A Practical Price Optimization Approach for Omnichannel Retailing

Pavithra Harsha

IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, pharsha@us.ibm.com

Shivaram Subramanian

IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, subshiva@us.ibm.com

Markus Ettl

IBM T. J. Watson Research Center, Yorktown Heights, NY 10598, msettl@us.ibm.com

Consumers are increasingly navigating across sales channels to maximize the value of their purchase. The existing retail practices of pricing channels either independently, or matching competitor prices, are unable to achieve the desired profitable coordination between channels. We engaged with three major retailers over two years and developed omnichannel pricing (OCP) solutions in partnership with IBM Commerce to overcome these challenges. We implement an integrated data processing and machine learning framework that enables estimation of location-specific cross-channel price elasticities and competitive effects. We develop an integrated OCP optimization formulation to profitably coordinate prices for non-perishable products offered across channels and store locations while satisfying practical constraints on volume and price. The resultant optimization formulations for discrete choice demand models are non-convex and NP-hard, and we prescribe practically efficient mixed-integer programs that can be used to recover (near) optimal solutions. An OCP implementation in two categories for a major retail chain projected a 7% profit lift while preserving the sales volume. This benefit was achieved by lowering online prices and optimally raising and lowering location specific store prices. This integrated pricing approach allows the retailer to be competitive and preserve market share without aggressively matching the low price of e-tail giants.

Key words: Omnichannel, pricing, multi-product, mixture, MNL demand, nested model, cross-channel, elasticity, censored demand estimation, machine learning

1. Introduction

Omnichannel retailing is a recent trend sweeping companies across the industry (Brynjolfsson et al. 2013, Bell et al. 2014). It is revolutionizing how companies engage with consumers by creating a seamless customer shopping experience across the retailer's multiple sales channels (Kahn 2018). This is because today's consumers navigate across the channels with ease to make purchases. Using smart phones, in-store shoppers can visit the mobile or web store of the same retailer or its competitors to find better deals and finalize a purchase. Omnichannel retailing also includes the use of advanced order fulfillment practices such as initiating ship-from-store fulfillment for e-commerce orders and offering a buy-online-pick-up-in store fulfillment option to enhance the convenience of

receiving a product. Traditional retailers require such capabilities to survive in a highly competitive marketplace that is witnessing a fast pace of online sales growth (US Census Bureau 2017 reported that online sales grew 14-16% compared to the previous year) and an ever-increasing market share gain by e-tailers whose price-transparent product offerings eat into e-tail margins as well as store sales (e.g. due to 'showrooming').

Many of today's large retailers started as single channel retailers and their supply chain was designed to ensure maximum efficiency and scale in that channel. These retailers subsequently opened additional sales channels, and supported common and channel-specific assortments, to increase their customer base. However, these channels largely operated independently of each other in 'silos', with limited transparency and data integration even within the organization. From the perspective of machine learning and optimization technologies, many retailers continue to rely on demand forecasting, pricing optimization, and inventory management methods that are channel specific. Such tools largely ignore the multi-channel shopping path of today's customers, as well as the potential efficiencies of omnichannel retailing.

As part of a joint partnership agreement with IBM Commerce, a leading provider of merchandizing solutions, we engaged with three major omnichannel retailers over a period of two years, who faced many of these challenges. These retail chains primarily operated two sales channels, brick-and-mortar stores, and online. Their existing retail practices were to price channels independently, or simply match channel and competitor prices. These practices failed to achieve the degree of coordination required between the channels in order to be profitable, achieve sales targets, satisfy brand-price image goals for products within an assortment, etc., and remain competitive with large e-tail giants. Given the increase in the number of digital channels (e.g., social, mobile) and the dynamic nature of the marketplace, the scale and speed of executing pricing decisions were equally critical. Our work focuses on developing an end-to-end unified commerce solution that overcomes these challenges by integrating data, machine learning and price optimization across the different sales channels, and keeping in mind the infrastructure and operational requirements of a deployable solution. Given this context, the contributions of this paper are as follows:

1. **Omnichannel machine learning framework to quantify cross-channel demand substitutions:** We develop an integrated data-analytics framework that combines multiple data streams to track concurrent channel-specific pricing variations over time, as well as relative price variations across the products, and locations. We then employ machine learning techniques on this integrated dataset to estimate own and cross-channel and cross-product price elasticity of demand by location. This framework enables us to rapidly model omnichannel demand for a large retail chains that manage many store locations. This scalability is achieved through a dimensionality reduction of cross-location and cross-channel interaction variables.

2. **Omnichannel price optimization:** We study the omnichannel pricing (OCP) problem of nonperishable products in the context of developing a regular (or baseline) pricing solution. The OCP objective is to maximize near-term profit across all channels and locations while also meeting critical long-term channel, volume, and price image goals.

   (a) *Single product pricing:* We show that the attraction choice model-based OCP is non-convex in the presence of certain simple linear pricing constraints that are employed in practice and is NP-Hard. We employ specialized ratio transformations, along with relaxation and linearization techniques (RLT), to derive an exact mixed integer program (MIP) reformulation of the non-linear OCP that can be solved to optimality using standard commercial solvers.

   (b) *Multi-product pricing:* Using nested attraction demand models, we additionally manage cross-product demand interactions within any channel. We exploit the concave structure of the nest attraction function and likewise obtain a tractable MIP approximation that also admits a variety of practical cross-product pricing constraints. We show that this MIP yields an effective upper bound that is asymptotically exact and can be used to obtain (near-) optimal solutions to the OCP problem.

3. **Implementation and business value assessment:** We perform a business value assessment as a part of our OCP implementation for one of the major retailers in the United States. For 100 products in the two product categories that we analyze, we benchmark the forecast accuracy of our censored data machine learning approach, and find that the estimated cross-channel price sensitivity to demand can be up to 50% of the own channel price elasticity. The OCP prices yield a projected profit lift of 7% over the retailer's legacy pricing system, while preserving the sales volume and satisfying several other critical business and pricing goals. The positive feedback from the retailer during this engagement and similar feedback from other retailers resulted in IBM Commerce deploying a proprietary version that has been in production since 2014. This solution was showcased in the smarter-commerce global summit, which included a presentation on its capabilities by the retailer.

1.1. Literature Review

Modeling the consumer preferences in an omnichannel environment is a first step that can pave the way for coordination of channel strategies. Recent papers in the marketing literature have explored consumer dynamics in a multi-channel environment (Zhang et al. 2010), in particular, consumer migration across sales channels such as web and catalog (Ansari et al. 2008) or online and brick channels (Chintagunta et al. 2012) respectively. Goolsbee (2001) finds significant cross-price elasticity between online sales and stores sales and concludes that channels cannot be treated as

4

Harsha, Subramanian and Ettl: *Omnichannel demand modeling and price optimization*
Article submitted to *INFORMS Journal on Optimization*; manuscript no. (Please, provide the manuscript number!)

separate markets. Similar to the consideration in the above papers, we estimate consumer channel preferences using discrete choice models. We propose using location-specific demands models, as price elasticity can depend on an area's household income, demography, etc (Mulhern et al. 1998).

From an operational perspective, there is substantial academic literature on single and multi-product pricing problems (for example, see the survey papers by Bitran and Caldentey 2003, Elmaghraby and Keskinocak 2003, Chen and Simchi-Levi 2012) but to the best of our knowledge, the focus has been on single channel pricing. In contrast, the focus of this paper is on an integrated multi-channel multi-location pricing problem in the presence of cross-channel and cross-product demand interactions and important operational considerations.

From the perspective of price optimization using customer choice models, several papers in the literature have analyzed a variety of parametric and non-parametric approaches in the context of cross-product demand substitution. *Discrete choice demand models* are one of the commonly used demand functions to model consumer choice in marketing, economics, and more recently, in the revenue management literature (McFadden 1974, Urban 1969). Some of its well-known examples include the *multinomial logit* (MNL) and the *multiplicative competitive interaction* (MCI) demand models. For the MNL demand model, Hanson and Martin (1996) show that the profit as a function of the prices is not *quasi-concave.* Aydin and Porteus (2008), Akçay et al. (2010) explored this problem further and show that the resultant profit function is unimodal in the price space. Meanwhile, Song and Xue (2007), Dong et al. (2009) proposed a market share variable transformation to recover an equivalent objective function that is concave in the space of the market share variables. This transformation idea for MNL demand models was later extended to a general class of attraction models by Schön (2010), and Keller et al. (2014). However, these papers ignore certain practical linear pricing restrictions that are critical to real-life applications. In Proposition 1, we demonstrate that this transformation even for a simple linear price monotonicity constraint yields a non-convex and non-linear formulation in the market share space.

Our work on single product pricing relates to the price optimization problem using a mixture of attraction demand models. Keller et al. (2014) recognize this as an open problem and develop a local optimal heuristic solution by employing a convex approximation of the demand model. Li et al. (2017) characterize the objective as a sum of quasi-concave functions and present alternative efficient methods that converge to a stationary point. On the other hand, by exploiting the sparse multi-location structure and the discrete nature of prices, we propose specialized transformations and tractable global optimization methods to solve large-scale problem instances that arise in practical omnichannel operations. Bront et al. (2009) propose a related assortment optimization formulation in the context of network revenue management for the mixed-MNL demand model, which can be viewed as a special case of the discrete pricing problem we study here. Furthermore,

our proposed formulation relies on certain specialized cuts (applicable to pricing problems), resulting in two orders of magnitude improvement in run times even for medium-sized problem instances (see Section 4.2 for details).

A few papers have explored the use of other parametric demand models in the context of multi-product pricing to maximize expected profit. Li and Huh (2011) extend the concavity result of MNL to nested MNL models when the price sensitivities are constant within a nest. Gallego and Wang (2014) relax this assumption and provide conditions under which the objective function is unimodal. Rayfield et al. (2015) provide an approximation method when prices are bounded by discretizing the intrinsic value of a nest. Davis et al. (2016) study this problem in a discrete price setting assuming a specific quality price ordering of products. They show that the number of feasible price vectors in a given nest is polynomial in the number of choices and price levels and derive an exact solution. In the multi-product OCP problem that we study, we use a similar nested attraction function. We extend the problem to a multi-location setting and optimize a mixture of nested attraction demand models. Unlike prior works, we develop a multi-objective MIP formulation that satisfies a variety of practical pricing constraints and yields effective upper and lower bounds that asymptotically converge and can be solved to achieve (near) global optimal solutions.

Subramanian and Sherali (2010) studies a multi-product pricing using a hybrid MNL demand model, and provide a piecewise linear approximation of market size, while managing a variety of practical business rules. Non-parametric approaches to multi-product pricing have been explored by Rusmevichientong et al. (2006) and Aggarwal et al. (2004) using heuristic approaches and approximation algorithms, and Ferreira et al. (2015) using reference price effects.

We briefly mention a few related papers on price optimization for a single channel. Caro and Gallien (2012) discuss an end-to-end development and implementation for clearance pricing at the physical stores of the fast-fashion retailer Zara. Fisher et al. (2017) study the competition-based dynamic pricing for a pure online retailer involving censored data estimation in the presence of competitor prices and stock-out situations. Lei et al. (2018) study the joint dynamic pricing and fulfillment problem for a pure e-commerce retailer.

Finally, there is emerging literature related to other important aspects of omnichannel retailing beyond pricing, such as inventory operations which include managing e-commerce ship-from-store fulfillments decisions (Zhu et al. 2017, Jalilipour Alishah et al. 2015, Govindarajan et al. 2018) and understanding the impact of buy-online-pickup-in-store implementations (Gallino and Moreno 2014, Gao and Su 2016).

2. Data framework to quantify cross-channel demand substitution

In an omnichannel environment, customers navigate across channels and retailers to finalize a purchase that maximizes their own benefit. Therefore, a fundamental aspect that omnichannel

demand models must capture is the channel switching behavior of consumers, i.e., the cross-channel substitution effects. For illustration purposes, let the set J denote the brick price zones of a retailer, which are geographical clusters of brick stores that offer the same price in order to avoid competitive shopping patterns across stores[1].We simply refer to them as brick store locations. Assuming price is the only driver of demand,

$$D_{B_j} := D_{B_j}(p_{B_j}, p_O) \; \forall j \in J, \qquad D_O := D_O(p_O, p_{B_1}, p_{B_2}, ...) \tag{2.1}$$

where D_{B_j}, D_O and p_{B_j}, p_O are the demands and prices for brick-store location B_j and the online channel O respectively. Here, store demand is a function of its own store price and the online price, and because of non-competitive nature of price zones, it is assumed to be independent of other store prices. Online channel demand, on the other hand, is a function of the online price and *all* the brick-and-mortar store prices because the online channel virtually connects all the physical stores.

Legacy forecasting systems for pricing treat the online channel just as an additional independent store location (from a data preprocessing standpoint and hence for all downstream operations). This prevents the modeling of cross-channel effects. While the legacy systems can estimate the impact of online causals on any store demand by including the online causals as modeling features (see Eq. 2.1), they fail to accurately quantify the interactions in the reverse direction. This is because the sheer number of physical store locations (ranging up to a few thousand), and the usage of location-specific pricing considerably increases the dimensionality and makes this task impractical. It also precludes an accurate quantification of location-specific impact of store causals on the online demand. Note that besides price, there are a variety of other location-specific demand influencers such as store promotions, competitors and local events, whose cross effects are additionally important to capture in an omnichannel demand model.

