditions (11) and those used to apply ICP in (5). First, the right-hand side of (11) converges to one as $\Delta\beta \to 0$; second, it can be seen that $\beta_1\frac{\log\beta_2 - \log\beta_1}{\beta_2 - \beta_1} < 1$ when $\beta_2 > \beta_1$, which can be interpreted that a higher level of inventory is required to apply ICP when the price coefficient is not known exactly.

When condition (11) is not satisfied, $\bar{R}'(p_2) > 0$ and the optimal revenue-maximizing price is found in the interval $[p_2, \infty)$ as the solution to the first order condition:

$$p^* = \frac{\log\frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1} \tag{12}$$

Similarly to the revenue-maximizing optimal price in the second line of (5), the optimal price defined in (12) does not depend on the inventory. That is, in this case the optimal markdown solution will have some unsold inventory at the exit date. Also, $\lim_{\beta_1 \to \beta_2} \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1} = \frac{1}{\beta_2}$, which coinsides with the optimal price defined in (5).

Interestingly, in the case of an uncertain price coefficient the optimal revenue-maximizing price is higher than the price that can be "naively" set to the reciprocal of the price coefficient mean value. In our case of the uniform distribution it can be proved by comparing $p^*$ defined in (12) to $\frac{2}{\beta_1+\beta_2}$ using their Taylor expansions. Indeed, after setting $x = \frac{\beta_2}{\beta_1} - 1$, it can be seen that

$$\frac{\log(x+1)}{x} > \frac{1}{1+x/2}$$

Finally, we summarize the solution to the optimal markdown pricing when the price coefficient is uniformly distributed in the $[\beta_1, \beta_2]$ interval, as follows:

$$p^* = \begin{cases} \frac{1}{\beta_2} \log \frac{\beta_2}{\beta_1} \frac{\alpha T}{S} & \text{if } \log \frac{\alpha T}{S} > \beta_1 \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1}; \\ \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1} & \text{otherwise.} \end{cases} \tag{13}$$

## 2.2 Joint inventory allocation and markdown optimization problem for multiple customer groups served from multiple FCs

It is a common practice that the demand for a limited life span item has to be fulfilled from several FCs. The demand is usually coming from several distinct groups with varying demand parameters. For example, the groups may be determined by the geographic locations with different socio-economic characteristics that affect their price elasticity. In this section we consider the problem of optimal markdown pricing of an item when its inventory is spread over several FCs. We assume that inventory at all FCs has the same exit date, after which it is salvaged at near-zero price. In addition, the cost associated with selling from each FC to each customer group may be distinct. In the example of the geographically dispersed customer groups, it could be delivery cost from an FC to a geographic location.

### 2.2.1 General case of multiple customer groups served from multiple FCs

Suppose there are $M$ FCs serving $L$ customer groups that may be defined either by their locations or other parameters. At each FC $m$, there is initial product inventory $S_m$. The product has its remaining life time $T$. Given the product demand function $d_\ell(p)$ at customer group $\ell$, and unit delivery cost to the customer group $\ell$ from FC $m$ as $c_{m\ell}$, the optimization problem is to find the group-specific prices $p_\ell$ and amount of inventory $x_{m\ell}$ delivered from $m$ to $\ell$ that maximizes profit, subject to FC inventory constraints.

Denote the total amount of inventory allocated to customer group $\ell$ as $S_\ell = \sum_{m \in M} x_{m\ell}$. Then the revenue-maximizing markdown price $p^*_\ell(S_\ell)$ for that group can be found by applying equation

(3), which now explicitly depends on the allocated inventory. The optimal expected revenue for that location becomes $\bar{R}^*_\ell(S_\ell) = \bar{R}(p^*_\ell, S_\ell)$. The total profit-optimization problem now can be formulated as follows

$$\max \sum_{\ell \in L} (\bar{R}^*_\ell(S_\ell) - \sum_{m \in M} c_{m\ell} x_{m\ell}) \tag{14}$$

subject to:

$$\sum_{\ell \in L} x_{m\ell} \leq B_m \tag{15}$$

$$S_\ell = \sum_{m \in M} x_{m\ell} \tag{16}$$

$$x_{m\ell} \geq 0 \tag{17}$$

where $B_m$ is the amount of inventory at FC $m$. That is, the constraint (15) determines the upper bound of the inventory delivered from FC $m$.

