MICHIGAN ROSS

Working Paper

Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments

Marshall Fisher
The Wharton School
University of Pennsylvania

Santiago Gallino
Tuck School of Business
Darthmouth College

Jun Li
Stephen M. Ross School of Business
at the University of Michigan

Ross School of Business Working Paper Series
Working Paper No. 1265
January 2017
Forthcoming at *Management Science*

Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments

Marshall Fisher
The Wharton School, University of Pennsylvania, fisher@wharton.upenn.edu

Santiago Gallino
Tuck School of Business, Dartmouth College, santiago.gallino@tuck.dartmouth.edu

Jun Li
Ross School of Business, University of Michigan, junwli@umich.edu

A retailer following a competition-based dynamic pricing strategy tracks competitors' price changes and then must answer the following questions: (1) Should we respond? (2) If so, respond to whom? (3) How much of a response? (4) And on which products? The answers require unbiased measures of price elasticity as well as accurate estimates of competitor significance and the extent to which consumers compare prices across retailers. There are two key challenges to quantify these factors empirically: first, the endogeneity associated with almost any type of observational data, where prices are correlated with demand shocks observable to pricing managers but not to researchers, and second, the absence of competitor sales information, which prevents efficient estimation of a full consumer-choice model. We address the first issue by conducting a field experiment with randomized prices. We resolve the second issue by exploiting the retailer's own and competitors' stockouts as a source of variation to the consumer choice set, in addition to variations in competitors' prices. We estimate an empirical model capturing consumer choices among substitutable products from multiple retailers. Based on the estimates, we propose and test a best-response pricing strategy through a carefully controlled live experiment that lasts five weeks. The experiment documents an 11 percent revenue increase while maintaining a margin above a retailer-specified target.

1. Introduction

The Internet has changed the way price information is disseminated. With just a few clicks, consumers are able to obtain price information from multiple retailers. This increased price transparency induces fierce competition among online retailers and requires real-time monitoring and quick responses to competition.[1] Many online retailers in turn adopt a competition-based pricing strategy wherein they constantly monitor competitors' prices and use this information as an input in setting their own prices. For example, they may always charge X dollars or X percent less or more than a target competitor or any competitor with the lowest price. Retailers miss several

[1] Coming soon: Toilet paper priced like airline tickets. *The Wall Street Journal*. September 5, 2012.

opportunities with such simple heuristics. Instead, they should ask; shouldn't my reaction depend on consumers' elasticity to prices? Shouldn't my reaction depend on the extent to which consumers compare prices across retailers? Shouldn't my reaction depend on changes in availability at competing retailers? Should I still match prices if it seems like the competitor made a pricing mistake? We address these questions in this paper.

Determining the best-response price requires knowing how demand reacts to price changes. This is a challenging task. Simply regressing historical sales on prices while controlling for observable product characteristics and seasonality usually suffers from endogeneity issues (see Villas-Boas and Winer 1999). Pricing managers often observe demand signals that we researchers do not, such as unobserved product characteristics or a temporal surge in demand due to manufacturer advertisements, and they may adjust prices based on these signals. If they increase prices when they see a demand surge, this creates a positive correlation that fallaciously implies that a higher price results in higher demand. Moreover, the relationship between demand and price will be further confounded by the price levels of substitutable products that the same retailer offers and the price levels of the same and similar products that competitors offer.

To determine the best-response price, we also need to understand the extent to which consumers compare prices across retailers. In a situation where consumers always choose the cheapest retailer for any product they buy, retailers need to either charge the lowest price in the market or accept no sales. However, in a situation where consumers consider both prices and retail channels when making purchase decisions, retailers may not need to match every single price that a competitor charges, especially when considering products with low price elasticities or less relevant competitors. Accurately assessing the level of consumer engagement in price comparison across retailers will enable targeted price responses that are efficient and effective.