We overcome this challenge by employing a geographical partitioning of the online store (transactions of which originate from a continuum of customer zip-codes) into discrete virtual online store locations (virtual stores, for brevity) using brick store locations J. Next, we assume that the customers within the zip-codes associated with any virtual store, choose to purchase from this virtual store or the physical store(s) in that location and are not influenced by the prices (and other causals) in other physical store locations, i.e.,

$$D_O(p_O, p_{B_1}, p_{B_2}, ...) = \sum_{j \in J} D_{O_j}(p_O, p_{B_1}, p_{B_2}, ...) = \sum_{j \in J} D_{O_j}(p_O, p_{B_j}). \tag{2.2}$$

This is a reasonable assumption because customers from one location typically do not have prior visibility into the prices offered by stores farther away[2]. By gainfully localizing online demand

---

[1] This method of clustering is commonly employed by retailers prior to price optimization (e.g., retail trade analysis).

[2] The assumption was empirically verified. Adding other location prices as features did not improve the out-of-sample forecast accuracy for either channel using the data and method described in Section 6.

using geographical partitions, we reduce the dimensionality of the price effects to be estimated from $O(|J|)$ to a distributed $O(1)$ across J locations, enabling us to reliably and tractably estimate location-specific cross-channel effects (here $|.|$ refers to cardinality of a set).

While partitioning online transactions to their respective virtual stores, it is important to track the final fulfillment destination of a sale because omnichannel retailers offer buy-online-pickup-instore options and execute ship-from-store fulfillments, and often, the legacy point-of-sales do not always record the differences in customer purchase, product sourcing, and delivery locations.

The proposed virtual stores can be created using appropriate geographical clustering tools (e.g., retail trade analysis). We now describe our approach for a retailer with a strong physical presence. First we cluster the omnichannel retailer's stores into zones using a k-means clustering algorithm based on the latitude-longitude coordinates of the stores. A volume weighted clustering will yield similar results as store counts in a region are correlated with volume of sales. We map each online store transaction to the zone whose centroid is nearest to its fulfilment destination zip code, thereby merging the spatial continuum of online transactions into discrete virtual stores. Next, this zone-tagged transaction data is aggregated by purchase channel and zone. The data at this granularity allows us to simultaneously track channel-specific pricing variations across time, as well as relative price variations across the substitutable products, sales channels and locations. We use this zone-tagged data to calibrate zone level cross-channel demand models described in the following section.

In Fig. 1 we provide a data-visualization example of the geo-spatial clustering a retailer's brick-and-mortar stores (more than 1500) into 50 zones. The figure also shows the zonal distribution of the sales (the volume is proportional to the pie size) and channel share between brick (red) and online (blue) for one product category. Observe the heterogeneity of the online channel share across zones (e.g., 4% to 11%). We find a geographically consistent channel preference pattern for other product categories as well.

2.1. Choice of demand models and its impact on price optimization solutions

There are a variety of demand model options that capture cross-channel effects including parametric vs. non-parametric models, and black-box machine learning methods versus more specialized methods. Each option has its own impact on the performance of the forecasting and optimization modules in the solution, in terms of accuracy, tractability, (near) optimality, and user experience.

For example, choosing a blackbox demand prediction model (e.g., regression trees, ensemble methods) has obvious benefits including readily available software packages, public benchmarks, and stable implementations, and potentially improved forecast accuracy. But consider the case where the feature set employed within the blackbox model includes a vector of channel prices that can each take any one of values from the set I. In an omnichannel system, the channel demands
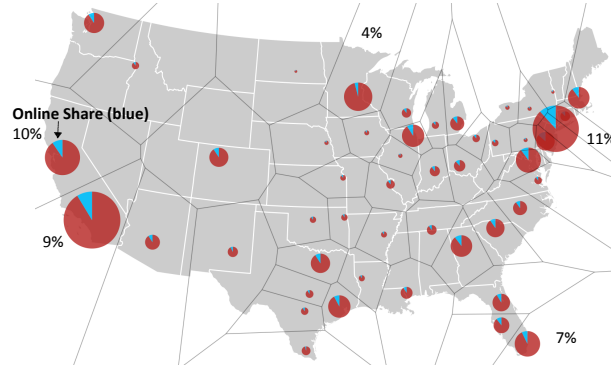
Figure 1   **Distribution of sales over 50 zones for a product category. Sales volume is proportional to the pie size. The pie in each zone shows the relative frequency of brick-and-mortar sales and online sales.**

are correlated. This means we require all channel price values to be specified in order to generate a single blackbox demand prediction for all channels (or even a single channel's demand). Therefore, assuming M denotes the set of channels, fully specifying the omnichannel demand (at all channel price vectors) requires generating $|I|^{|M|}$ predictions using the calibrated demand model, unless specialized constraints are added to limit this combinatorial explosion (e.g., see Ferreira et al. 2015 for the use of a reference anchor price). On the forecasting side, when the underlying estimation problem is non-convex (e.g., as with deep learning) and solved heuristically (e.g. using stochastic gradient methods), the prediction model can change whenever it is retrained. The resultant omnichannel demand solution can recommend drastically different channel prices (even with a near identical dataset), which can be viewed by the user as an erratic response from the application.

The use of simpler parametric models that limit the combinatorial explosion (by way of specification) and adopt simple convex estimation methods (e.g., log-log, hybrid MNL models wherein total sales are spread down to channel choice) can also adversely impact downstream optimization and model tractability, making it difficult to find (near) optimal prices. The ability to find optimal solutions is not only of theoretical importance. In practice, heuristic approaches can induce an inconsistent pricing response from the application (e.g., profit increases after a constraint is added) resulting in unsatisfactory user experience and loss in credibility. We show using an actual OCP price-matching application in Appendix E that this capability directly influences user-experience, which in turn can decide whether the deployed application will be accepted or rejected by users.

Our proposed optimization approaches to demand estimation and pricing take into account these requirements, while matching the forecast accuracy achieved by alternative methods.

3.   Omnichannel demand model for a non-perishable product

In the next two sections, we focus on analytical models for a single non-perishable product and then in Section 5 we extend the models to the multi-product setting.

Consider an omnichannel retailer selling a single non-perishable product using M *sales* channels to customers in J locations. Let $V \subset M$ be the set of virtual channels like website, mobile, social, which are partitioned into virtual stores by location $j \in J$. Let p_{jm} be the price of the product sold in channel $m \in M$ and location $j \in J$ and $\mathbf{p}_j$ be the corresponding vector of prices in all channels at location j. Note that p_{jm} is the same across $j \in J$ for virtual channels $m \in V$. Let $\mathbf{D}_j(\mathbf{p}_j)$ be the vector of demands originating from location $j \in J$ in all the channels. As motivated in Section 2, we assume that the demand in a specific channel and location depends on the attributes of all channels at that location. We refer to this representation as the omnichannel demand model.

We use discrete choice demand functions to model consumer channel demand in an omnichannel environment as a product of market size and the channel choice probability as follows:

$$D_{mj}(\mathbf{p}_j) = \tau_j \ \frac{f_{mj}(p_{mj})}{1 + \sum_{m' \in M} f_{m'j}(p_{m'j})} \quad m \in M, j \in J. \tag{3.1}$$

Here τ_j is the market size of location j and $f_{mj}(p_{mj})$ is the attraction function of customers in location j to channel m. The market size term represents the measure of consumers interested in the product and the channel choice probability term represents the relative attractiveness of a channel-choice over all choices that includes the no-purchase option, whose attractiveness without loss of generality is normalized to 1. If the attractiveness of a channel drops (for example, due to a channel price increase) then that channel share of the product reduces, and it get distributed among the other channels. This attraction structure models the cross-channel substitution (i.e., switching) behavior of consumers across the M sales channels and the no purchase option.

Examples of the attraction function for demand models include the MNL demand model where $f_{mj}(p_{mj}) = e^{a_{mj} + b_{mj} p_{mj}}$, the MCI demand model where $f_{mj}(p_{mj}) = a_{mj} p_{mj}^{b_{mj}}$, and the linear attraction demand model where $f_{mj}(p_{mj}) = a_{mj} + b_{mj} p_{mj}$. Here, a_{mj}, b_{mj} are constants that ensure the negative price elasticity of demand.

A discrete choice function is operationally convenient because of its parsimony in the number of coefficients to be estimated. The number of coefficients in the discrete choice demand model is $O(|M|)$ for purchasing choices in set M (this models $O(|M|^2)$ cross-channel interactions).

The standard methods to estimate discrete choice models require at least some historical information about every choice, which in our setting, would also include the no purchase option (Domencich and McFadden 1975, Berkson 1953). Omnichannel retailers rarely have complete information about lost sales and must calibrate their demand models using incomplete data. We employ an integrated MIP based loss-minimization method proposed by Subramanian and Harsha (2017) to jointly calibrate the market size and choice probability parameters when the lost sales data is censored. Their method performs imputations endogenously in the MIP by estimating optimal values for the

probabilities of the unobserved censored choice. This is a computationally fast single-step method. Compared to its MLE (maximum likelihood estimator) based counterparts like the Expectation Maximization (EM) method (Talluri and Van Ryzin 2004), this method can simultaneously calibrate market-size covariates (e.g., τ_j dependent on temporal causals) besides attraction functionxs, a critical feature with real data. We incorporated model enhancements such as variable selection using LASSO penalties, and sign constraints on coefficients to enable an automated machine learning environment that is required for operational deployment.

In the next two sections, we use discrete choice demand models to analyze certain omnichannel retail price optimization problems. We return to demand forecasting in Section 6.1 where we present the specific censored-data parameter estimation formulation employed to calibrate an MNL omnichannel demand model using historical sales data, incorporating features such as promotions, temporal effects (seasonalities and holidays) and competitor prices besides channel prices.

4. Omnichannel price optimization (OCP) for a non-perishable product

In this section, we formulate the multi-objective OCP model for a non-perishable product in order to identify the most profitable prices in all channels and locations, subject to various retailer's product category goals, channel strategy, sales targets and practical business rules.

We assume that there are well established replenishment policies, and that out-of-stock inventory effects are negligible. This is a reasonable assumption for non-perishable goods (e.g., basic items such as office stationery, printer supplies, etc.). Mathematically, it allows one to view the integrated pricing problem across the retail chain as a single period pricing problem without inventory effects.

Let $\mathbf{c}_j$ denote the selling cost vector at location j across all the channels. Along with the notation introduced earlier in Section 3, we formulate the general non-linear omnichannel price optimization problem denoted by OCP as follows:

$$\text{OCP:} \quad \max_{\mathbf{p}_j} \quad \sum_{j \in J} (\mathbf{p}_j - \mathbf{c}_j)^T \mathbf{D}_j(\mathbf{p}_j) \tag{4.1}$$

$$\sum_j \mathbf{A}_{kj} \mathbf{D}_j(\mathbf{p}_j) \geq u_k \qquad \forall\, k = 1, ..., K \tag{4.2}$$

$$\sum_j \mathbf{B}_{lj} \mathbf{p}_j \leq v_l \qquad \forall\, l = 1, ..., L \tag{4.3}$$

$$p_{m,j} = p_{m,j'} \qquad \forall\, m \in V,\ j, j' \in J \tag{4.4}$$

$$p_{mj} \in \Omega_{mj} \qquad \forall\, m \in M,\ j \in J. \tag{4.5}$$

The decision variables in the above OCP formulation are the prices in all locations and channels, and the objective is to maximize the total profitability of the retailer across the retail chain.

Constraints (4.2–4.3) are generic polyhedral constraints on demands and prices defined with known matrices $\mathbf{A}_k, \mathbf{B}_l \in \mathbb{R}^{|M| \times |J|}$ and vectors $\mathbf{u} \in \mathbb{R}^K, \mathbf{v} \in \mathbb{R}^L$. These generic constraints encapsulate the retailer's goals and critical pricing business rules that are required for operations. We provide examples of these constraints in this section below. Constraint (4.4) ensures the same price across all the virtual stores. This constraint is particularly relevant within our omnichannel framework because we explicitly partitioned the virtual channels by location in order to model cross-channel effects, and this constraint binds them back together from the view of the customer. Discrete pricing constraints, which are typical in retail operations, are encapsulated in constraint (4.5).

Some examples of the generic business rules used in practice are as follows:

Volume (or sales goal) constraints

$$\sum_{m \in M_k, j \in J_k} D_{mj}(\mathbf{p}_j) \geq u_k, \tag{4.6}$$

where $M_k \subset M$ and $J_k \subset J$ and depending on the choice of M_k, J_k these constraints can be employed to support a retailer's global or channel and location-specific sales goals. For example, constraint (4.6) can ensure that the total sales volume by channel does not drop below a user-specified threshold, u_k, thereby balancing profitability and market share objectives. Such constraints act as a practical safeguard to prevent the drastic price increases that can occur while optimizing prices for weakly elastic products (products with elasticity between -1 and 0, typical for basic items). Raising prices for basic products results in a market share erosion over time and is clearly undesirable in the longer term.

Generalized price monotonicity constraints

$$p_{mj} \leq \gamma_{mm'} p_{m'j} + \delta_{mm'j} \quad \forall j \in J \text{ and for some } m, m' \in M. \tag{4.7}$$

Constraints (4.7) enforces business rules related to multi-choice retail pricing. It can be used to ensure that prices in certain channels are cheaper than others by a specified percentage $\gamma_{mm'}$ and/or a constant $\delta_{mm'j}$. This constraint can also account for the variation in unit-cost across channels, i.e., the overhead cost of operating a physical store. An extension of constraint (4.7) is the *price-matching* constraint across the retail chain where the inequality is replaced by an equality and setting $\gamma_{mm'j} = 1$, $\delta_{mm'j} = 0$. Here, consumers can buy the same product anywhere in the retail chain at the same price. Inequalities (4.7) can also be used to impose a *volume measure* constraint, which is also critical in retail pricing. For example, a 24-pack bulk case of white board markers sold online versus a 6-pack case of markers sold in-store. Here, $\gamma_{mm'}$ is a scaling factor between channels that is employed to achieve price parity per unit measure.