In practically all cases allocating a greater amount of the inventory to be sold to the customers cannot reduce the revenue as extra inventory can always be left unsold. On the other hand, in order to derive more revenue by selling more inventory within a given time period, the sales price should be lowered, which results in diminishing marginal return of inventory allocation. Mathematically speaking, the second derivative of the optimally priced revenue as a function of allocated inventory is negative implying that the function is concave. Therefore, after the sign of the objective function (14) is changed, the optimization problem (14–17) becomes equivalent to a well-studied min-cost network flow problem with convex costs. One of the solution methods can be found in (Bertsekas et al. 1987).

We propose another algorithm described below as Algorithm 1 that iteratively solves the Optimal Inventory Allocation problem. The algorithm takes a hyper-parameter $\delta_0$ to determine its initial step length.

**Algorithm 1** Optimal Inventory Allocation

```
Construct an initial feasible solution x_{mℓ} satisfying constraint (15)
n ← 1
while n < N and termination criteria not satisfied
    δ ← δ_0/n
    obtain f_{mℓ} as the optimal solution to MCNF(δ)
    x_{mℓ} ← f_{mℓ}
    S_ℓ ← Σ_{m∈M} x_{mℓ}
    b_m ← B_m − Σ_{ℓ∈L} x_{mℓ}
end while
return x_{mℓ}
```

At each iteration of the Algorithm 1, we solve the min-cost network flow problem, MCNF($\delta$) described next. We start by defining the critical product allocation amount $\sigma_\ell$ as a solution to the

following equation:

$$\frac{\partial \bar{R}^*_\ell(S_\ell)}{\partial S_\ell}|_{\sigma_\ell} = \min_m c_{m\ell} \tag{18}$$

That is, $\sigma_\ell$ can be thought of as the maximal possible allocation of the inventory that can still improve the profit generated by the customer group $\ell$. We will use $\sigma_\ell$ to scale the amount of inventory to be allocated to the customer group. Next, we design the network consisting of three layers of nodes as shown in the example in Figure 1(a) for two FCs and two customer groups. The nodes in the top layer correspond to the FCs; the nodes in the bottom and intermediate layers correspond to the customer groups. The latter serve to "consolidate" the allocations from the FCs to the customer groups. There are arcs connecting each FC to each customer group node that is reachable from the FC as not all FCs may serve all customer groups. The cost of each arc in this layer is the delivery or service cost of the unit of the product from the FC to the customer group. The network flow problem starts with the current feasible solution $x_{m\ell}$ units of flow in the top arc layer. Consequently, the amount of flow sent from the consolidator nodes to the customer group nodes is $S_\ell = \sum_{m\in M} x_{m\ell}$. The cost of the arc from the "consolidator" node to the customer group node is $-\partial \bar{R}^*_\ell(S_\ell)/\partial S_\ell$. Note the negative sign that is used to solve the minimization problem. The top-layer nodes serve as flow sources with capacity $B_m$ (framed in red in the diagram). The bottom-layer nodes serve as what is called "sinks" in the network flow terminology. The capacity of each sink is set to the interval $[(S_\ell - \sigma_\ell\delta_n)^+, S_\ell + \sigma_\ell\delta_n]$. That is, the flow is feasible but there are constraints to limit the change of the total allocated product for each customer group within progressively tighter boundaries. At each call to the MCNF algorithm, the new network flow $f_{m\ell}$ is returned that decreases the total cost of the flow in the network thus increasing the total profit. The Optimal Inventory Allocation algorithm terminates when it fails to improve the objective function above a certain threshold or the maximal number of iterations is reached. There are several well-studied fast polynomial-time algorithms to solve the min-cost network flow problem. Some of them can be found in (Ahuja et al. 1993).

Below, we make certain observations regarding the solution to the problem. We define the cost of an undirected cycle in the directed flow network as the sum of all arc costs in the cycle with the cost of an arc traversed in the reverse direction as negative of the original forward arc cost. It is illustrated in Figure 1(b) where the reverse arcs are shown by dotted lines. Assigning a positive flow to the reverse arc means reducing the allocation of the inventory to the customer group. Most of the min cost network flow algorithms work by iteratively finding a negative cost cycle and increasing the flow along the cycle within the flow constraints. Borrowing the terminology from the network simplex method, we call a network degenerate if it has a zero-cost cycle. The network degeneracy may occur in practice, for example, when the retailer has flat shipping rates that are the same for all origin-destination pairs. It is easy to see that in this case the marginal revenue $\partial \bar{R}^*_\ell(S_\ell)/\partial S_\ell$ in the optimal solution is the same for each customer group $S_\ell$. The marginal revenue exceeds the shipping cost when there is no leftover inventory and equal to the shipping cost when there is some unsold inventory left at the end of the sales season.