We partnered with a leading Chinese online retailer, which we will refer to hereafter as the "retail partner," to address these challenges. We begin by developing a demand model to understand how consumers make choices when given a set of substitutable products from multiple retailers. Our model resolves a key challenge many retailers face when attempting to implement a choice model to understand consumer purchase decisions: absence of competitor sales information. We exploit the retailer's own and competitors' stockouts as a source of variation for identification, in addition to competitive price variations. Because stockouts alter the consumer choice set, they provide additional moment conditions that are instrumental in estimating consumers' preferences for retailers and their level of engagement in comparing prices across retailers. Next, we conduct a randomized price experiment to obtain unbiased estimates of price elasticities addressing the endogeneity challenge. In this experiment, we randomly assign prices to each product under study using a fractional factorial design. After obtaining model estimates using data generated during the one-month experiment period, we solve a constrained optimization problem to define optimal price responses to competitor price changes. Finally, in collaboration with our retail partner, we evaluate the performance of our best-response pricing algorithm through a carefully controlled field experiment. The daily categorical revenue of the treatment group increases by 11 percent following our methodology, compared to the control group for the same before and after periods.

1.1. Contributions

Our paper contributes to the operations management literature and to retail practice in a number of ways. In what follows, we discuss our contributions and the relevant literature.

Revenue Management and Dynamic Pricing in Retailing. Competition-based dynamic pricing is a recent development driven by the competitive nature of online retailing. Conventionally, dynamic pricing has been applied mostly in settings with perishable inventory and a finite selling season. For example, Caro and Gallien (2012) study the development and implementation of a clearance pricing model for the fast-fashion retailer Zara. Harsha et al. (2016) propose a new revenue management model to maximize total profit in omni-channel retailing and document on average 6 percent improvement in markdown revenue through a proprietary pilot with a major US retailer. Cohen et al. (2016) propose a linear integer programming approximation to solve the dynamic promotional pricing problem, and show that their method can improve profit by 3 percent using sales data from a grocery retailer.

In our setting, however, the need for dynamic pricing arises not from constrained capacity or a perishable product or service, but from constantly changing market conditions due to competition. This dynamic market environment poses new challenges to, and opportunities for, retail pricing. The longstanding literature of retail pricing focuses mostly on pricing and promotion decisions for a single brick-and-mortar retailer holding competitive prices constant (e.g., Caro and Gallien 2012, Cohen et al. 2016) or a long-term competitive pricing strategy (e.g., Lal and Rao 1997). In these traditional retail settings, the pressure for frequent competitive responses is less prevalent due to the customer's high search costs and the retailer's high menu cost of changing prices. Our work expands the existing literature on this topic, since these two factors are not present in an online setting, which makes competition-based dynamic pricing a very relevant issue.

Consumer Choice Modeling. We propose a parsimonious choice model that captures the key tension involved in this competitive environment. A critical feature of the model is to describe how consumers choose products among all competing options, including products offered by our retail partner and its major competitors. In particular, modeling and estimating the substitution across retailers is essential to define the correct responses to competitors' price and availability changes.

Our model follows the choice model framework pioneered by Guadagni and Little (1983). This framework has been increasingly applied to study consumer behavior and firms' pricing decision in retailing. For example, Kök and Xu (2011) solve the joint assortment and pricing problem under a nested logit model. Honhon et al. (2010) solves the assortment planning and inventory decisions problem under arbitrary customer preferences and substitution behavior. More recently, Farias et al. (2013) propose a nonparametric approach to predict revenue from offering a particular assortment of choices. Jagabathula and Rusmevichientong (2016) solve the profit-maximizing assortment and prices using a nonparametric choice model.

What distinguishes our approach from a standard choice model is the incomplete information on choice decisions. In particular, we do not observe choices made on competitors' products, a common challenge almost all retailers face. If we did observe these choices, it would be straightforward to apply a standard multinomial logit model or some variation of it to estimate how consumers make choices among all options, where each option is a retailer-product pair. Absence of competitors' sales information, the classical approach of incorporating a consumer' purchase of a competitor's product is to model it as part of the outside option (Vulcano et al. 2012, Haensel and Koole 2011). Such an approach, however, does not allow us to explicitly estimate consumers' retailer preferences and the extent to which customer engage in price comparison across retailers. Both components are key to identify substitution patterns across competing retailers. Though variations in product features such as prices assist identification (Newman et al. 2013), they may provide only weak identification in settings with limited price variations or high price correlation among competitors. Our identification strategy exploits temporary variations in consumer choice sets through our retail partner and competitor stockouts, in addition to competitive price variations.