Price bounds

$$\underline{\mu}_{mj} \leq p_{mj} \leq \overline{\mu}_{mj} \quad \forall j \in J, \ m \in M. \tag{4.8}$$

12

Harsha, Subramanian and Ettl: *Omnichannel demand modeling and price optimization*

Article submitted to *INFORMS Journal on Optimization*; manuscript no. (Please, provide the manuscript number!)

Here, $\underline{\mu}_{mj}$ and $\overline{\mu}_{mj}$ are upper and lower bounds that are often imposed as a percentage of historically offered prices or as a percentage of competitor prices to ensure the competitiveness of the retailer.

Discrete prices

$$p_{mj} \in \Omega_{mj} \quad \forall j \in J, \ m \in M. \tag{4.9}$$

Ticket prices are naturally discrete (e.g., dollars and cents). Often 'magic number' endings (e.g., ending with '9') are important to a retailer and are encoded as a business rule. Furthermore, constraint (4.9) can be employed to generate a price ladder that proactively excludes trivial in-store price changes to minimize the substantial labor cost incurred in changing the sticker prices.

In general, we classify all location-specific business rules as *inter-channel* constraints and the rest as *inter-location* constraints. Because of our data framework (2.2), inter-channel constraints only have location-specific variables. This contributes to a sparse block-diagonal structure of the multi-location optimization problem.

4.1. Continuous relaxation of OCP and its non-convexity

Consider the OCP problem with the discrete pricing restrictions relaxed. If this continuous relaxation is convex then the solution of a simple gradient-based method yields a true upper bound, which in turn serves as a certificate of the solution quality for the post-facto rounded solution (with minor violations of constraints such as volume goals). We show that the continuous relaxation of the OCP problem is non-convex. Consequently, simple gradient-based rounding schemes cannot offer a reliable quality guarantee.

Market share transformations are commonly used for discrete choice models to achieve convexity in the continuous pricing problem (see Section 1.1 for related references). The market share variables are defined as follows for each $j \in J$:

$$\theta_{mj} = \frac{f_{mj}(p_{mj})}{1 + \sum_{m' \in M} f_{m'j}(p_{m'j})} \qquad \forall m \in M, \text{ and} \tag{4.10}$$

$$\bar{\theta}_j = 1 - \sum_m \theta_{mj}, \tag{4.11}$$

with a one-to-one transformation to the price variables, given by $p_{mj} = f_{mj}^{-1}\left(\frac{\theta_{mj}}{\bar{\theta}_j}\right)$. In the following proposition, we show that the market share transformations are not sufficient for proving convexity of the continuous pricing OCP problem discussed above.

PROPOSITION 1. *Under the market share transformations, the resultant price monotonicity (or a volume measure) constraints (4.7) are non-linear and non-convex.*

We provide an example to prove the above proposition in Appendix A.

Furthermore, even in the absence of these constraints, the objective function of the continuous multi-location OCP problem is not unimodel in the price space and convex in the market-share

space (unlike the single-location multi-choice setting, see Section 1.1). In Fig. 2 we plot the values
of the objective function Eq. (4.1) for an OCP instance having a single virtual channel, say online,
and two locations, with constraint (4.4) ensuring that the online price across locations is the same.
We observe from the figure that the objective function in this example is non-convex and has
multiple peaks. Constraint (4.4) is similar to the price monotonicity constraint (4.7), except that
it is across locations and not choices (see Proposition 1).
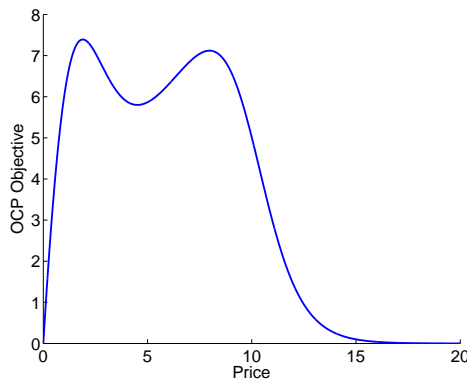
Figure 2 **Example of a OCP objective for a single virtual channel and two locations as a function of the virtual**
channel price for an MNL demand model with $a_{11} = 10, a_{12} = 1, b_{11} = 1, b_{12} = 1, \tau_1 = 1$ **and** $\tau_2 = 10$**.**

Recognizing the apparent non-convexity and non-linearity of the OCP problem in the price and
market share space, a relevant question is if there are alternative transformations that recover a
convex continuous relaxation of the OCP problem having a limited number of channels (e.g., with
one virtual channel and one physical channel). While this remains an open question, we prove the
following proposition in Appendix B.

PROPOSITION 2. *OCP with multiple virtual channels and at least two locations is NP-hard for*
general attraction functions that satisfy the following mild regularity condition:

$$\lim_{p_m \to \infty} p_m \frac{f_{mj}(p_m)}{1 + f_{mj}(p_m)} = 0 \quad \forall m \in V, j \in J. \tag{4.12}$$

4.2. An exact mixed-integer programming approach for OCP

We present an empirically tractable global optimization approach using mixed-integer program-
ming (MIP) for the OCP problem. Unlike the discrete choice optimization models in the literature,
the proposed method gainfully operates both in the price and the market share space to address
the non-linearity concerns discussed above.

Let the set I_{mj} denote the index set of the feasible discrete prices for channel $m \in M$ and location
$j \in J$. We denote the corresponding prices by $\bar{p}_{mji}$ for $i \in I_{mj}, m \in M, j \in J$. Let z_{mji} be a binary
variable which is nonzero only if the price in channel $m \in M$ at location $j \in J$ is $\bar{p}_{mji}$. Note that for

a virtual channel $v \in V$ the prices across all locations are the same. Therefore, the corresponding prices $\bar{p}_{vi}$, the price index set I_v and binary variable z_{vi} are location independent, and $\bar{p}_{vji} = \bar{p}_{vi}$, $I_{vj} = I_v$ and $z_{vji} = z_{vi}$.

Using these definitions, the key term to linearize in the OCP problem is the demand function $D_{mj}(\mathbf{p}_j)$ which depends on prices in all the channels. A naive way of linearizing this function requires the introduction of $|I|^{|M|}$ additional binary variables when $I_{mj} = I \; \forall m \in M$ (note that there are already $|I||M|$ binary variables due to discrete prices). This results in a MIP that explodes in size very quickly making it impractical to solve. We present an exact alternative linearization that exploits the special structure of the discrete choice model and does not require any additional binary variables, resulting in a computational tractable MIP.

PROPOSITION 3. *Let* $q_{mji} = \tau_j(\bar{p}_{mji} - c_{mj})f_{mj}(\bar{p}_{mji})$, $r_{mji} = f_{mj}(\bar{p}_{mji})$, $\alpha_{kmji} = A_{kmj}\tau_j r_{mji}$ *and* $\beta_{kmji} = B_{lmj}\bar{p}_{mji}$. *Then, the OCP problem can be reformulated, and is equivalent, to the following MIP formulation.*

$$\max_{z,y,x} \sum_{j \in J} \sum_{m \in M} \sum_{i \in I_{mj}} q_{mji} x_{mji} \tag{4.13}$$

$$\sum_{j \in J} \sum_{m \in M} \sum_{i \in I_{mj}} \alpha_{kmji} x_{mji} \geq u_k \qquad \forall \, k \in K \tag{4.14}$$

$$\sum_{j \in J} \sum_{m \in M} \sum_{i \in I_{mj}} \beta_{lmji} z_{mji} \leq v_l \qquad \forall \, l \in L \tag{4.15}$$

$$y_j + \sum_{m \in M} \sum_{i \in I_{mj}} r_{mji} x_{mji} = 1 \qquad \forall \, j \in J \tag{4.16}$$

$$x_{mji} \leq z_{mji} \qquad \forall \, i \in I_{mj}, m \in M, \; j \in J \tag{4.17}$$

$$\sum_{i \in I_{mj}} x_{mji} = y_j \qquad \forall \, m \in M, \; j \in J \tag{4.18}$$

$$\sum_{i \in I_{mj}} z_{mji} = 1 \qquad \forall \, j \in J, m \in M \tag{4.19}$$

$$z_{vi} = z_{vji} \qquad \forall \, j \in J, v \in V \tag{4.20}$$

$$z_{mji} \in \{0,1\} \qquad \forall \, i \in I_{mj}, m \in M, j \in J \tag{4.21}$$

$$y_j, x_{mji} \geq 0 \qquad \forall \, i \in I_{mj}, m \in M, \; j \in J \tag{4.22}$$

Proof: In the above MIP formulation, constraints (4.19) and (4.21) model the discrete nature of channel prices as in (4.5) using the binary variables z_{mji}. The general business rules on prices and the uniform virtual price constraint of the OCP problem given by constraints (4.3–4.4) can be re-written with the new notation as constraints (4.15) and (4.20) respectively. The objective and constraint (4.2) of the OCP problem can be rewritten as

$$\max_{z_{mji}, z_{vi}} \quad \sum_{j \in J} \frac{\sum_{m \in M} \sum_{i \in I_{mj}} q_{mji} z_{mji}}{1 + \sum_{m \in M} \sum_{i \in I_{mj}} r_{mji} z_{mji}}, \text{and} \tag{4.23}$$

$$\sum_{j \in J} \sum_{m \in M} \sum_{i \in I_{mj}} \frac{\alpha_{kmji} z_{mji}}{1 + \sum_{m \in M} \sum_{i \in I_{mj}} r_{mji} z_{mji}} \geq u_k \qquad \forall\, k \in K. \qquad (4.24)$$

We use the fractional programming transformations proposed by Charnes and Cooper (1962) to overcome the non-linearity arising from the ratio terms. Let

$$y_j = \frac{1}{1 + \sum_{m \in M} \sum_{i \in I_{mj}} r_{mji} z_{mji}} \quad \forall\, j \in J. \qquad (4.25)$$

Because z_{mji} are binary variables and r_{mji} are non-negative constants, $0 \leq y_j \leq 1 \ \forall j \in J$. We define

$$x_{mji} = y_j z_{mji}. \qquad (4.26)$$

A direct outcome of this transformation is that the objective function (4.1) and constraints (4.2) can be re-written as (4.13) and (4.24) respectively, wherein Eq. (4.25) is expressed as (4.16).

Next, we use the reformulation and linearization technique (RLT) proposed by Sherali and Adams (1999) to eliminate the non-linearities in Eq. (4.26) and linearize it exactly (in the integer sense) using constraints (4.17–4.18). First, it is easy to see that $0 \leq x_{mji} \leq 1 \ \forall m \in M, i \in I_{mj}$. Next, if $z_{mji} = 0$, constraint (4.17), along with $x_{mji} \geq 0$ ensures $x_{mji} = 0$, and if $z_{mji} = 1$, constraint (4.18) together with constraint (4.19) ensure $x_{mji} = y_j$, which is feasible to constraint (4.17). $\qquad \square$

Observe that the transformed OCP formulation can incorporate a variety of important and complex business rules that are employed in practice. The above formulation is a linear MIP and a commercial optimization software package can be used to solve this problem to optimality.

In our numerical computations, we observed that the addition of RLT constraints (4.18), which are known for providing tighter relaxation structures (Sherali and Adams 1999), yielded a considerable improvement in the computational performance over alternative methods. For example, a commonly used approach is to use constraints $y_j - x_{mji} \leq 1 - z_{mji}$ and $x_{mji} \leq y_j \quad \forall i \in I_{mj}, m \in M, j \in J$ instead of the RLT constraint (4.18) that can be gainfully used in the pricing context for linearization (see for example Bront et al. 2009's approach for an assortment problem.). This results in $2|I_{mj}| - 1$ more constraints for each $m \in M$ and $j \in J$. A simulation using synthetic data over 100 instances having uniformly random chosen intercept and price coefficients for an MNL demand model with 2 channels, 20 locations and 20 price levels uniformly discretized between 60 and 140, a global volume constraint and default CPLEX solver settings results in a run time reduction from 12.7 seconds to 0.22 seconds using the RLT constraint, a factor of 57 in improvement.

In Fig. 3, we plot the average run time of the proposed MIP (with 1% optimality tolerance) as function of the number of locations for different combinations of virtual and physical channels. We observe the runtime increase is non-linear in the number of locations and in the number channels. The largest MIP instance (2 channels, 512 locations) required more than 10,000 binary variables. It takes less than 5 seconds to solve the MIP when the number of locations is as large as 512, 128 and 16 in the following three cases respectively: (1) single virtual channel, (2) single virtual

with one physical channel and (3) 4 virtual with one physical channel, while the toughest problem instance (5 channels, 128 locations) solves within 21 minutes on average (10 seconds per location). In Section 6 we report that solving practical OCP instances take no more than 3 seconds. These large instances remain tractable due to the sparse nature of constraints (4.16–4.19) (see Crowder et al. 1983 for reasons on MIP sparsity aiding tractability). For larger instances, with many channels and locations, decomposition methods that exploit the block diagonal structure of the problem maybe viable.
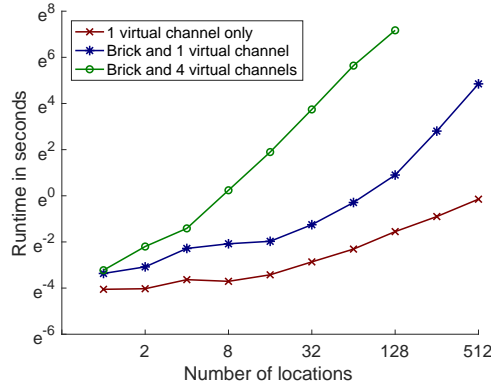

Figure 3   Average run times of the OCP MIP over 30 simulated instances.