When the network is non-degenerate, the links between the top and the middle layers with the

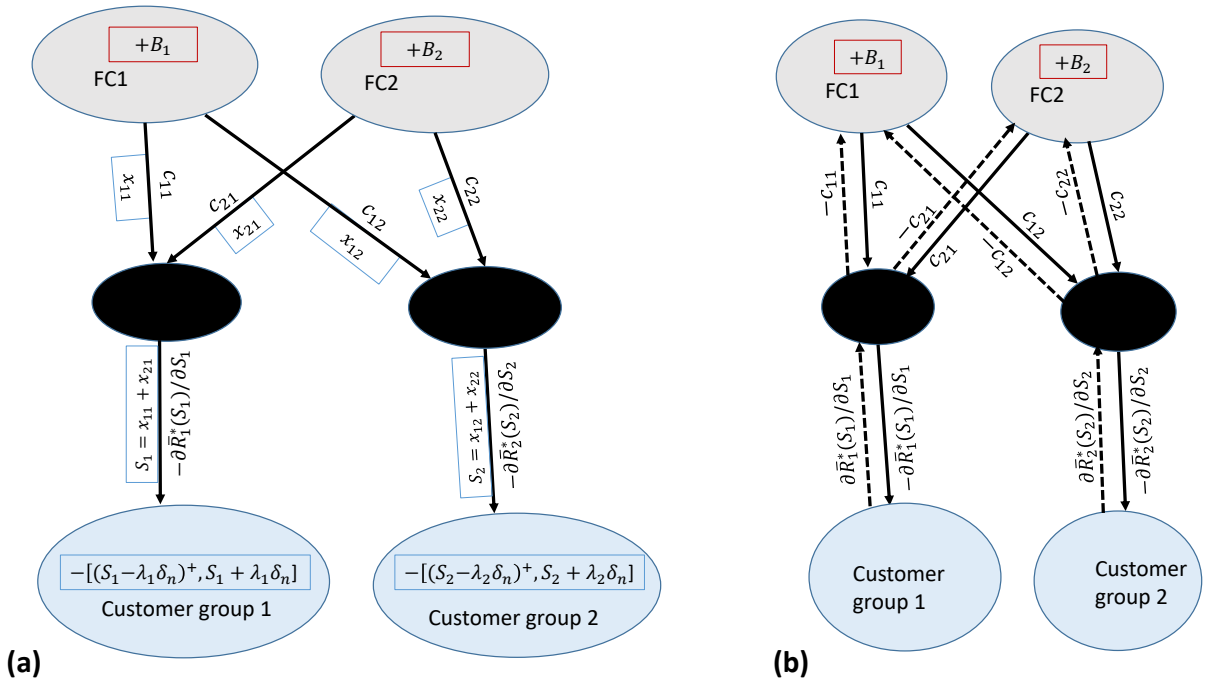


Figure 1: Illustration of network flow graph

non-zero flow, that is, with $x_{m\ell} > 0$, cannot form a cycle. For example, in the network shown in Figure 1 at most three of the four links in the top layer can carry the positive flow. It can be easily proved that in this case a non-zero flow cycle will have a positive cost in one of the directions, which means that reducing the flow along this cycle improves the objective function implying the suboptimality of the solution. One implication of this observation is that if the shipping costs are sufficiently different, the number of links between the FCs and intermediate consolidating nodes does not exceed $|M| + |L| - 1$. Another observation is that each customer group is supplied from at least one FC. Since by the assumption the demand function stays positive for any price, then for any delivery cost values, there is always a positive (although very low) inventory allocation that can make this allocation profitable. In other words, if there is zero allocation of inventory, a sufficiently small reallocation from another customer group will carry overall positive marginal profit. One implication of this observation and the number of customer groups supplied from two or more FCs is at most one less than the number of FCs.

These properties of the optimal solution may provide the following two managerial insights:

- Since in many cases the number of FCs in the supply chain is significantly lower than the number of customer groups, additional constraints may need to be imposed on planning the supply to increase its robustness by connecting the customer groups to more than one FC. One such measure could be imposing hard or soft constraints on link capacities.