What further distinguish our work from this literature is the use of field experiments to demonstrate the strength of the method. We design two experiments: one is designed to correct for price endogeneity in estimating demand models, and the other to test the applicability and effectiveness of our methodology in real competitive business settings. We now discuss them in detail.

Field Experiments to Test Price Elasticities and the Best Response Pricing Algorithm. In the retail setting, we are aware of two papers that use field experiment to study the impact of price variations: Gaur and Fisher (2005) and Ferreira et al. (2016). These papers focus on how demand varies with prices for products sold by the retailer, holding competition constant. The key differences between our work and theirs include the presence of competition, stockouts, and the fast-changing online environment, which calls for dynamic responses.

We first conduct a randomized price experiment in the field to obtain unbiased measures of price elasticities, thereby overcoming several problems that arise in the observational data of our retail partner. Specifically, demand can appear inelastic to price due to the lack of price variations, which happens very often when selling millions of products online. It also happened that, even if the price of a product itself varied, it followed competitors' prices closely such that there is little price variation comparatively. In this situation, it is impossible to distinguish how demand responds to changes in one retailer's own price versus changes in competitors' prices.

Moreover, price elasticity estimates can be biased upward when ignoring the fact that retail managers make price decisions based on private demand signals. The effectiveness of commonly used instrumental variables, including cost shifters, average prices in other geographically disparate locations or lagged prices, tend to be compromised by a combination of factors unique in online retailing. For example, cost shifters do not change nearly as fast as prices for many products, while for other products with almost constant prices, lagged and current prices are highly collinear. Lastly, prices of the same product in other locations are not always available, as some products are not sold in other locations or are off-shelf for the majority of time under study.

It is also important to note that with the presence of competition, price randomization alone will not necessarily guarantee unbiased estimation of elasticities unless competitors' actions are properly accounted for. Ignoring competitors' reactions to our retail partner price changes would bias the estimation because prices can still be correlated with unobserved demand shocks through correlation with competitors' prices. This is why we account for changes in competitors' prices and product availability by using the consumer choice model.

We then conduct a second field experiment to test the best-response pricing algorithm based on the estimated consumer choice model. We assign products to two groups to minimize substitution across groups while allowing substitution within each group. We conduct the experiment by alternating the treatment between the two groups in one geographical region, while using two other similar but disparate regions where the retailer operates as comparisons. This design leads to a difference-in-differences-in-differences estimator, which allows us to correct for the potential differences in demand trends between the control and treatment groups using comparison groups subject to similar but independent demand. In the experiment, we exert care in framing and communicating the experiment to the pricing managers such that (1) price managers in other regions are unaware of the ongoing experiment and (2) the experiment is not framed as a test of an algorithm to replace current practice, but rather as a decision support tool. The experiment demonstrates an 11 percent revenue increase while maintaining a margin above a retailer-specified target.

In sum, our methodology has stood the test of a real competitive business environment and demonstrated tangible revenue improvements. Our work helps navigate and evaluate the trade-offs involved in bridging theoretical, empirical, and field work. It thus presents a scalable and replicable methodology to set dynamic prices in online retailing.

2. Empirical Setting

Our retail partner is a leading Chinese online retailer that originally focused on consumer packaged goods but evolved over time to be a hypermarket. Founded in July 2008, the firm achieved sales of

$1.9 billion in 2013 and was identified by a Deloitte survey in 2011 as the fastest growing technology company in the Asia-Pacific region, with a three-year revenue growth of 19,218 percent.[2] In our study, we focus on one particular product category: baby-feeding bottles.

Pricing decisions are present in every category the retailer offers. We initially focused on one category to allow us to consider all the different factors affecting the pricing decision. Because experimentation is an important component of our research approach, the category must be one with which the retailer is willing to experiment. For these reasons, the retailer suggested we start with baby-feeding bottles before rolling out to other categories with similar features.