5. Omnichannel multi-product pricing problem

This far we analyzed omnichannel models in the setting of a single non-perishable product. Retailers usually offer product assortments consisting of differentiated non-perishable product alternatives. In this case, the consumer choices are the substitutable channel-product pairs. This setting can be modeled using the same demand model described in Eq. (3.1), by extending the set M to include all channel-product pairs. However, this choice model suffers from the independence from irrelevant alternatives (IIA) property. The implied proportional substitution across alternatives (Train 2009) can lead to unrealistic demand predictions in certain settings. In this section, we consider the nested attraction demand model, which extends the choice model in Eq. (3.1) and overcomes the IIA limitation.

The nested model is a hierarchical discrete choice model where the assumption is that consumers first select a nest and then the choices within the nest. Suppose M denotes the set of nests and N_m the choices in each nest, the demand for a choice n in a nest m has the form:

$$D_m^n(\mathbf{p}) = \tau \; \frac{U_m^{\gamma_m}}{1 + \sum_{m' \in M} U_{m'}^{\gamma_{m'}}} \; \frac{f_m^n(p_m^n)}{\sum_{n' \in N_m} f_m^{n'}(p_m^{n'})}, \tag{5.1}$$
$$\text{where } U_m = \sum_{n' \in N_m} f_m^{n'}\left(p_m^{n'}\right), \; n \in N_m, \; m \in M.$$

Here U_m is referred to as the intrinsic value of a nest and is the total attractiveness of a nest, while the parameter $\gamma_m \in (0,1]$ is a nest coefficient that measures the degree of inter-nest heterogeneity.

When $\gamma_m = 1$, the nested model reduces to the choice model in Eq. (3.1) and therefore here we focus on the case when $\gamma_m < 1$ that is relevant to our omnichannel setting.

The groups of channel-product pairs that form nests depend on channel attributes (e.g. ease of purchase, user experience, product variety) as well as product attributes (e.g., brand, quality, etc.). To fix ideas in the paper, we let the upper-level nests M correspond to channels and the no-purchase option, and N_m correspond to the differentiated products within the channel assortment (for example, the 'endless aisle' in the online channel and the space-limited assortments in stores). This nested model structure allows us to quantify cross-product demand interactions using the purchase probability of the products within each channel (lower level), as well as the cross-channel effects using channel purchase probability of demand (upper level), wherein the channel attractiveness depends on the total attractiveness of the channel-specific assortment.

To calibrate uncensored nested attraction models, we refer the reader to Train (2009) for (1) a joint upper and lower level MLE estimator (preferred) and (2) two-step bottom-up estimator that exploits the fact that the choice probabilities can be decomposed into marginal and conditional probabilities, resulting in an estimator that is consistent but not efficient. We are not aware of any estimation method for the censored nested attraction model. We can adapt and extend the two-step bottom-up method to the censored data setting as follows. First the parameters for each nest at the lower level are independently estimated from historical sales data using an MLE method (as in the uncensored case). Next using the resultant estimated intrinsic values for all the nests, the upper level parameters are estimated using the procedure described and detailed in Sections 3 and 6.1 respectively (in lieu of MLE in uncensored case), where the inclusive value terms are treated as explanatory variables and the no-purchase choice is censored.

5.1. A mixed-integer programming approach for the nested attraction model

The canonical formulation of the OCP problem using the nested attraction demand model is similar to the single item case presented in Section 4 except that the decision variables now include prices of all the products. Additional important business constraints such as volume goals at the product and channel level, as well as inter-product price constraints due to brand and volume differences (which have a form similar to constraint 4.7) that are equally relevant in the multi-product setting. For brevity, we only present the linear transformations of the objective function for the single location case. Incorporating all the business rules required for operations, and an extension to the multi-location case are relatively simple after this transformation, and we follow the steps in Section 4.2 with appropriate changes.

Let the set I_{mn} denote the index set of feasible discrete prices for choice $n \in N_m$ within nest $m \in M$. We denote corresponding feasible prices by $\bar{p}_{mni}$ for $i \in I_{mn}, n \in N_m, m \in M$. Let z_{mni}

be a binary variable which is nonzero only if the price in nest $m \in M$ and choice $n \in N_m$ is $\bar{p}_{mni}$. Assuming $q_{mni} = \tau(\bar{p}_{mni} - c_{mn})f_{mn}(\bar{p}_{mni})$ and $r_{mni} = f_{mn}(\bar{p}_{mni})$, the single location objective function of the OCP problem with the nested model can be written as:

$$\text{OCPN:} \quad \max_{z \in \{0,1\}} \quad \sum_{m \in M} \frac{U_m^{\gamma_m}}{1 + \sum_{m \in M} U_m^{\gamma_m}} \frac{\sum_{n \in N_m} \sum_{i \in I_{mn}} q_{mni} z_{mni}}{\sum_{n \in N_m} \sum_{i \in I_{mn}} r_{mni} z_{mni}} \tag{5.2}$$

$$\text{where } U_m = \sum_{n \in N_m} \sum_{i \in I_{mn}} r_{mni} z_{mni} \tag{5.3}$$

Linearizing OCPN is especially challenging given the two-levels of variables (channel-level and product-level). We employ multiple ideas to derive two different global optimal approximations to this difficult problem that we allude to in this paragraph and detail further below with propositions. First, we introduce channel and product level market share variables and use RLT to partially linearize Eq. (5.2). Second, we define η_m as the normalized intrinsic value of nest m between $(0,1]$ (recall that U_m refers to the intrinsic value). We discretize this quantity and refer to the discrete samples as knots. Third, using Special-Ordered-Set Type 2 (SOS2) variables we rewrite η_m exactly as a convex combination of sampled knots in $[0,1]$ and 'transfer' this value from the lower level (U_m) to the upper level model $(U_m^{\gamma_m})$. The resultant linearized problem with these three ideas approximates OCPN and is asymptotically optimal as the number of knots used increases (see Proposition 4 below).

We enhance this formulation further with the knowledge that the upper level intrinsic value $(U_m^{\gamma_m})$ is a concave function and therefore a piecewise linear approximation provides a lower bound, while the tangents to the curve yield an upper bound. We show that the resultant problem, with these cuts in lieu of a value transfer, is a relaxation of OCPN problem that is also asymptotically optimal as the number of knots used increases (see Proposition 5 below). The benefit of this formulation (despite having more constraints) is that it yields an upper bound to the original OCPN's optimal objective function value, besides performing better computationally. We now provide the reformulations and the propositions.

PROPOSITION 4. *OCPN-PWL approximates OCPN using a piecewise linear approximation of the normalized intrinsic value and is asymptotically exact as the number of knots increases.*

$$\text{OCPN-PWL:} \quad \max_{z \in \{0,1\}, x, y, \theta} \quad \sum_{m \in M} \sum_{n \in N_m} \sum_{i \in I_{mn}} q_{mni} x_{mni} \tag{5.4}$$

$$\sum_m \theta_m + \bar{\theta} = 1 \tag{5.5}$$

$$\sum_{n \in N_m} \sum_{i \in I_{mn}} r_{mni} x_{mni} = \theta_m \qquad \forall m \in M \tag{5.6}$$

$$Y_m^{\min} \le y_m \le Y_m^{\max} \qquad \forall m \in M \tag{5.7}$$

$$\sum_{i \in I_{mn}} x_{mni} = y_m \qquad\qquad \forall n \in N_m, m \in M \tag{5.8}$$

$$Y_m^{\min} z_{mni} \le x_{mni} \le Y_m^{\max} z_{mni} \qquad\qquad \forall i \in I_{mn}, n \in N_m, m \in M \tag{5.9}$$

$$\theta_m, \bar{\theta}, x_{mni} \ge 0 \qquad\qquad \forall i \in I_{mn}, n \in N_m, m \in M \tag{5.10}$$

$$\sum_{n \in N_m} \sum_{i \in I_{mn}} r_{mni} z_{mni} = \sum_{s \in S} \bar{U}_{ms} w_{ms} \qquad \forall m \in M \tag{5.11}$$

$$\sum_{s \in S} \bar{R}_{ms} \rho_{ms} = \theta_m \qquad\qquad \forall m \in M \tag{5.12}$$

$$\sum_{s \in S} \rho_{ms} = \bar{\theta} \qquad\qquad \forall m \in M \tag{5.13}$$

$$\rho_{ms} \le w_{ms} \qquad\qquad \forall m \in M, s \in S \tag{5.14}$$

$$\{w_{ms} \forall s \in S\} \in SOS2 \qquad\qquad \forall m \in M. \tag{5.15}$$

where

- $Y_m^{\min} = \frac{(U_m^{\min})^{\gamma_m - 1}}{1 + \sum_{m'} (U_{m'}^{\max})^{\gamma_{m'}}}$ *and* $Y_m^{\max} = \frac{(U_m^{\max})^{\gamma_m - 1}}{1 + \sum_{m'} (U_{m'}^{\min})^{\gamma_{m'}}}$ *where* $U_m^{\max}$ *and* $U_m^{\min}$ *are the maximum and minimum intrinsic values for a nest obtained by setting the discrete prices at the minimum and maximum values respectively (which can be further strengthened by considering intra nest constraints, if any);*

- $\bar{\eta}_{ms} \, \forall s \in S$ *are the chosen knots that discretize interval* $\left[\frac{U_m^{\min}}{U_m^{\max}}, 1\right]$ *for each* $m \in M$;

- $\bar{U}_{ms} = U_m^{\max} \bar{\eta}_{ms}$, *and* $\bar{R}_{ms} = (U_m^{\max} \bar{\eta}_{ms})^{\gamma_m}$;

- *the auxiliary decision variables* $w_{ms} \forall s \in S$ *are modeled as SOS2 variables wherein at most two adjacent members (assuming the* $\bar{\eta}_{ms}$*'s are ordered) can be non-zero. We do not expand on the form of the SOS2 constraints as all standard optimization packages allows this level of specification and manage these variables internally.*

We describe the transformations used to derive OCPN-PWL and refer the reader to Appendix C for the proof on asymptotic optimality. We first introduce the market share transformations both in the upper and the lower level. Let

$$\frac{U_m^{\gamma_m}}{1 + \sum_{m' \in M} U_{m'}^{\gamma_{m'}}} = \theta_m, \quad \frac{1}{1 + \sum_{m' \in M} U_{m'}^{\gamma_{m'}}} = \bar{\theta}, \quad \frac{\theta_m}{\sum_{n \in N_m} \sum_{i \in I_{mn}} r_{mni} z_{mni}} = y_m \text{ and } y_m z_{mni} = x_{mni}.$$

We linearize the bilinear term exactly using RLT constraints because y_m is bounded, in particular, using the definitions of $U_m^{\max}, U_m^{\min}, Y_m^{\max}, Y_m^{\min}$, we know

$$\frac{(U_m^{\min})^{\gamma_m - 1}}{1 + \sum_{m'} (U_{m'}^{\max})^{\gamma_{m'}}} = Y_m^{\min} \le y_m \le Y_m^{\max} = \frac{(U_m^{\max})^{\gamma_m - 1}}{1 + \sum_{m'} (U_{m'}^{\min})^{\gamma_{m'}}}. \tag{5.16}$$

With these definitions and transformations, (5.2) can be linearized exactly to objective (5.4) and constraints (5.5–5.10).