- Most min-cost flow algorithms also compute so-called node potentials, or dual costs, for the nodes with binding supply constraints. One interpretation of the dual cost is the improvement

of the objective function per unit of extra supply at the node. This information may be useful when there is an additional opportunity of transporting inventory between the nodes. Usually, it involves a significant fixed cost the problem becomes what is known as the network design problem. Its solution may serve as an additional decision tool for planning load balancing between the FCs.

### 2.2.2 Multiple customer groups served from multiple FCs with the price coefficient uniformly distributed

In this section, we derive a closed-form expression for the marginal revenue to be used in Algorithm 1 in the special case of uniformly distributed price coefficient described in section 2.1.1.

In order to find the derivative of the revenue-maximizing function of the allocated inventory, $\partial \bar{R}^*_\ell(S_\ell)/\partial S_\ell$, we substitute the expression for the optimal price $p^*$ from (13) to the expression (8) for the expected revenue $\bar{R}$ to get the optimal revenue as a function of the allocated inventory $S$, $\bar{R}^*(S)$ and take its derivative:

$$\partial \bar{R}^*(S)/\partial S = \begin{cases} \frac{1}{\beta_2} \log \frac{\alpha T}{S} - \frac{\beta_1}{\beta_2} \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1} & \text{if } \log \frac{\alpha T}{S} > \beta_1 \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1}; \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Note that when the allocated inventory exceeds the critical threshold $\alpha T \exp(-\beta_1 \frac{\log \frac{\beta_2}{\beta_1}}{\beta_2 - \beta_1})$, adding more inventory does not affect the optimal markdown revenue.

## 3 The hierarchical demand model

In this section, we consider several approaches to estimating the demand model parameters based on the sales history of multiple merchandise items and the hierarchical structure of the retail merchandise. The latter allows for grouping of different but similar items into the increasingly larger groups while moving from the individual items at the leaves of the hierarchy tree to its root. Another issue guiding our choice of demand modeling is the requirement for the model to exhibit certain properties, which narrows the choice of the demand model among the family of the of the general predictive models. The main limitation comes from the fact that the predicted demand must be a monotonically decreasing function of the product price, which makes the most commonly used machine learning approaches like ensemble methods (Random Forest, Gradient Boosting) or deep learning neural networks inapplicable. A new type of approach based on building a decision tree and applying isotonic regression as described in (Aouad et al. 2020) is designed to build a monotonic function of price but it does not provide the consistent price elasticity measure. Therefore, we limit our scope to the main two groups of the parametric methods: discreet choice models based on the Multinomial Logit (MNL) and regression methods using generalized linear models (GLM). The latter includes log-linear, Poisson, and Negative Binomial regressions. In both groups, the prediction is based on a linear function of several attributes including price, seasonality

or seasonal time shocks, promotion lifts, and post-promotion fatigue or pantry-loading. The MNL-based methods consider this linear function as the product utility and predict demand as a market share. Their common disadvantage is the dependency on the estimation of the so-called no-purchase option coefficient, which is usually an unobservable variable. On the other hand, these methods account for the demand transference among similar products and have superior performance for the categories of products with high substitutability. The GLM-based methods usually do not explicitly model the demand transference. However, recently in the marketing science literature, there were reports of the models that account for the presence of similar products and their demand cannibalization effects ((Rooderkerk et al. 2013)) using the similarities of their customer attributes such as brand, color, size, etc. By focusing specifically on the sales history of the items, we leave the potential impact of accounting for other data such as social media and the clickstream out of the scope of the model.

The main hurdle in applying these methods to predict the sales of a particular item at a particular store or by a specific customer group is that most of these sales are very low volume and thus have a very low signal-to-noise ratio. In addition, the price changes may be very infrequent and in a very small range, which makes the estimation of the price sensitivity even more difficult. Furthermore, the sales history of an individual item may contain the number of observations too low to provide sufficient statistical power for the estimation of several parameters. All these factors make parameter estimation very unreliable when it is based on the data from the sales of an isolated item. It can, however, be mitigated by considering multiple items sharing similar parameters or the same item sold at different locations. As locations and merchandise items usually form a certain hierarchy coming from the entire chain down to price zones and individual stores when the location hierarchy is considered and to departments, categories, classes, and individual items for the merchandise hierarchy. This hierarchy coupled with clustering the items at the lower level of the merchandise classification provides a convenient tool for estimating the demand model parameters at the levels of the hierarchy with sufficient statistical power. (Cohen et al. 2020) provides a methodology to determine at which level each parameter is estimated in addition to building clusters of similar items for the parameter estimation.