This category presents attractive features for our study. It includes a group of relatively homogeneous products that can be characterized by a number of well-defined product attributes: country of origin, brand, bottle size, bottle shape and material, nipple size, nipple shape and material, and price point. The fact that feeding bottles have well-defined product characteristics makes it easier to identify competing or substitutable products, which plays a key role in the pricing decision. In addition, although there are innovations and new product launches in the baby-feeding-bottle category, the life cycle of the products is long compared to the time span that the product will be used. It is also the case that during the course of our analysis, there were no new product introductions or innovations. The baby-feeding-bottle category presents a relatively small number of brands and manufacturers that do not engage in exclusivity deals with retailers. This means all competing retailers can carry any product across different brands. Finally, most customers will not engage in repeated purchases in a short period of time (e.g., daily or weekly) since the product will outlast the baby's need. Therefore, it is unlikely consumers will anchor prices based on their purchase histories. Moreover, inter-temporal substitution is not a pressing concern, nor is stockpiling behavior, which may be present for other categories, such as toilet paper and laundry detergent.

Although the characteristics of the baby-feeding-bottle category make it very appealing for our purposes, it is important to note these features are not unique to this category. Many other product categories share similar characteristics to which our methodology also apply, such as small appliances, hardware tools, and kitchenware. Hence, the methodology we introduce can be used to create a broadly applicable pricing tool for a variety of product categories and retail settings.

3. Consumer Choice Model

In this section, we first present the general framework of our choice model, which describes how consumers make choices among substitutable products that multiple competing retailers offer. Then we discuss a strategy for estimating model parameters in the absence of competitors' sales data.

[2] Deloitte news release: Asia Pacific's Top 10 Fastest Growing Technology Firms for 2011. December 1, 2011.

3.1. Choice Model Framework

Facing a choice set of J products offered by R retailers, a consumer i obtains utility u_{ijrt} from purchasing product j at retailer r at price p_{jrt} on day t, or utility u_{i0t} by not purchasing from one of the R retailers, where

$$\begin{aligned} u_{ijrt} &= \alpha_j + \alpha_r + \beta_j \log p_{jrt} + \epsilon_{ijrt}, j = 1, 2, ..., J, , r = 1, 2, ..., R, \\ u_{i0t} &= X_{0t}\gamma + \epsilon_{i0t}. \end{aligned}$$

A consumer will purchase product j from retailer r at price p_{jrt} if $u_{ijrt} = \max\{u_{ijrt}, u_{i0t}; j = 1, 2, ..., J, r = 1, 2, ..., R\}$, and will choose the outside option if $u_{i0t} = \max\{u_{ijrt}, u_{i0t}; j = 1, 2, ..., J, r = 1, 2, ..., R\}$. The intercept α_j corresponds to the constant utility obtained from purchasing product j regardless of which retailer the product is purchased from. The intercept α_r is the additional utility obtained by purchasing a product from retailer r, which can be understood as a customer's retailer preference. For instance, a consumer would assign a higher utility to a retailer who offers more convenient online checkout, a reliable delivery program, or a lenient return policy. The higher the value of α_r, the larger the premium a customer is willing to pay to buy the product from retailer r. Note that only the *differences* across these retailers preferences are identifiable. Hence, we normalize $\alpha_r = 0$ for our retail partner. Product-specific price sensitivity is captured by the parameter β_j. We do not explicitly model shipping costs because all retailers offer a generous shipping policy in this context—free shipping for a small minimum spending per order (¥29 to ¥39), thanks to low labor costs in China.[3] However, in a context where shipping costs vary significantly across competitors and orders, one could include shipping cost sensitivity in the model.

The outside option in our model includes purchasing from other channels, including both online and brick-and-mortar retailers, and not purchasing at all. We allow the utility of the outside option to vary across days of the week and holiday periods as purchase intention could vary between weekdays and weekends or between holidays and "regular" days (Perdikaki et al. 2012, Lu et al. 2013). These covariates, i.e., day-of-week dummies and holiday dummies, are represented by X_{0t}. Finally, ϵ_{ijrt} represents consumer i's utility shock of purchasing product j at retailer r on day t. Distribution assumptions and correlation patterns of ϵ_{ijrt} will be discussed subsequently.