We now linearize $U_m^{\gamma_m} = \frac{\theta_m}{\bar{\theta}}$ approximately using constraint (5.3). We refer to Fig. 4 for the following discussion. We introduce a piecewise linear mapping of the concave function η^{γ_m} for
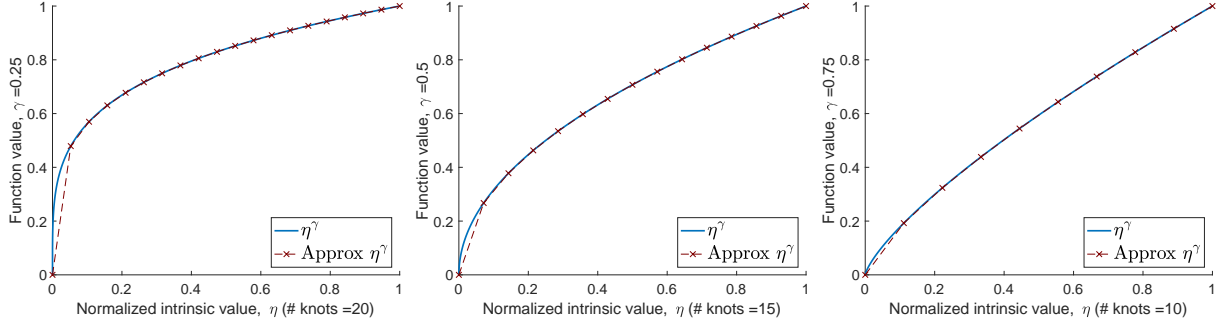
Figure 4   **True and approximate value of the projected function η^γ, where $\eta \in [0,1]$ is the normalized nest attractiveness value, for varying number of uniformly spaced knots.**

$\eta \in [0,1]$ using a normalized nest intrinsic value η, which we define as the actual nest intrinsic value, U_m divided by $U_m^{\max}$. For any nest m, $\eta^{\gamma_m} \in \left[\frac{U_m^{\min}}{U_m^{\max}}, 1\right]$. With $\bar{\eta}_{ms} \ \forall s \in S$ being the chosen knots, and $\bar{U}_{ms} = U_m^{\max} \bar{\eta}_{ms}$, the lower level piecewise linear mapping given by Eq. (5.3) can be exactly written as Eq. (5.11) because $\eta_m = \eta_{ms} w_{ms}$ and the SOS2 nature of the variables w_{ms} with (5.15). For the upper level mapping, with $\bar{R}_{ms} = (U_m^{\max} \bar{\eta}_{ms})^{\gamma_m}$, $U_m^{\gamma_m}$ is approximated as

$$U_m^{\gamma_m} \sim \sum_{s \in S} \bar{R}_{ms} w_{ms}. \tag{5.17}$$

Substituting $U_m^{\gamma_m} = \frac{\theta_m}{\bar{\theta}}$, we get, bilinear variables of the form $w_{ms}\bar{\theta}$. We now re-apply RLT to linearize this term (denoted by ρ_{ms}) to obtain constraints (5.12–5.14). Note that this linearization does not preserve exactness as the auxiliary w_{ms} variables are SOS2 variables (and not binary), but constraint (5.12) ensures U_m lies on the same piecewise-linear segment. Exactness is not preserved using a finite number of knots, but it is achieved asymptotically as the number of knots increase (see detail of proof in Appendix C).

We use Fig. 4 to depict the piecewise linear approximation for different numbers of knots employed and different γ values. Observe that the approximation is relatively poor when η is close to 0, although η is lower bounded by $\frac{U_m^{\min}}{U_m^{\max}} > 0$. For basic (weakly elastic) items, we observe this lower bound is higher compared to highly elastic products. The approximation improves for larger values of γ and by increasing the number of knots.

COROLLARY 1. *Suppose there exist a set of polyhedral price constraints in OCPN for which the number of feasible price vectors is polynomial in the size of the choice set N_m and the index set I_{mn}. Then the intrinsic values corresponding to these price vectors can be enumerated and knots can be located at every such normalized intrinsic value. In this special case, it suffices that $w_{ms} \ \forall s \in S$ are modeled as binary variables in order for OCPN-PWL to be an exact transformation of OCPN.*

We state the above corollary without proof. It is clear that by changing w_{ms} to binary variables, the aforementioned RLT transformations for ρ_{ms} become exact.

$(|M|,|N|,$ $|J|)$	Tangent Approx	1% Target Opt Tolerance			5% Target Opt Tolerance			10% Target Opt Tolerance		
		Run Time avg. (sec)	Realized Opt Gap (%)	# Success Instances	Run Time avg. (sec)	Realized Opt Gap (%)	# Success Instances	Run Time avg. (sec)	Realized Opt Gap (%)	# Success Instances
(2, 5, 1)	No	1.1	1.1	22	0.3	2.2	30	0.2	2.5	30
(2, 10, 1)	No	3.9	1.4	11	0.5	2.4	30	0.4	2.7	30
(2, 20, 1)	No	13	1.6	11	1.0	2.4	30	1.0	3.0	30
(2, 30, 1)	No	26	1.5	11	1.8	2.5	30	1.6	3.8	30
(3, 30, 1)	No	58	1.6	6	4.4	2.7	30	4.4	2.7	30
(4, 30, 1)	No	103	1.8	5	18	2.7	30	17	3.4	30
(5, 30, 1)	No	115	1.9	2	24	2.6	30	23	3.2	30
(2, 30, 1)	Yes	13	1.5	10	1.7	2.9	30	1.5	3.6	30
(5, 30, 1)	Yes	81	3.1	0	17	4.1	29	13	4.8	30
(2, 30, 5)	Yes	-	-	-	143	4.6	29	56	8.0	30
(2, 30, 10)	Yes	-	-	-	521	4.6	26	262	5.0	30
(2, 30, 20)	Yes	-	-	-	2447	4.9	22	890	5.5	30

Table 1   **Computational results for 30 instances of OCPN with discrete prices and a global volume goal.**

PROPOSITION 5. *Define OCPN-Tangent as OCPN-PWL with constraint (5.12) replaced by constraints (5.18) given below:*

$$\sum_{s \in S} \bar{R}_{ms} \rho_{ms} \leq \theta_m \leq \bar{R}_m^t \left[(1 - \gamma_m)\bar{\theta} + \gamma_m \sum_{s \in S} \frac{\bar{\eta}_{ms}}{\bar{\eta}_m^t} \rho_{ms} \right] \qquad \forall t \in T, \qquad (5.18)$$

where the tangent to concave normalized intrinsic value curve $\eta_m^{\gamma_m}$ is evaluated at $\bar{\eta}_m^t \ \forall t \in T$ and $\bar{R}_m^t = (U_m^{\max} \bar{\eta}_m^t)^{\gamma_m}$. OCPN-Tangent is a relaxation of OCPN using piecewise linear and tangent bounds and is asymptotically exact as the number of knots increases.

OCPN-Tangent provide an alternative way of linearizing $\eta_m^{\gamma_m}$ such that the optimal objective of OCPN-Tangent is a guaranteed upper bound to OCPN. The intuition to constraint (5.18) is as follows. Observe that the piecewise-linear mapping curve is a lower bound to the true function $(U_m^{\max} \eta_m)^{\gamma_m}$, which is concave. Any tangent is any upper bound to this function, and hence $\theta_m \leq \bar{R}_{mt} \bar{\theta} \left[1 - \gamma_m + \frac{\gamma_m \eta_m}{\bar{\eta}_m^t} \right]$ where the tangent is evaluated at $\bar{\eta}_m^t$. We know $\eta_m = \sum_{s \in S} \bar{\eta}_{ms} w_s$. Substituting and summarizing, we get, the constraint. The proof of the proposition is in Appendix D.

This single location OCPN formulation can be directly extended to the multi-location setting by simply summing the objective and constraint contributions across locations. The corresponding multi-location solution of both the OCPN-PWL and the OCPN-Tangent would approximately satisfy the volume goals that are present.

5.2. Computational performance of the proposed methods for the OCPN problem

Table 1 provides average run time results for the single and multi-location constrained OCPN problem over 30 random instances with model parameters motivated from real data. We vary the number of channels $|M|$, the products per channel $|N|$ and the number of locations $|J|$. The γ_m values were randomly chosen between (0.1,1.0) and the price coefficients between (-2.0, -0.1). We employed 20 discrete prices uniformly distributed within +/- 50% of the average value. We applied a global volume goal to preserve the total sales volume that is calculated at the average product prices. We did not implement any inter-product price constraints that reduced the feasible space

and improved run times. We used only simple bound-preprocessing ideas to obtain and tighten the bounds for $U_m^{\min}$ and $U_m^{\max}$ for this analysis. We used CPLEX 12.6.2 as a solver with a pre-specified target optimality tolerance that was varied, while the bound node limit was set to 100K.

First, we tested the OCPN-PWL formulation (in Table 1, Tangent Approx = No). Here we observe that all instances solve successfully at 10% and 5% optimality tolerance[3], while only some instances solve successfully at 1% tolerance, suggesting that the proposed MIP can achieve near optimal solutions (within 2-5% of optimality) for the single-location instances within 30 seconds of run time. Upon computing the maximum approximation error in the true (non-linear) objective function value and OCPN-PWL model evaluated objective function value, and the corresponding global volume values across all instances (in Table 1, Tangent Approx = No), we find this error to be less than 0.1% (not in table), suggesting that the quality of the SOS2 approximation using 20 knots is adequate.

To evaluate the impact of adding tangential cuts (5.18), we re-solved some of the instances using 10 tangents per nest and the OCPN-Tangent formulation (in Table 1, Tangent Approx = Yes). On average, the MIP converged about 30% faster compared to the case without tangents. Furthermore, it also yields an upper bound on the true optimum (best upper bound achieved by the MIP solver). We therefore employed the tangents to tackle the harder multi-location instances. We could solve all multi-location instances at the 10% optimality setting. Between 73-97% of the instances were solved at a 5% tolerance setting and therefore we did not test the 1% tolerance setting. The largest problem setting (20, 2, 30) was solved to within 5.5% of MIP optimality under 15 minutes on average (or 30 seconds per product). A lower bound for each instance was calculated by evaluating the true non-linear objective of OCPN at the optimal MIP prices. This yielded an average global optimality gap of 10.8%, where the average violation of the global volume goal was less than 0.6%.

6. OCP implementation for a major US retailer

In this section, we report results of an OCP implementation for a major U.S. omnichannel retailer. We worked with IBM Commerce and engaged with the retailer over the course of 8 months to demonstrate the business value of the integrated omnichannel regular pricing over their existing channel independent pricing on representative product categories, based on historical data. The details of the different steps are provided below but first, we describe the retailer and their business process (while maintaining their anonymity).

Retailer and their current business process: The retailer sells a variety of products, including office supply product categories. They operate a network of over 1500 stores across the United

---

[3] This means the realized optimality gap is less than the target optimality gap. In general, this may not happen with other stopping criteria like maximum node limit, which we also imposed.

States. The online channel is used to complete sales transactions that are routed through their website, as well as mobile and paper-catalog orders. The organizational structure of the retailer results in two different divisions separately managing the planning and operations of the two channels. Both divisions use a regular price optimization (RPO) solution to manage prices for many non-perishable products, referred to as UPCs (universal product code). The prices for the remaining products are controlled partly by the manufacturer or are price-matched with certain competitors. The incumbent RPO solution first produces weekly demand forecasts using models that are calibrated with the latest available weekly sales and promotion data and are independent of the other channels' or competitors' data. Next it identifies regular or baseline price for the (non-perishable) products that maximizes the retailer's profitability over a specified finite horizon (often multiple weeks at a time) subject to some business constraints. Small price changes (e.g., less than 30%) are typical. One price is recommended for every geographical cluster of brick stores identified by the retailer as a 'price zone'. The entire online channel is treated as a separate price zone. These prices can be overlaid with various promotions.

Business Problem: More than 20% of UPCs that are priced by the retailer are sold in both channels. They contribute to a significant portion of the retailer's category revenue (details provided below). Due to the rapid growth of the online channel, the retailer was primarily concerned about how to profitably coordinate prices between the two channels while accounting for competitor effects and preserving market share. For the business value assessment, they suggested we ignore cross-product effects (since it was negligible for the UPCs provided in the historical data) and that we do not enforce a price match between the brick and online channels.

Data Summary: To support the business value assessment, we were provided with 52 weeks of U.S. sales transaction and promotion data from July 2012 for the brick and online channels for two categories: (1) inkjet cartridges and (2) markers and highlighters. The top 50 UPCs in terms of historical volume that were sold in both the channels (channel volume share of at least 1%) were selected for the business value assessment. Some statistics about the data are summarized in Table 6.

Category	No. of UPCs	Avg. Final Price Brick	Online	% of category revenue	Online volume share
Inkjet Cartridges	50	$36.4	$32.1	30%	12%
Markers and Highlighters	50	$8.7	$8.3	42%	12%

Table 2   Summary of the 2012-13 data provided.

In the inkjet cartridges and the markers and highlighters category, the 50 UPCs that were selected contributed to about 30%, and 42% of the category revenues respectively, and have a 12%

online volume share in each case. Although the historical online share of the retailer sales in 2012-13 was relatively low, we note that this number has been steadily increasing every year. The inkjet cartridges category consists of products that are more expensive than the markers and highlighters. The retailer sells products in the brick channel at a slightly higher price than the online channel, despite a vast majority of customers buying in-store. Although one would attribute it to the higher holding costs in-store, this information was not provided to us. We were only provided with the wholesale cost values for the different UPCs and these were the same across the sales channels.

We implemented a scalable data method to extract, transform and load, high volumes of transactional data streams. Forty geographical price zones were created based on this data using the omnichannel framework described in Section 2. The total sales rates across the 40 zones were not evenly distributed across locations, and we found that the top 10 zones contributed to 54% of the total sales, and the top 20 zones accounted for 83% of the total sales.

For each of the UPC-zone pairs, we obtained by channel, the weekly aggregated sales, volume weighted weekly average ticket price, discounts and promotions, and weekly holidays and seasonality factors. For a subset of UPCs, we were also provided a time series of three online competitor's prices in order to assess the competitor price impact on profitability and sales. We observed that the products exhibited a relatively steady sales rate, which is typical of non-perishable basic products. The retailer varied their store and online prices over time, with relatively more frequent online price changes. Their store prices varied across locations and differed from the online price. These pricing variations enabled us to estimate own- and cross-channel price elasticity values from the sales data using our data framework.