We use these ideas to build estimation methods taking advantage of the existing hierarchical structure as well as the clusters built for estimation purposes. The output of the methods is the probability distributions of some of the parameter estimates.

## 3.1 The predictive demand model

### 3.1.1 Assumptions

We assume the availability of the general sales data aggregated over a short time period like a day or week rather than transaction-level data reflecting individual purchases. That is, we assume that for each item $i$ there are observations from historic sales data that consist of triples $(s_{it}, p_{it}, r_{it})$ representing, respectively, sales, price, and a binary promotion flag for the dated time period $t$. In many cases, zero sales are not part of the data and in addition to that, the out-of-stock (OOS)

periods are not clearly flagged. The combination of these two factors makes it impossible to distinguish between the zero demand and OOS observations. In addition, sometimes promotions are not clearly marked either or there could be several types of promotions with different promotion lifts marked by the same flag. In the next section, we propose a simple heuristic to distinguish among these cases.

We also assume that the merchandise hierarchy is known and adequately reflects the similarity between the retail items. In the next section, we describe how multiple items with a varying magnitude of demand and prices can be pooled together. We assume that textual descriptions of the items are present from which it is possible to extract attributes as described in (Kanani et al. 2015).

### 3.1.2 Attributes of the demand model

In this section, we describe the attributes together with the demand output variable and how they are used by the model. We also describe some pre-processing procedures that are carried out to form an observation pool at higher levels of the hierarchy. Below, we use the notation $x_i(t)$ to denote the value of variable $x$ observed for product $i$ at time $t$.

- **demand:** Since multiple items that are pooled together may have exhibit similar behavior but have different mean demand, e.g., for the same item at different stores, it is necessary to estimate the so-called base demand, which can be either approximated by the mean demand if the item price changes are not significant or estimated as an intercept-type constant.

  Since it is not possible to distinguish between zero-demand periods and unavailability of a slow-moving product when the zero-inventory periods are not reliably identified in the data or appear as missing data, we consider the sufficiently long sequences of the missing or zero-sales periods as OOS. The length threshold is determined by the probability threshold of encountering a certain number of consecutive zero-demand periods, which is computed as $N_{thresh} = -\frac{\log p_{thresh}}{\lambda}$ where $\lambda$ is the parameter of the Poisson distribution. Since the observed demand $d_i(t)$ is effectively zero-truncated, the maximum likelihood estimator of the parameter $\lambda$ is obtained from solving the following equation:

  $$\frac{\lambda}{1 - e^{-\lambda}} = \bar{d_i}$$

  where $\bar{d_i}$ is the sample mean of the observed demand for product $i$ (Johnson et al. 2005). The threshold probability value is usually selected to be in the order of 0.01, meaning that the probability that a sequence of zero-demand periods with the length exceeding the threshold due to zero demand does not exceed 1%. For example, if the sample mean of the observed zero-truncated demand is 2.3, then the estimated mean demand is about 2, which means that the probability of zero demand in three or more consecutive periods is below 1%. Therefore, it is reasonable to assume that these would be the OOS periods.

- **price:** In order to pool together similar merchandise items with different prices such as different size packaging of the same product, we replace their price values with relative deviation from the average or from the initial non-discounted price: $\tilde{p}_i(t) = \frac{p}{p_0} - 1$. The estimated coefficient of this variable is the price elasticity of demand when the demand is modeled as exponentially dependent on price.

- **promotion flag:** As promotions come in various kinds and flavors, there should ideally be a vector of promotion flags reflecting which campaigns were run and which promotion tools were used in each case. However, very few retailers keep historical records of the promotions. In some cases when there are no promotion indicators or only some of the promotions are identified, promotion and regular sales can be separated by applying an unsupervised or, respectively, semi-supervised clustering algorithm (Zhou et al. 2004) in the two-dimensional price-demand space. We denote the given or inferred promotion Boolean variable as $r_i(t)$.

- **post-promotion:** This is the well-known effect caused by the so-called promotion fatigue and pantry loading leads to sales dip following the promotion period. We define this variable as 0-1, or Boolean, indicating whether there was a promotion in the previous week.