The majority of prior studies involving consumer choice models restrict the attention to a model where the parameter β is a constant that does not vary across products. In these models, the estimates of price sensitivity are driven primarily by demand and price variations *across* products. In our work, however, we conduct an experiment to introduce price variations *within* each product,

[3] The exchange rate of Chinese Yuan to US Dollar as of Aug 1, 2014 is 6.18 to 1.

thereby allowing us to measure the extent to which price sensitivities vary across products, and meanwhile addressing potential concerns of price endogeneity.

There are several arguments in the literature for why price sensitivity might vary by product. First, there are many examples of price premiums charged for products with higher expected quality. This suggests that either higher expected product quality reduces price sensitivity or that less-price-sensitive consumers are drawn to higher quality products (Erdem et al. 2002). Second, product uncertainty may affect price sensitivity. The direction of the effect can go both ways. When consumers are uncertain about product quality, they may use price as a signal and thus exhibit lower price sensitivity (Gaur and Fisher 2005). On the other hand, if consumers are risk averse, they may derive greater disutility from a given price, thus inducing higher price sensitivity for uncertain products (Tellis and Gaeth 1990). Lastly, availability of alternative choices may lead to higher price sensitivity (Nelson 1974). Hence, products offered at more venues may exhibit higher price sensitivity, and popular products may exhibit higher price elasticity than niche products.

An alternative to letting price elasticity vary by product is to let it vary by customers through a random coefficient model, where the price coefficient $\widetilde{\beta}_i$ is consumer specific and is drawn from a distribution whose parameters are to be estimated. The advantage of this model is that it explicitly incorporates consumer heterogeneity. However, how price elasticity varies across products is dictated by product specific intercepts (see Train 2009, Chapter 6, for details). In contrast, the model with product-specific price elasticity allows for greater degrees of freedom and is more sensitive to demand and price variations associated with each specific product; as we shall see in the estimation results, price elasticities vary significantly across products. Of course, achieving this requires greater price variations within each product to retain the statistical power, which we obtain thanks to our randomized price experiment. A more ambitious model may incorporate both consumer heterogeneity and product specificity at the same time. However, such a model requires estimating at least J random coefficient distributions (both means and standard deviations), which suffers from over-fitting issues when applied to a relatively sparse experimental data set.

3.2. Extent of Price Comparison

The utility shocks ϵ_{ijrt} are not completely independent of each other because the R options associated with a particular product j are essentially the same product. Even though the purchasing utility could vary depending on the retailer's platform, the consumption utility associated with these products is the same. Consequently, it is reasonable to assume that a consumer who likes product j at retailer r should also like the same product offered by other retailers. In other words, the utility shocks ϵ_{ijrt} for a product are likely correlated across retailers. To allow such correlation, we assume the utility shocks $\epsilon_{it} = (\epsilon_{i0t}, \epsilon_{ijrt}; j = 1, 2, ..., J, r = 1, 2, ..., R)$ have a cumulative distribution given by:

$$\exp\Big(- e^{-\epsilon i0t} - \sum_{j=1}^{J} \Big(\sum_{r=1}^{R} e^{-\frac{\epsilon_{ijrt}}{1-\lambda}} \Big)^{1-\lambda} \Big).$$

Under such a joint distribution, the marginal distribution of each utility shock ϵ_{ijrt} follows a univariate extreme value distribution. In other words, our model establishes a nested structure where each product is a nest. The parameter λ can be intuitively understood as an indicator of correlation for utility shocks for the same product offered by different retailers. As λ increases, the correlation increases.[4] A value of $\lambda = 0$ indicates no correlation, and the model reduces to a standard multinomial logit model. As the value of λ approaches 1, utility shocks approach perfect correlation, which means that all idiosyncratic utility shocks associated with the same product are identical across retailers. In this case, every consumer will always buy from the retailer that offers the lowest price (assuming for a moment that retailer preferences α_r are identical). That is, consumers engage in most extensive price comparison when deciding where to buy a product. The larger λ, the more likely prices (or prices normalized by channel quality α_r) will drive retailer choice. The smaller λ, the more likely the choice of retailers will be proportional to their market share according to the independence of irrelevant alternatives (IIA) property, which asserts that the ratio of probabilities of choosing two alternatives is independent of the availability or attributes of a third option. For this reason, the larger λ, the more concerned retailers should be about monitoring and following competitors' price movements. Under this proposed model, the probability of purchasing product j from retailer r on day t can be written as follows (Train 2009):