6.1. Demand Estimation using Machine Learning

We use the zone-tagged historical data to calibrate the omnichannel demand model described in Eq. (3.1) based on the MNL attraction function at the UPC-zone level. Model selection and cross-validation on a variety of training instances yielded: (1) the following exponential market-size model $\tau_t(\boldsymbol{\gamma})$ for a zone that predicts the customer arrivals for any week t in the selling season as a function of coefficients $\boldsymbol{\gamma}$:

$$\ln\left(\tau_t(\boldsymbol{\gamma})\right) = \gamma_0 + \sum_q \gamma_{1,q}\text{TEMPORAL-VARIABLES}_{q,t}, \tag{6.1}$$

and (2) the following channel attraction model $f_{m,t}(\boldsymbol{\beta}_m)$ to predict channel purchase probabilities in week t as a function of coefficients $\boldsymbol{\beta}_m$:

$$\ln\left(f_{m,t}(\boldsymbol{\beta}_m)\right) = \beta_{m0} + \beta_{m1}\text{PRICE}_{m,t} + \sum_r \beta_{m2,r}\text{PROMOTION-VARIABLES}_{m,r,t} \tag{6.2}$$

$$+ \sum_w \beta_{m3,w}\text{COMPETITOR PRICES (optional)}_{m,w,t}.$$

The promotion variables included discounts and other promotional indicators and the temporal variables included seasonality, holiday effects and trend. Competitor market share data were

unavailable and competitor prices were introduced as attributes in the channel-specific utilities to approximately account for their impact on location and channel-specific demand.

Because lost sales data was unavailable in the training data, the weekly market-sizes are also unknown. Given this the γ and $\boldsymbol{\beta}_m$ coefficients in Eq. (6.1) and Eq. (6.2) are jointly estimated using an L1 loss-minimization model, which is an omnichannel adaptation of the censored data estimation model proposed in Subramanian and Harsha (2017) along with user-acceptance constraints. In this approach, the censored lost share value for any time period is treated as an auxiliary decision variable that is bounded between 0 and 1. The resultant nonconvexity and nonlinearity in the estimation problem is effectively managed using a piecewise-linear (PWL) approximation and yields a tractable linear MIP formulation. An MNL demand model calibrated in this manner can predict the weekly market sizes as well as channel purchase probabilities required by our pricing models. We now describe the notation used, the data inputs, and key decision variables, and then discuss the L1 loss-minimization model formulation.

The user specifies a discrete sampling of possible lost share values/knots in $[\epsilon, 1-\epsilon]$, denoted by $\phi_k, \forall k \in K$, where ϵ is a small positive number (e.g., 10^{-3}). In period t, the observed sales for choice m is denoted by $\bar{s}_{mt}$ and supposing that the lost share is ϕ_k, the corresponding lost sales for period t due to no-purchase is obtained as $\bar{\zeta}_{kt} = \frac{\phi_k}{1-\phi_k} \sum_{m \in S_t} \bar{s}_{mt}$, and the market size is $\bar{\eta}_{kt} = \frac{1}{1-\phi_k} \sum_{m \in S_t} \bar{s}_{mt}$. We introduce auxiliary SOS2 decision variables z_{kt} to denote the probability that the lost share in period t is ϕ_k. Recall that this means at most two adjacent members $z_{kt} \; \forall k \in K$ (assuming the ϕ_k's are ordered) can be non-zero. Using these auxiliary variables, we approximate the logarithm of the lost sales and the market size to be PWL functions in the z-space, i.e., $\ln \zeta_t \approx \sum_{k \in K} \ln(\bar{\zeta}_{kt}) z_{kt}$ and $\ln \eta_t \approx \sum_{k \in K} \ln(\bar{\eta}_{kt}) z_{kt}$. The corresponding true functions are non-linear near when lost shares are near 0 or 1 and linear in the mid-range. This feature can be exploited to reduce the number of ϕ_k knots used, which otherwise for simplicity can be chosen to be equidistant. The primary decision variables of interest are the continuous bounded-constrained channel attraction and market size coefficients $\boldsymbol{\beta}_m$ and γ respectively with user specified lower and upper bounds ($\mathbf{l}_m$ and $\mathbf{u}_m$ for $\boldsymbol{\beta}_m$ and $\mathbf{l}_s$ and $\mathbf{u}_s$ for γ) and L1 penalties ($\boldsymbol{\rho}_m$ and $\boldsymbol{\rho}_s$). The model used is as follows:

$$\min_{\boldsymbol{\beta}, \boldsymbol{\gamma}, z} \sum_{t \in \mathcal{T}} \sum_{m \in M} \left| \ln(\bar{s}_{mt}) - \sum_{k \in K} \ln(\bar{\zeta}_{kt}) z_{kt} - \ln(f_{m,t}(\boldsymbol{\beta}_m)) \right| \qquad \textbf{(OCDE)}$$

$$+ \sum_{t \in \mathcal{T}} \left| \sum_{k \in K} \ln(\bar{\eta}_{kt}) z_{kt} - \ln(\tau_t(\boldsymbol{\gamma})) \right| + \sum_{m \in M} \boldsymbol{\rho}_m^T |\boldsymbol{\beta}_m| + \boldsymbol{\rho}_s^T |\boldsymbol{\gamma}|$$

$$\sum_{k \in K} z_{kt} = 1 \qquad \forall\, t \in \mathcal{T} \qquad (6.3)$$

$$z_{kt} \geq 0 \qquad \forall\, t \in \mathcal{T}, k \in K \qquad (6.4)$$

$$\{z_{kt} \; \forall\, k \in K\} \in \text{ SOS2} \qquad \forall\, t \in \mathcal{T} \qquad (6.5)$$

$$\mathbf{l}_m \leq \boldsymbol{\beta}_m \leq \mathbf{u}_m \qquad \forall m \in M \qquad (6.6)$$

$$\mathbf{l}_s \leq \boldsymbol{\gamma} \leq \mathbf{u}_s \qquad (6.7)$$

The objective function of the omnichannel demand estimation (OCDE) minimizes the sum of the absolute deviations between the imputed and predicted values of the MNL log-channel shares (first term, where the imputed share is $\frac{\bar{s}_{mt}}{\zeta_t}$) and the log-market size values (second term, where imputed size is η_t) over all the periods in the training data set. The remaining terms represent the L1 (LASSO) penalty that helps to automate variable selection and obtain a sparse model. Constraints (6.3–6.5) ensure that one lost share probability value is chosen for any given period t and the PWL lost sales and market size imputations are obtained endogenously using this probability and observed sales. Coefficient bounds (6.6–6.7) enforce sign constraints and are useful practical safeguards in an automated machine learning environment.

The OCDE optimization models were evaluated on a Windows-7 computer having 8GB RAM, and an Intel Core i7 processor. We used CPLEX 12.6.2 with a branch and bound node limit setting of 10,000 and optimality gap of 1%. Using 10 uniform PWL segments ($|K| = 10$, $\{\phi_k, k \in K\} = \{0.001, 0.1, 0.2, ..., 0.9, 0.999\}$) resulted in a run time of 1.5 seconds on average per instance (i.e., at the UPC-Zone level). We used sign constraints on parameter values (e.g., $\beta_{m1} \leq 0, \beta_{m2,k}, \beta_{m3,j} \geq 0$) and appropriate L1 penalty values in [0.1,1]. Introducing such penalties result in a biased estimator and makes standard significance testing difficult. On the other hand, a tuned LASSO penalty selects variables that enhances out-of-sample forecast accuracy.

With the resultant MNL model obtained from the calibrated parameters and the weekly feature values, we can predict the weekly demand (same as sales in our setting). The preferred method for the retailer to track the forecast accuracy of the MIP-based machine learning approach was using the weighted mean absolute percentage error (WMAPE) metric where t represents the week index:

$$WMAPE = \frac{\sum_t |\text{predicted sales}(t) - \text{actual sales}(t)|}{\sum_t \text{actual sales}(t)} * 100 \tag{6.8}$$

Table 4 reports the achieved out-of sample WMAPE metric for eight-week ahead predictions of weekly sales at the UPC-zone level. The WMAPE value compares well with the estimates of Fisher and Vaidyanathan (2014) who report an out-of-sample sales-weighted MAPE of 40.1% at the store-UPC level and 25.8% at the chain-UPC level for store-sales of automobile tires. Overall, the achieved forecast accuracy by our machine learning approach at this fine level of aggregation satisfied the retailer's expectation.

Inkjet Cartridges		Markers and highlighters	
Brick	Online	Brick	Online
29	36	36	24
(23,31)	(24,45)	(26,44)	(20,27)

Table 3   Eight week average and range (25th and 75th quantile) of the out-of-sample WMAPE fit over the UPC-zone pairs in each category.

From an implementation perspective, continual model re-training is required for a variety of practical reasons: (1) changing customer preferences and demographic trends; (2) new types of promotions; (3) newer product versions and/or channels introduced with limited historical data. Automated re-training of the demand model is triggered prior to every price optimization using the latest available data and with sufficient history to capture price, promotion and temporal effects, so that the OCDE model is used to progressively learn the parameters over time and reduce prediction error. For products with limited data, prior coefficient values with ridge (L2) penalty are used until the estimated coefficient values stabilize. Bootstrapping and Thompson sampling (Russo and Van Roy 2014) based active learning methods can be used to enhance convergence when there is limited price variation in the training data.

MIP-based censored MNL vs. MLE-based uncensored hybrid MNL: We compare our MIP-based censored data estimation accuracy to the uncensored hybrid MNL demand model (Subramanian and Sherali 2010) employed in the retail industry to model consumer choice. Recall that in the hybrid model, total sales are spread down to the purchase choices using an MNL model excluding the no-purchase option (there is no 1 in the denominator of the MNL). This simplifies the estimation procedure into two independent estimations: (1) channel purchase probability (using MLE), and (2) total sales (e.g., using GLMNET library in R statistical software package), where the latter depends on all the channel features as well. The practical drawbacks of hybrid MNL are the non-intuitive cross price elasticities (e.g., positive price elasticity for substitutive items) and an increased degree of non-linearity and non-convexity in the resultant price optimization model.

Table 4 compares the forecast quality of the OCDE model with that of the hybrid demand model for the different scenarios (a standard error of 1% or lower is observed in the reported WMAPEs). We also included UPCs from other product categories that were not part of the BVA. We can see that the OCDE approach either matches or improves upon the WMAPE achieved by the hybrid approach in 9 of the 10 channel-level scenarios. Furthermore, the omnichannel model is strictly better in all the product categories for the online channel.

Category	# UPCs	Omnichannel		Hybrid	
		Brick	Online	Brick	Online
Inkjet Cartridges	50	29	36	29	38
Markers & Highlighters	50	36	24	37	27
Category 3	147	8	12	9	17
Category 4	114	23	21	21	26
Category 5	76	18	16	19	18

Table 4   **Eight week average out-of-sample WMAPE fit over the UPC-zone pairs in each category.**

Estimated own and cross-channel price elasticities: We present the average (and percentile variations) of the same-channel and cross-channel price elasticity values evaluated at the average

channel price in Table 5 across the UPC-Zone data instances. These elasticities range between -2.0 to 0. The relatively low elasticity values are typical of essential consumer products. For example, Krugman and Wells (2008) report price elasticities for various essential and luxury product types and a value of -0.5 for stationery goods. It can also be seen from Table 5 that the cross-channel price elasticities are significant for these categories, and as high as 50% of the own channel price elasticity. We also observe that the cross-elasticities are asymmetric in that the impact of brick prices on the online sales tends to be higher than the impact of the online prices on brick sales. It is indicative of the heterogeneity of the customers shopping in the different channels as well as the volume share of these channels. With the brick channel contributing to 88% of the market share, the absolute change in volume of brick sales though is much higher than that for the online channel. As the online share rises over time, we can expect the online price to exert an increasing influence on the brick channel sales.

Channel \ Price	Inkjet Cartridges		Markers and highlighters	
	Brick price	Online price	Brick price	Online price
Brick sales	-0.66	0.06	-1.19	0.04
	(-1.84,-0.4)	(0,0.22)	(-1.87,-0.01)	(0, 0.05)
Online sales	0.31	-1.04	0.19	-0.78
	(0,1.46)	(-1.99,-0.01)	(0,0.39)	(-1.96,-0.01)

Table 5   **Average and the range (10th and 90th quantile) of the own and cross-channel price elasticities across the UPC-Zone data instances.**

Note that the cross-channel elasticities (off diagonal entries in Table 5), and the impact of brick prices on online sales at a location specific level (lower off-diagonal entry), cannot be computed using traditional pricing models employed in the industry. The value assessment presented in the following section are projections based on the calibrated demand model discussed here.

6.2. Business value assessment

For each of the 50 UPCs in the two product categories we jointly optimize prices in all the physical stores and the online channel using the estimated zone-level omnichannel demand models and the OCP MIP model. For the business case, we explicitly focused on the multi-period regular pricing problem (a typical RPO setting) wherein the goal is to find an optimized price for a product over the last 8 consecutive out-of-sample prediction weeks. The OCP model was extended to include multiple time periods and its demand variations, via a constraint like (4.4) that runs across time[4].