- **similarity:** This is the set of variables reflecting the amount of the demand for this item that is cannibalized by other items in the current assortment and the effects of their prices. Following (Rooderkerk et al. 2013), we define two variables here, one for the total of the assortment item similarity, the other for the total of item prices weighted by their similarity. The latter reflects the competitive effect of the prices of the other items. The similarity between the items is computed based on the item attributes extracted from their textual descriptions. We denote the similarity and price similarity variables as $u_i(t)$ and $v_i(t)$, respectively.

- **seasonality:** We consider seasonality as an external modulation process, which is independent of the pricing and promotion controls applied by the retailer. Therefore, together with the seasonal events such as holidays and seasons of the year, we include other demand shocks including local events and extreme weather with the impacts that are measurable either through historic observations or other means. As the majority of the peak sales periods is determined by annual holidays, some of which may fall on different weeks of the calendar year (e.g., Easter), we mark the corresponding holiday periods by their occurrences rather than calendar dates. Since in most cases the peak sales happen in the weeks preceding the holidays and the holiday and post-holiday periods are often characterized by the drops in sales, we define seasonality variables for three or four weeks per holiday. Assuming each holiday-related week is indexed by its index $h$, we define a set of dummy variables $w_h(t)$ indicating whether period $t$ is the time period associated with the holiday.

We denote the vector of the explanatory variables as

$$\vec{x}_i(t) = (1, \tilde{p}_i(t), r_i(t), r_i(t-1), u_i(t), v_i(t), \{w_h(t) | h \in H\}) \tag{20}$$

and the vector of the coefficients as

$$\vec{\beta}_i = (\beta_i^0, \beta^{price}, \beta^{promo}, \beta^{postPromo}, \beta^{sim}, \beta^{priceSim}, \{\beta_h^{holiday}(t)|h \in H\}) \tag{21}$$

Note that only base demand coefficient $\beta_i^0$ is specific to product $i$.

If the demand is modeled as Poisson distribution with the mean:

$$\lambda_i(t) = E(d_i(t)|x_i(t)) = e^{\vec{\beta}_i'\vec{x}_i(t)} \tag{22}$$

Then the probability mass function is given by:

$$p(d_i(t)|\vec{x}_i(t); \vec{\beta}_i) = \frac{\lambda_i(t)^{d_i(t)}}{d_i(t)!} e^{-\lambda_i(t)}$$

Given a series of $T$ observations $(d_i(t), x_i(t))$, $t = 1, \ldots, T$, for product set $M$, we use the regularized regression to find the estimate of the $\beta$ coefficients by maximizing the difference between the logarithm of the likelihood and the regularization term:

$$l(\vec{\beta}) = \sum_{i \in M} \sum_{t=1}^{T} (d_i(t) \vec{\beta_i}' \vec{x}_i(t) - e^{\vec{\beta_i}' \vec{x}_i(t)}) - \frac{T|M|}{2} \alpha \|\vec{\beta}_i\|_2^2 \tag{23}$$

where $\alpha$ is the regularization penalty parameter estimated via cross-validation. This technique, similar to ridge regression, can reduce overfitting. More specifically, while maximizing the logarithm of the likelihood expressed by the first term, the minimization of the $L_2$ norm of the coefficient vector leads to suppressing high values of the estimated parameters.

Since the log-likelihood function is concave, the Newton-Raphson method can be applied to find the maximum likelihood estimator (MLE) of $\vec{\beta}_i$. There are multiple widely available implementations to compute this MLE. E.g., see (Seabold and Perktold 2010).

In case when the mean demand $\bar{d}_i$ is easy to estimate for each product from the sales data, it can be entered into the equation (22) as exposure:

$$\lambda_i(t) = E(d_i(t)|x_i(t)) = \bar{d}_i e^{\vec{\beta}'\vec{x}_i(t)} \tag{24}$$

In this case, the estimated coefficient vector $\vec{\beta}$ defined in (21) becomes:

$$\vec{\beta} = (\beta^{price}, \beta^{promo}, \beta^{postPromo}, \beta^{sim}, \beta^{priceSim}, \{\beta_h^{holiday}(t)|h \in H\}) \tag{25}$$

and becomes the same for all products.

The variable vector defined in (20) becomes:

$$\vec{x}_i(t) = (\tilde{p}_i(t), r_i(t), r_i(t-1), u_i(t), v_i(t), w_h(t)|h \in H) \tag{26}$$

Therefore, the number of estimated parameters is reduced by $|M|$.