$$Pr_{jrt} = \frac{\exp\big(\frac{\alpha_j + \alpha_r + \beta_j \log p_{jrt}}{1-\lambda}\big) \Big(\sum_{s=1}^{R} \exp\big(\frac{\alpha_j + \alpha_s + \beta_j \log p_{jrt}}{1-\lambda}\big) \Big)^{-\lambda}}{\exp(X_{0t}\gamma) + \sum_{j=1}^{J} \Big(\sum_{s=1}^{R} \exp\big(\frac{\alpha_j + \alpha_s + \beta_j \log p_{jrt}}{1-\lambda}\big) \Big)^{1-\lambda}}. \tag{3.1}$$

3.3. Stockouts as a Source of Identification

The model discussed so far could be estimated using a standard nested logit framework (Train 2009) if we were able to observe competitors' sales, which, unfortunately, we are not. However, we do observe the assortment competitors carry as well as their prices and availability. In what follows, we first illustrate the identification challenges that incomplete sales information poses. Then we show how own and competitors' stockout occasions serve as a useful source to identify retailer preferences and the extent of price comparison. We will use a moment condition framework to illustrate how stockouts provide the information needed for identification; however, the model is estimated using

[4] λ is not exactly equal to the correlation, but it can be used as a proxy for it.

maximum likelihood estimator consistent with what is commonly used for the family of logit models. Moment estimators and maximum likelihood estimators yield similar results (Berry et al. 1995). When sales levels are low, however, such that market shares in specific instances are not good proxies for theoretical probabilities, the maximum likelihood estimator is more efficient.

For illustration, we use a simple case where there are only two products, 1 and 2, and two competing retailers, A and B (normalize $\alpha_A = 0$). We also assume for now that all utility shocks are independent, i.e., $\lambda = 0$. For simplicity, we drop subscript t and remove the covariates matrix X_{0t} and assume that the mean utility of the outside option equals zero. Thus, the model reduces to a standard multinomial logit model where:

$$\begin{aligned}
u_{i1A} &= \alpha_1 + \beta_1 \log p_{1A} + \epsilon_{i1A}, \\
u_{i2A} &= \alpha_2 + \beta_2 \log p_{2A} + \epsilon_{i2A}, \\
u_{i1B} &= \alpha_1 + \alpha_B + \beta_1 \log p_{1B} + \epsilon_{i1A}, \\
u_{i2B} &= \alpha_2 + \alpha_B + \beta_2 \log p_{2B} + \epsilon_{i2A}, \\
u_{i0} &= \epsilon_{i0}.
\end{aligned}$$

Suppose we only observe retailer A's sales data. Given market size M and sales y_{1A}, y_{2A}, one can calculate the market share of product 1 at retailer A, $s_{1A} = \frac{y_{1A}}{M}$, and the market share of product 2 at retailer A, $s_{2A} = \frac{y_{2A}}{M}$. We then let market shares be equal to purchase probabilities obtained through Equation 3.1. By taking the ratios of these market shares, we obtain the following two moment conditions:

$$\frac{s_{1A}}{s_{2A}} = \frac{\exp(\alpha_1 + \beta_1 \log p_{1A})}{\exp(\alpha_2 + \beta_2 \log p_{2A})}, \tag{3.2}$$

$$\frac{s_{1A}}{1 - s_{1A} - s_{2A}} = \frac{\exp(\alpha_1 + \beta_1 \log p_{1A})}{1 + \exp(\alpha_1 + \alpha_B + \beta_1 \log p_{1B}) + \exp(\alpha_2 + \alpha_B + \beta_2 \log p_{2B})}. \tag{3.3}$$

By rewriting Equation 3.2, we have

$$\log\left(\frac{s_{1A}}{s_{2A}}\right) = \alpha_1 - \alpha_2 + \beta_1 \log p_{1A} - \beta_2 \log p_{2A}. \tag{3.4}$$

From Equation 3.4, we are able to identify three (sets of) parameters: $\alpha_1 - \alpha_2$, β_1, β_2, where β_1 and β_2 are identified through variations of two products' prices, p_{1A} and p_{2A}. Recall that we are able to identify price coefficients without bias because prices p_{1A}, p_{2A} are randomly assigned in our experiment and, more importantly, assigned in a way that allows us to identify two separate price sensitivities without encountering multi-collinearity among these two price series. Unfortunately, we are unable to separate the intercepts α_1, α_2 from this moment condition.