The retailer specified the following goals and business rules for the value assessment using the historical sales data: (1) price bounds to ensure that recommended prices were within historical

---

[4] More precisely, we replace the index j and set J in the OCP-MIP with indicies j, t and sets J and T, where $j \in J$ and $t \in T$. We also include the equality constraint: $z_{mji} = z_{mjti} \ \forall i \in I_{mj}, m \in M, j \in J, t \in T$.
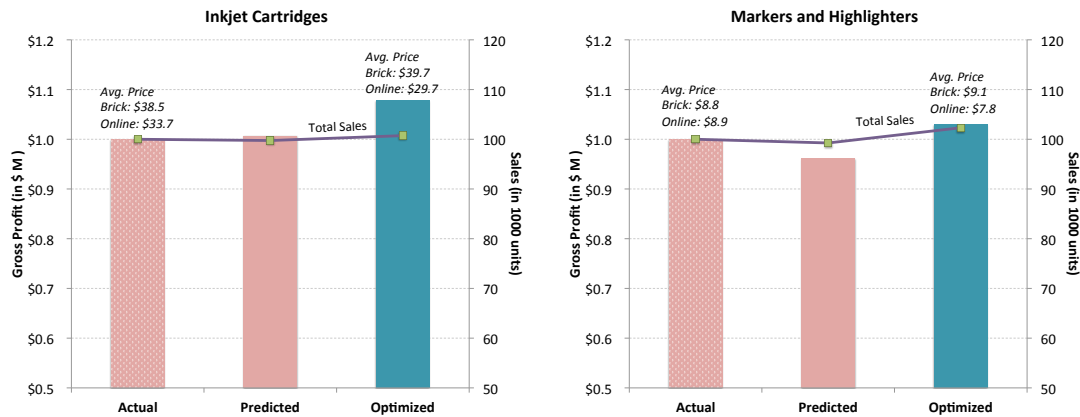
Figure 5   Re-normalized gross-profit and sales during an 8-week period for actuals from historical data, predicted using omnichannel demand models at actual prices and the optimization.

values allowing a 10% variation (20 price points per ladder with magic number endings); (2) a global volume goal to preserve sales volume that required the predicted volume at the optimal prices to be no less than the predicted volume at the baseline (actual) prices; and (3) a price balancing constraint that required the average channel price at a location to be with a preset factor of the corresponding baseline value. The motivation for the last constraint is to prevent a myopic response of raising prices at any location that can gradually erode market share.

The OCP optimization models developed as a JAVA API were evaluated on a Windows-7 computer having 8GB RAM, and an Intel Core i7 processor. CPLEX 12.6.2 with its out-of-box parameter settings was used to solve the resultant MIPs to global optimality. The average problem size after preprocessing was about 30K rows, 20K columns, and 550 binary variables. The average runtime per UPC was 1.7 seconds, and no more than 3 seconds in the worst case. Thus, the solution response is fast enough to perform multiple price updates within a day making it a suitable solution for dynamic omnichannel pricing.

The results of this optimization are presented in Fig. 5 along with two baselines. The first baseline (actual) represents the historically realized KPIs. The second baseline (predicted) represents the predicted KPIs using the actual prices and the calibrated omnichannel demand model. To protect the retailer's data privacy, the actual gross profits are normalized to \$1M and the sales volume to 100K units and the results for predicted and optimized are relative to these normalized actuals.

We observe that the predicted and realized metrics are close to each other for both categories, i.e., within 1% in terms of sales volume and gross profit each for inkjet cartridges, and within 1% in terms of sales volume and 4% in terms of gross profit for markers and highlighters. This baseline result is important because it boosted the confidence of all the stakeholders in the accuracy of the calibrated omnichannel demand model and was accepted as a valid model for the BVA projections. The optimized metrics project a 7% gross profit lift in both categories over the predicted value. This
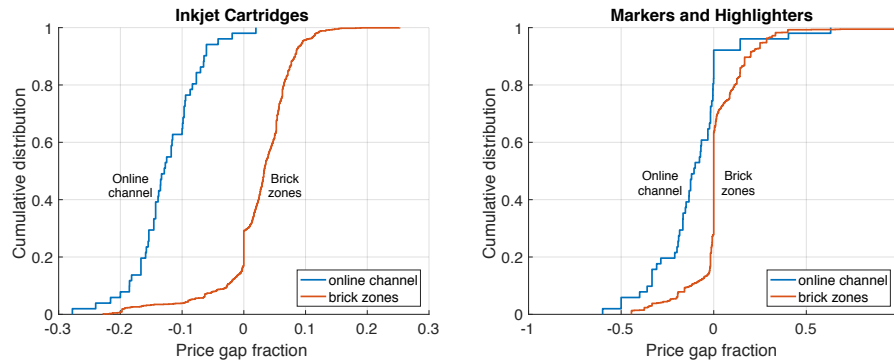
Figure 6   **Cumulative distribution of UPC (UPC-Zone pairs for brick) as a function of the relative change of the optimized channel price compared to the actual price.**

is notable because it is achieved while also increasing the corresponding aggregate sales volume by 1% and 3% respectively. This gain is obtained by lowering the average optimized online prices from actuals by about 12%, while the average optimized store prices are increased by 2.5% in each of the product categories.

Fig. 6 displays the cumulative distribution of the UPCs (UPC-Zone pairs for the brick stores) as a function of the relative change of the optimized channel price from the actual price. For example, the y-axis value at the 'zero' point on the x-axis gives us the total proportion of UPCs (or UPC-Zone pairs) whose optimized channel prices are at or below the actual price. In the inkjet category, we observe that the optimization increased brick prices in 70% of the zones, while retaining or lowering brick prices in 30% of the locations. On the other hand, for the markers, only 37% of the brick locations witness a price rise (due to the higher own brick price elasticity of markers over the inkjet category). In both categories, the optimization predominantly lowered the online channel price for all UPCs compared to the brick zones. This is due to the higher elasticity of the online channel relative to the stores. Such a pricing solution boosts online sales, allowing the retailer to be more competitive with online retailers in the marketplace without trying to aggressively match the low price of e-tail giants.

The above results were presented at the retailer site to a team of pricing analysts and senior executives, including their Vice-President for revenue management. OCP's demonstrated ability to accurately model omnichannel customer behavior and integrate competitive channel pricing strategies was highly valued by IBM Commerce and their retail customers. The overall feedback was positive, and with similar experiences with other retailers in a variety of product categories. Subsequently, a proprietary version of the solution was deployed into production by IBM Commerce in 2014 and showcased as one of the retail analytics success stories in an IBM smarter-commerce global summit the same year. It included a presentation on its capabilities by a partner retail chain. This work was formally recognized by IBM as one of the research accomplishments in 2015.

Appendix A: Proof of Proposition 1

We prove this proposition with a contradiction. Let us assume the general price monotonicity (or a volume measure) constraint (4.7) is convex using market share variables. Consider a special case of this constraint: $p_{mj} \leq \gamma p_{m'j}$, where $\gamma \neq 1$ is a constant. We drop the location subscript j for ease of notation. Suppose that $f_m(p) = e^{a_m - b_m p}$, $f_{m'}(p) = e^{a_{m'} - b_{m'} p}$ and that $b_m \neq b_{m'}$. The price monotonicity constraint can be written as follows in the market share variables:

$$\frac{a_m}{b_m} - \frac{1}{b_m} \log\left(\frac{\theta_m}{1 - \theta_m}\right) \leq \frac{\gamma a_{m'}}{b_{m'}} - \frac{\gamma}{b_{m'}} \log\left(\frac{\theta_{m'}}{1 - \theta_{m'}}\right)$$

This constraint results in an irreducible non-linearity and is non-convex unless $\gamma = 1$ and $b_m = b_{m'}$. Hence the contradiction.   □

Appendix B: Proof of Proposition 2

We achieve this result by performing a reduction from the *2-class logit assortment optimization problem* (2CL) which is shown to be NP-complete by Rusmevichientong et al. (2010). The inputs and the problem are as follows:

2-Class Logit (2CL) Problem:

Inputs:

- The set of products indexed by $1, 2, \cdots, n$.
- The profits $w_1, w_2, \cdots, w_n$, where $w_i \in \mathbb{Z}_+ \forall i = 1, 2, \cdots, n$.
- The preference weights $(v_1^g, \cdots, v_n^g)$, where $v_i^g \in \mathbb{Z}_+ \forall i = 1, 2, \cdots, n, g = 1, 2$.
- The relative weights of the two classes, $\alpha_1, \alpha_2 \in \mathbb{Q}^+$ such that $\alpha_1 + \alpha_2 = 1$.
- The target profit $K \in \mathbb{Z}_+$.

Question: Is there an assortment $S \subset \{1, \cdots, n\}$ whose expected profit is at least K, that is

$$\alpha_1 \frac{\sum_{i \in S} w_i v_i^1}{1 + \sum_{i \in S} v_i^1} + \alpha_2 \frac{\sum_{i \in S} w_i v_i^2}{1 + \sum_{i \in S} v_i^2} \geq K \tag{B.1}$$

We similarly describe the 2-location OCP problem as follows:

Two location OCP (2L-OCP) Problem:

Inputs:

- The set of channels indexed by $I = 1, 2, \cdots, n$.
- The attraction functions $(f_1^g(p_1), \cdots, f_n^g(p_n))$, with the following mild regularity assumption:

$$\lim_{p_i \to \infty} p_i \frac{f_i^g(p_i)}{1 + f_i^g(p_i)} = 0 \ \forall i \in I, g = 1, 2. \tag{B.2}$$

   The regularity condition ensures that there exists a finite, possible large, price where the revenue contribution of a virtual channel can be reduced to 0 or close enough to 0.

- The cost of offering the item on a channel is $c_i \ \forall i \in I$.
- The weights across locations, τ_1, τ_2.
- The target profit $K \in \mathbb{Z}_+$.

Question: Is there a price vector $(p_1^*, ..., p_n^*) \in \Omega_1 \times \cdots \times \Omega_n$ such that Ω_i is a finite set of discrete prices, whose expected profit is at least K, that is

$$\tau_1 \frac{\sum_{i \in I} (p_i - c_i) f_i^1(p_i)}{1 + \sum_{i \in I} f_i^1(p_i)} + \tau_2 \frac{\sum_{i \in I} (p_i - c_i) f_i^2(p_i)}{1 + \sum_{i \in I} f_i^2(p_i)} \geq K \tag{B.3}$$

Clearly the 2L-OCP problem is in NP. Consider an arbitrary instance of 2CL. We map every item $1, \cdots, n$ to a distinct virtual channel in the 2-location OCP problem, i.e., $I = 1, \cdots, n$ and set $c_i = 0 \ \forall i \in I$ and $\tau_g = \alpha_g$ for $g = 1, 2$. We consider an attraction function with two finite discrete price offers $(\bar{p}_i, \bar{p}'_{\max})$ for each channel $i \in I$ where $\bar{p}_i = w_i$

32

Harsha, Subramanian and Ettl: *Omnichannel demand modeling and price optimization*
Article submitted to *INFORMS Journal on Optimization*; manuscript no. (Please, provide the manuscript number!)

and $f_i^g(\bar{p}_i) = v_i^g$ for $g = 1, 2$ and a very large but finite price $\bar{p}'_{\max}$ is chosen, constant across channels, such that the following inequalities are satisfied:

$$\bar{p}'_{\max} > \sum_{i \in I, g = 1, 2} w_i v_i^g \tag{B.4}$$

$$\bar{p}'_{\max} \frac{f_i^g(\bar{p}'_{\max})}{1 + f_i^g(\bar{p}'_{\max})} \leq \frac{1}{2n} \delta \ \forall i \in I, g = 1, 2 \tag{B.5}$$

where $\delta = \min_{g=1,2} \frac{1}{1 + \sum_{i \in I} v_i^g}$. Note that because of Eq. (B.2), a finite $\bar{p}'_{\max}$ can be identified.

We first show that given an assortment S for the 2CL problem, its objective $2CL_{Obj}(S)$ can be used to upper and lower bound the OCP objective when channels in S offer price $\bar{p}_i$ respectively for each $i \in I$ and the rest offer $\bar{p}'_{\max}$. We denote this OCP objective by $OCP_{Obj}(S)$.

$$OCP_{Obj}(S) = \tau_1 \frac{\sum_{i \in S} \bar{p}_i f_i^1(\bar{p}_i) + \sum_{i \notin S} \bar{p}'_{\max} f_i^1(\bar{p}'_{\max})}{1 + \sum_{i \in S} f_i^1(\bar{p}_i) + \sum_{i \notin S} f_i^1(\bar{p}'_{\max})} + \tau_2 \frac{\sum_{i \in S} \bar{p}_i f_i^2(\bar{p}_i) + \sum_{i \notin S} \bar{p}'_{\max} f_i^2(\bar{p}'_{\max})}{1 + \sum_{i \in S} f_i^2(\bar{p}_i) + \sum_{i \notin S} f_i^2(\bar{p}'_{\max})} \tag{B.6}$$

$$\leq 2CL_{Obj}(S) + \tau_1 \sum_{i \notin S} \frac{1}{2n} \delta + \tau_2 \sum_{i \notin S} \frac{1}{2n} \delta \tag{B.7}$$

$$\leq 2CL_{Obj}(S) + \frac{\delta}{2} \tag{B.8}$$

The first inequality is because $\bar{p}'_{\max}$ satisfies (B.5). Moreover, because $\bar{p}'_{\max}$ satisfies (B.4),

$$2CL_{Obj}(S) \leq OCP_{Obj}(S). \tag{B.9}$$

Now suppose there exists an assortment S such that $2CL_{Obj}(S) \geq K$ then from (B.9), it is easy to conclude that, $OCP_{Obj}(S) \geq K$. But suppose there exists no assortment S such that $2CL_{Obj}(S) \geq K$, i.e., $2CL^*_{Obj} < K$ where $2CL^*_{Obj} = \max_S 2CL_{Obj}(S)$, then because of the following observation, we conclude that $OCP^*_{Obj} < K$ where $OCP^*_{Obj} = \max_S OCP_{Obj}(S)$. The key observation is that $2CL^*_{Obj}$ is either an integer (less than or equal to $K - 1$) or

$$\lceil 2CL^*_{Obj} \rceil - 2CL^*_{Obj} > \delta, \tag{B.10}$$

because $v_i^g \in \mathbb{Z}_+$ and $\alpha_1 + \alpha_2 = 1$. In either case, now $2CL^*_{Obj} + \frac{\delta}{2} < K$ and hence $OCP^*_{Obj} < K$. This concludes the reduction, showing that 2L-OCP is NP-complete (its optimization counterpart is NP-Hard). □