### 3.2 Hierarchical Estimation

In this section, we describe the hierarchical approach to estimating the coefficients of the demand model defined in the previous section. We assume there are at least three levels of the hierarchy that we call category (about 100–1,000 individual SKU/UPC items), class (about 10-100 items) and individual items that may include several SKU/UPC of varying size/color but of the same style. The terminology may differ among retailers but this hierarchy is almost always present. Most of the demand model coefficients are estimated at the class level except for the base demand $\beta^0$, which is estimated at the item level. The price elasticity is also estimated at the class level except for the classes with strong seasonality. For those classes, the post-seasonal drop in the demand often coincides with significant price discounts to the extent of a positive correlation between price and demand changes, which may result in a positive sign of the estimated price coefficient. This is a well-known endogeneity effect when correlation exists between an explanatory variable and the error term (or unobserved factors) in a model. This effect has been extensively studied and several methods have been proposed for its correction. E.g., see (Berry et al. 1995) for the automobile prices and (Mumbower et al. 2014) for the airline prices. However, each of these and other approaches depends on the availability of the so-called instrumental variables and requires specific knowledge of product characteristics, which are not always available. Instead, we estimate the price elasticity using other products from the same category that do not exhibit strong seasonality or if necessary go even higher in the hierarchy tree. Another application of the hierarchical estimation approach is prediction of the sales parameters of the new products that can inferred from considering several similar products at a certain level of the hierarchy.

#### 3.2.1 Uncertainty Estimation

We propose two methods to estimate the parameter uncertainty based on Hessian at the maximum of the log-likelihood function and using the Markov Chain Monte Carlo (MCMC) method.

- **Hessian-based:** Since the Newton-Raphson method effectively approximates the log-likelihood function as a paraboloid with its shape determined by the Hessian of the function, the variance of the parameter estimations can be computed as the negative of their diagonal elements in the inverse of the Hessian, which provides the means for building the confidence interval by assuming the normal distribution around the estimation. For example, the 95%-confidence intervals can be found as

  $$\hat{\beta}^j \pm 1.96\sqrt{-H_{jj}^{-1}}$$

  where $\hat{\beta}^j$ is the MLE for the $j$-th parameter and $H_{jj}$ is the $j$-th diagonal element of the Hessian. After that, the probability distribution can be modeled as normal with $(\hat{\beta}^j, \sqrt{-H_{jj}^{-1}})$ parameters or uniform distribution for better tractability.

- **MCMC-based:** We utilize Markov Chain Monte Carlo (MCMC) sampling as described in (Cho et al. 2020). The samples are drawn from a prior distribution with the distribution of

the next sample dependent on the last sample to form a Markov chain that converges to the posterior distribution $g(\vec{\beta}|d_i(t), \vec{x}_i(t), t = 1, \ldots, T, i \in M)$ (Gelman et al. 2013). At the higher hierarchy level the parameters are estimated using an uninformative prior distribution. At the lower levels, the prior distribution is the posterior from the higher level. The posterior distribution at the item level is used to model the parameter uncertainty.

From the practical implementation perspective, these two methods have their advantages and disadvantages. Newton-Raphson MLE methods may be relatively fast but computationally unstable whereas MCMC-type methods are usually stable but may be computationally slow especially for large samples. In some practical implementations, it may be beneficial to use a hybrid approach by applying Newton-Raphson methods at the higher hierarchy levels and MCMC at the item level with Gaussian prior obtained from the resulting Hessian at the higher level.

# 4 Summary and Future Work

Parameter uncertainty is an important issue to consider in order to develop a practically applicable prescriptive solution based on a parametric predictive model since in most applications it is impossible to obtain a sufficiently precise estimator of the model parameters. The joint markdown pricing and inventory allocation model described in this paper can be readily implemented as the retail decision-making process to optimize the sales performance when multiple warehouses of fulfillment centers are used to supply the merchandise to multiple groups of customers with highly heterogeneous demand parameters. Finally, the demand model described in this paper serves as a necessary part of the entire solution bridging the gap between the observed data and optimization model by providing the interval estimates of the model parameters.

The network flow optimization model laid out in the paper can be used to provide additional managerial insight in situations when the retailer can benefit from additional one-time actions like load balancing between the FCs by transporting large quantities of inventory at a fixed cost or improving the fault-tolerance of the supply chain between the FCs and the retail locations by increasing the density of the connections.