Since β_1, β_2 can be identified from Equation 3.4, we are left with *three* parameters, $\alpha_1, \alpha_2, \alpha_B$, that need to be estimated and we have only *two* moment conditions given by Equations 3.2 and 3.3. We would need at least one more moment condition to fully identify all three parameters. In what follows, we show how stockouts provide this additional moment condition. Suppose that on a different day, product 1 stocks out at Competitor B. We then have the following moment condition with regard to the newly observed market shares of product 1 and 2 by retailer A, s'_{1A}, s'_{2A}:

$$\frac{s'_{1A}}{1-s'_{1A}-s'_{2A}} = \frac{\exp(\alpha_1+\beta_1\log p_{1A})}{1+\exp(\alpha_2+\alpha_B+\beta_2\log p_{2B})}. \tag{3.5}$$

We now have three moment conditions 3.2, 3.3 and 3.5, and exactly three parameters, $\alpha_1, \alpha_2, \alpha_B$, to identify. We then demonstrate how the extent of price comparison λ is identified. We again illustrate the logic with an example employing two products and two retailers, but now utility shocks can be correlated across retailers selling the same product and the correlation is parameterized by λ. Consider again two days: on the first day all products are in stock at both retailers, and on the second day product 1 stocks out at Competitor B. Market shares on the first day are denoted as s_{1A}, s_{2A}, and markets shares on the second day are denoted as s'_{1A}, s'_{2A}. We have:

$$\frac{s_{1A}}{s_{2A}} = \frac{\exp\left(\frac{V_{1A}}{1-\lambda}\right)\left(\exp\left(\frac{V_{1A}}{1-\lambda}\right)+\exp\left(\frac{V_{1B}}{1-\lambda}\right)\right)^{-\lambda}}{\exp\left(\frac{V_{2A}}{1-\lambda}\right)\left(\exp\left(\frac{V_{2A}}{1-\lambda}\right)+\exp\left(\frac{V_{2B}}{1-\lambda}\right)\right)^{-\lambda}},$$

$$\frac{s_{1A}}{1-s_{1A}-s_{2A}} = \frac{\exp\left(\frac{V_{1A}}{1-\lambda}\right)\left(\exp\left(\frac{V_{1A}}{1-\lambda}\right)+\exp\left(\frac{V_{1B}}{1-\lambda}\right)\right)^{-\lambda}}{1+\sum_{j=1,2}\exp\left(\frac{V_{jB}}{1-\lambda}\right)\left(\exp\left(\frac{V_{jA}}{1-\lambda}\right)+\exp\left(\frac{V_{jB}}{1-\lambda}\right)\right)^{-\lambda}},$$

$$\frac{s'_{1A}}{s'_{2A}} = \frac{\exp\left(V_{1A}\right)}{\exp\left(\frac{V_{2A}}{1-\lambda}\right)\left(\exp\left(\frac{V_{2A}}{1-\lambda}\right)+\exp\left(\frac{V_{2B}}{1-\lambda}\right)\right)^{-\lambda}},$$

$$\frac{s'_{1A}}{1-s'_{1A}-s'_{2A}} = \frac{\exp\left(V_{1A}\right)}{1+\exp\left(\frac{V_{2B}}{1-\lambda}\right)\left(\exp\left(\frac{V_{2A}}{1-\lambda}\right)+\exp\left(\frac{V_{2B}}{1-\lambda}\right)\right)^{-\lambda}},$$

where $V_{jr} = \alpha_j + \alpha_r + \beta_j \log p_{jr}, j = 1,2, r = A,B$. Again, the identification of $\beta_j, j = 1,2$ comes from the variation in prices $p_{jA}, j = 1,2$. These four moment conditions allow us to identify four parameters $\alpha_1, \alpha_2, \alpha_B, \lambda$.