Appendix C: Proof of Proposition 4

All transformations to obtain the OCPN-PWL formulated are explained below Proposition 4. The transformed model uses two approximations: (1) the PWL approximation of the normalized intrinsic value curve using SOS2 variables, w_{ms}; and (2) the RLT constraints with the same SOS2 variables to remove the non-linearity $\rho_{ms} = \bar{\theta} w_{ms}$. With the increase in the number of knots, the approximation due to (1) is by definition exact. What remains to show is that $\sum_{s \in S} \bar{R}_{ms} \rho_{ms}$ converges to term $\sum_{s \in S} \bar{R}_{ms} \bar{\theta} w_{ms}$ when the knots increase, i.e., when $|\bar{U}_{ms1} - \bar{U}_{ms2}| \to 0$ for any $s1, s2$ that are adjacent indicies. Because $w_{ms} \ \forall s \in S$ are SOS2 variables, without loss of generality, say $w_{ms1}, w_{ms2} \geq 0$ and the rest are all zero. Therefore, $w_{ms1} + w_{ms2} = 1$ and $\rho_{ms1} + \rho_{ms2} = \bar{\theta}$. Now substituting for ρ_{ms2} and w_{ms2} in the respective terms, they reduce to $(\bar{R}_{ms1} - \bar{R}_{ms2})\rho_{ms1} + \bar{R}_{ms2}\bar{\theta}$ and $(\bar{R}_{ms1} - \bar{R}_{ms2})\bar{\theta} w_{ms1} + \bar{R}_{ms2}\bar{\theta}$, which converge in the limit as $|\bar{R}_{ms1} - \bar{R}_{ms2}| \to 0$ when $|\bar{U}_{ms1} - \bar{U}_{ms2}| \to 0$. □

Appendix D: Proof of Proposition 5

We first show that the new formulation is always an upper bound. Because all the transformation steps are exact (in the integer sense) except the lower and upper bounds on R_m, it suffices to show that even with the RLT relaxation,

and the introduction of the variable ρ_{ms}, the lower and upper bounds to the true value of R_m given in Eq. (5.3) hold. This means we have to show that the following holds:

$$\sum_{s \in S} \bar{R}_{ms} \rho_{ms} \leq \bar{\theta}(U_m^{\max} \eta)^{\gamma_m} \tag{D.1}$$

$$\bar{\theta}(U_m^{\max} \eta)^{\gamma_m} \leq \bar{R}_m^t \left[(1 - \gamma_m)\bar{\theta} + \gamma_m \sum_{s \in S} \frac{\bar{\eta}_{ms}}{\bar{\eta}_m^t} \rho_{ms} \right] \qquad \forall t \in T \tag{D.2}$$

Because $w_{ms} \; \forall s \in S$ are SOS2 variables, without loss of generality, say $w_{ms1}, w_{ms2} \geq 0$ and the rest are all zero. Therefore, $w_{ms1} + w_{ms2} = 1$ and $\rho_{ms1} + \rho_{ms2} = \bar{\theta}$. Note that here $\eta = \bar{\eta}_{ms1} w_{ms1} + \bar{\eta}_{ms2} w_{ms2}$.

Now the left hand side (LHS) of constraint (D.1) can be simplified to $\bar{\theta} \left[\bar{R}_{ms1} \frac{\rho_{ms1}}{\bar{\theta}} + \bar{R}_{ms2} \frac{\rho_{ms2}}{\bar{\theta}} \right] = \bar{\theta}(U_m^{\max})^{\gamma_m} \left[\bar{\eta}_{ms1}^{\gamma_m} \frac{\rho_{ms1}}{\bar{\theta}} + \bar{\eta}_{ms2}^{\gamma_m} \frac{\rho_{ms2}}{\bar{\theta}} \right]$. We can deduce that it is always less than the right hand side (RHS) of constraint (D.1) as the RHS is a concave function and LHS is a piecewise linear lower bound.

Now consider the RHS of constraint (D.2). This can be simplified as follows: $\bar{R}_m^t \bar{\theta} \left[(1 - \lambda_m) + \lambda_m \frac{\bar{\eta}_{ms1} \rho_{ms1} + \bar{\eta}_{ms2} \rho_{ms2}}{\bar{\theta} \bar{\eta}_{mt}} \right]$. This simplifies to a tangent to the concave function on the LHS and hence is always an upper bound. Thus, this proves that even with the RLT relaxation of the SOS2 variables, constraints (D.1–D.2) relax the function at its true value and hence the optimal solution of this formulation is an upper bound on OCPN.

The proof of asymptotic optimality follows the same steps as Proposition 4, except that here we use it both on the lower and the upper bound. $\qquad \square$

Appendix E: Price matching and the practical value of global optimality

Consider a retailer who would rather price-match the channels if the profitability gain (by not price-matching) is insufficient. The retailer can conveniently specify this tradeoff via a profitability threshold, where in, a price match constraint is accepted only if the resultant drop in profitability is within the threshold.

A natural way of implementing this feature is to solve the OCP problem with, and without the price-matching constraint and then pick the preferred solution. Employing an exact solution approach turns out to be critical in this context. The use of a local-optimum based heuristic approach to solving OCP with and without the price matching constraint, can result in incorrect profitability gaps, producing 'false-positive' price-matching recommendations. Note that for such heuristic methods, a price-matching constraint can operate like a cutting-plane that deletes a local (but not global) optimal solution to the unconstrained problem (i.e., without the price-matching constraint), potentially yielding an improved profitability objective. In such cases, the heuristic approach is likely to approve price-matching, whereas, the unconstrained profit value it achieves may have been far from optimality. Numerical testing clearly demonstrated that such false-positives were not uncommon. On the other hand, an optimal approach always generates the correct price-coordination recommendation, and the application produces stable and predictable responses from a user perspective.

Acknowledgments

34

Harsha, Subramanian and Ettl: *Omnichannel demand modeling and price optimization*
Article submitted to *INFORMS Journal on Optimization*; manuscript no. (Please, provide the manuscript number!)

References

Aggarwal, Gagan, Tomás Feder, Rajeev Motwani, An Zhu. 2004. Algorithms for multi-product pricing. *Automata, Languages and Programming*. Springer Berlin Heidelberg, 72–83.

Akçay, Yalçın, Harihara Prasad Natarajan, Susan H Xu. 2010. Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science* **56**(8) 1345–1361.

Ansari, Asim, Carl F Mela, Scott A Neslin. 2008. Customer channel migration. *Journal of Marketing Research* **45**(1) 60–76.

Aydin, Goker, Evan L Porteus. 2008. Joint inventory and pricing decisions for an assortment. *Operations Research* **56**(5) 1247–1255.

Bell, D. R., S. Gallino, A. Moreno. 2014. How to win in an omnichannel world. *Sloan Management Review* .

Berkson, Joseph. 1953. A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *Journal of the American Statistical Association* **48**(263) 565–599.

Bitran, Gabriel, René Caldentey. 2003. An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* **5**(3) 203–229.

Bront, Juan José Miranda, Isabel Méndez-Díaz, Gustavo Vulcano. 2009. A column generation algorithm for choice-based network revenue management. *Operations Research* **57**(3) 769–784.

Brynjolfsson, E., Y. J. Hu, M. S. Rahman. 2013. Competing in the age of omnichannel retailing. *Sloan Management Review* .

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

Charnes, Abraham, William W Cooper. 1962. Programming with linear fractional functionals. *Naval Research logistics quarterly* **9**(3-4) 181–186.

Chen, Xin, David Simchi-Levi. 2012. *The Oxford handbook of pricing management*, chap. Pricing and inventory management. Oxford University Press, 784–822.

Chintagunta, Pradeep K, Junhong Chu, Javier Cebollada. 2012. Quantifying transaction costs in online/offline grocery channel choice. *Marketing Science* **31**(1) 96–114.

Crowder, Harlan, Ellis L Johnson, Manfred Padberg. 1983. Solving large-scale zero-one linear programming problems. *Operations Research* **31**(5) 803–834.

Davis, James M, Huseyin Topaloglu, David P Williamson. 2016. Pricing problems under the nested logit model with a quality consistency constraint. *INFORMS Journal on Computing* **29**(1) 54–76.

Domencich, Thomas A, Daniel McFadden. 1975. *Urban travel demand-a behavioral analysis*. North-Holland.

Dong, Lingxiu, Panos Kouvelis, Zhongjun Tian. 2009. Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management* **11**(2) 317–339.

Elmaghraby, Wedad, Pinar Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science* **49**(10) 1287–1309.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2015. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* **18**(1) 69–88.

Fisher, Marshall, Santiago Gallino, Jun Li. 2017. Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science* **64**(6) 2496–2514.

Fisher, Marshall, Ramnath Vaidyanathan. 2014. A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science* **60**(10) 2401–2415.

Gallego, Guillermo, Ruxian Wang. 2014. Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* **62**(2) 450–461.

Gallino, Santiago, Antonio Moreno. 2014. Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science* **60**(6) 1434–1451.

Gao, Fei, Xuanming Su. 2016. Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science* **63**(8) 2478–2492.

Goolsbee, Austan. 2001. Competition in the computer industry: Online versus retail. *The Journal of Industrial Economics* **49**(4) 487–499.

Govindarajan, Aravind, Amitabh Sinha, Joline Uichanco. 2018. Joint inventory and fulfillment decisions for omnichannel retail networks.

Hanson, Ward, Kipp Martin. 1996. Optimizing multinomial logit profit functions. *Management Science* **42**(7) 992–1003.

Jalilipour Alishah, Elnaz, Kamran Moinzadeh, Yong-Pin Zhou. 2015. Inventory fulfillment strategies for an omni-channel retailer.

Kahn, Barbara E. 2018. *The Shopping Revolution: How Successful Retailers Win Customers in an Era of Endless Disruption*. Wharton Digital Press.

Keller, Philipp W, Retsef Levi, Georgia Perakis. 2014. Efficient formulations for pricing under attraction demand models. *Mathematical Programming* **145**(1-2) 223–261.

Krugman, Paul, Robin Wells. 2008. *Microeconomics*. 2nd ed. New York: Worth Publishers.

Lei, Yanzhe, Stefanus Jasin, Amitabh Sinha. 2018. Joint dynamic pricing and order fulfillment for e-commerce retailers. *Manufacturing & Service Operations Management* **20**(2) 269–284.

Li, Hongmin, Woonghee Tim Huh. 2011. Pricing multiple products with the multinomial logit and nested logit models. *Manufacturing & Service Operations Management* **13**(4) 549–563.

Li, Hongmin, Scott Webster, Nicholas Mason, Karl Kempf. 2017. Product-line pricing under discrete mixed multinomial logit demand. *Manufacturing & Service Operations Management* .

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics* 105–142.

Mulhern, Francis J, Jerome D Williams, Robert P Leone. 1998. Variability of brand price elasticities across retail stores: Ethnic, income, and brand determinants. *Journal of Retailing* **74**(3) 427–446.

Rayfield, W Zachary, Paat Rusmevichientong, Huseyin Topaloglu. 2015. Approximation methods for pricing problems under the nested logit model with price bounds. *Journal on Computing* **27**(2) 335–357.

Rusmevichientong, Paat, David Shmoys, Huseyin Topaloglu. 2010. Assortment optimization with mixtures of logits. Tech. rep., Tech. rep., School of IEOR, Cornell University.

Rusmevichientong, Paat, Benjamin Van Roy, Peter W Glynn. 2006. A nonparametric approach to multi-product pricing. *Operations Research* **54**(1) 82–98.

Russo, Daniel, Benjamin Van Roy. 2014. Learning to optimize via posterior sampling. *Mathematics of Operations Research* **39**(4) 1221–1243.

Schön, Cornelia. 2010. On the product line selection problem under attraction choice models of consumer behavior. *European Journal of Operational Research* **206**(1) 260–264.

Sherali, Hanif D, Warren P Adams. 1999. Reformulation-linearization techniques for discrete optimization problems. *Handbook of combinatorial optimization*. Springer, 479–532.

Song, Jing-Sheng, Zhengliang Xue. 2007. Demand management and inventory control for substitutable products. *Working paper* .

Subramanian, Shivaram, Pavithra Harsha. 2017. Demand modeling in the presence of unobserved lost sales. Submitted, IBM Research.

Subramanian, Shivaram, Hanif D. Sherali. 2010. A fractional programming approach for retail category price optimization. *Journal of Global Optimization* **48**(2) 263–277.

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

Train, Kenneth E. 2009. *Discrete choice methods with simulation*. 2nd ed. Cambridge university press.

Urban, Glen L. 1969. A mathematical modeling approach to product line decisions. *Journal of Marketing Research* 40–47.

US Census Bureau. 2017. Quarterly retail e-commerce sales. URL https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf.

Zhang, Jie, Paul W Farris, John W Irvin, Tarun Kushwaha, Thomas J Steenburgh, Barton A Weitz. 2010. Crafting integrated multichannel retailing strategies. *Journal of Interactive Marketing* **24**(2) 168–180.

Zhu, Y, A Deshpande, B Quanz, L Cao, A Koc, X Liu, Y Li, A Hampapur, C Milite, M Desimone, et al. 2017. Improving e-commerce order fulfillment for peak times via incorporating fulfillment network load balancing. *IIE Annual Conference. Proceedings*. 1806–1810.