# References


Ahuja RK, Magnanti TL, Orlin JB (1993) *Network Flows: Theory, Algorithms, and Applications* (USA: Prentice-Hall, Inc.), ISBN 013617549X.

Aouad A, Elmachtoub A, Ferreira K, McNellis R (2020) Market segmentation treesl. Available at arXiv:1906.01174v2 .

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Bertsekas D, Hosein P, Tseng P (1987) Relaxation methods for network flow problems with convex arc costs. *SIAM J. Control and Optimization* 25(5):1219–1243.

Caro F, Gallien J (2012) Clearance pricing optimization for a fast-fashion retailer. *Operations Research* 60(6):1404–1422.

Chen M, Chen ZL (2015) Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information. *Production and Operations Management* 24(5):704–731, ISSN 1937-5956, URL http://dx.doi.org/10.1111/poms.12295.

Cho S, Ferguson M, Pekgun P, Vakhutinsky A (2020) Estimating personalized demand with unobserved no-purchases using a mixture model: An application in the hotel industry. Available at SSRN: https://ssrn.com/abstract=3700177 or http://dx.doi.org/10.2139/ssrn.3700177.

Cohen M, Perakis G (2018) Promotion optimization in retail. Available at SSRN: https://ssrn.com/abstract=3194640 or http://dx.doi.org/10.2139/ssrn.3194640.

Cohen M, Zhang R, Jiao K (2020) Data aggregation and demand prediction. Available at SSRN: https://ssrn.com/abstract=3411653 or http://dx.doi.org/10.2139/ssrn.3411653.

Ferreira K, Lee B, Simchi-Levi D (2015) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 69–88.

Gallego G, van Ryzin G (1997) A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research* 45:24–41.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis*. 3rd edition.

Harsha P, Perakis G (2010) Tractable markdown optimization for single and multiple items under uncertainty. *INFORMS Manufacturing and Services Operations Management (MSOM) Conference*.

Harsha P, Subramanian S, Uichanco J (2019) Dynamic pricing of omnichannel inventories. *Manufacturing & Service Operations Management* 21(1):47–65, URL http://dx.doi.org/10.1287/msom.2018.0737.

Ito S, Fujimaki R (2017) Optimization beyond prediction: Prescriptive price optimization. *KDD 2017*, 1833–1841 (Halifax, NS, Canada).

Johnson NL, Kemp AW, Kotz S (2005) *Univariate Discrete Distributions (third ed.)* (Wiley-Interscience).

Kanani P, Wick M, Pocock A (2015) Attribute extraction from noisy text using character-based sequence tagging models. *Machine Learning for eCommerce workshop* (NIPS).

Lei Y, Jasin S, Sinha A (2018) Dynamic joint pricing and order fulfillment for e-commerce retailers. *Manufacturing & Service Operations Management* 20(2).

Lei YM, Jasin S, Uichanco J, Vakhutinsky A (2021) Randomized product display (ranking), pricing, and order fulllment for e-commerce retailers .

Mai T, Jaillet P (2019) Robust multi-product pricing under general extreme value models. Papers 1912.09552, arXiv.org.

Mumbower S, Garrow LA, Higgins MJ (2014) Estimating flight-level price elasticities using online airline data: A first step toward integrating pricing, demand, and revenue optimization. *Transportation Research Part A: Policy and Practice* 66:196–212.

Rigby D (2014) Online shopping isnt as protable as you think. https://hbr.org/2014/08/online-shopping-isnt-as-profitable-as-you-think.

Rooderkerk RP, van Heerde HJ, Bijmolt THA (2013) Optimizing retail assortments. *Marketing Science* 32(5):699–715, URL http://dx.doi.org/10.1287/mksc.2013.0800.

Seabold S, Perktold J (2010) statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Smith SA, Achabal DD (1998) Clearance pricing and inventory policies for retail chains. *Management Science* 44(3):285–300.

US Census (2017) *Latest quarterly e-commerce report. http://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf.*

Vakhutinsky A, Kushkuley A, Gupte M (2012) Markdown optimization with an inventory-depletion effect. *Journal of Revenue and Pricing Management* 11:632–644.

Vakhutinsky A, Mihic K, Wu SM (2019) *Journal of Pattern Recognition Research* 14:1–21.

Zhou D, Bousquet O, Lal TN, Weston J, Schlkopf B (2004) Learning with local and global consistency. *Advances in Neural Information Processing Systems 16*, 321–328 (MIT Press).