It is worthwhile to understand the intuition in addition to the mathematical derivation. When a competitor stocks out, the extent to which the retailer's total sales increase provides information about the competitor's market share, or consumers' preference for the competing retailer α_r. This is because the higher the competitor's market share, the more overflow we would observe when this competitor stocks out. Additionally, if more of the unsatisfied demand spills over to the same product that stocks out at the competitor, rather than to other products, it suggests tastes are strongly correlated for the same product offered by different retailers, and that more customers shop across retailers. That is, when a product stocks out at a competitor, customers originally interested in the product will not simply buy a different product, but instead will be more likely to buy the same product from other retailers.

To summarize, let $z_{jrt} = 1$ denote that product j is in stock at retailer r on day t, and $z_{jrt} = 0$ otherwise. We can rewrite the probability of purchasing product j at retailer r on day t:

$$Pr_{jrt} = \frac{z_{jrt}\exp\left(\frac{\alpha_j+\alpha_r+\beta_j\log p_{jrt}}{1-\lambda}\right)\left(\sum_{s=1}^{R} z_{js}\exp\left(\frac{\alpha_j+\alpha_s+\beta_j\log p_{jrt}}{1-\lambda}\right)\right)^{-\lambda}}{\exp(X_0\gamma)+\sum_{r=1}^{R}\sum_{j=1}^{J} z_{jrt}\exp\left(\frac{\alpha_j+\alpha_r+\beta_j\log p_{jrt}}{1-\lambda}\right)\left(\sum_{s=1}^{R} z_{jst}\exp\left(\frac{\alpha_j+\alpha_s+\beta_j\log p_{jrt}}{1-\lambda}\right)\right)^{-\lambda}}. \tag{3.6}$$

Comparison with Price Variations as a Source of Identification. Using logic similar to that above, one could prove that price variations at competitors would also lend additional moment conditions for identification, which is consistent with Newman et al. (2013). However, this identification strategy can sometimes be weak, especially if price variations are low, due either to limited self-price variations or highly correlated prices across competitors. It is because under these conditions, besides availability, price is the only time-variant product feature that is present in this context for identification of retailer preferences and extent of price comparison, and the same variation is also used to identify other parameters, specifically, price elasticities β_1, β_2. All the other product features such as size, brand, or shape, are time invariant.

It is important to note that although prices and availability are both time-variant product features, they enable identification in different ways. While price variation affects only the utility levels obtained from each option, availability variation alters the composition of the choice set by changing what the set offers. The latter generates sizable changes in the choices available to consumers, and this is especially valuable when market shares are small. On the other hand, price variations usually do not affect choices as significantly as availability variations, and price variations are also used to identify other parameters as well, as previously noted.

Using simulation, we compare parameter estimation accuracy across 36 scenarios with varying levels of price variation, price correlation between the retailer and the market, and stockout probabilities. Details of the simulation are discussed in the electronic companion EC.2. We find that if price variation is low and the retailer's price is highly correlated with that of competition, then ignoring stockouts leads to biased and imprecise parameter estimates.

In a nutshell, stockouts allow us to exploit temporal changes in the consumer choice set. Conlon and Mortimer (2013) also use stockouts as for identification but with complete information on choice decisions. As we have shown in this section, this source of variation is even more critical with partially observed choice decisions.

Other Considerations for Identification. First, the identification depends on the exogeneity of stockouts. If we observe correlation between stockouts of a product and the unobserved utility shock of this product, the exogeneity condition would be violated. This would be the case, for example, if a product has been popular for some time and has sold out as a result of this popularity. If utility shocks are serially correlated, then the utility shock of the product is likely to be high during the stockout as well. We test this by examining the demand patterns of each product for autocorrelation and find no evidence of such.[5]

Second, it is possible that retailers adjust prices of other available items when they experience stockouts for some products. If the choice model explicitly accounts for competitors' prices of all major products, like we do, such correlation is not a threat to identification.

Third, in contexts where a non-negligible number of customers choose to wait when experiencing a stockout—for example, when the iPhone was first introduced— stockouts can still be used as an identification strategy, but the model needs to explicitly account for the waiting decision. In the current model, if a consumer experiences a stockout at a retailer and decides to wait until the product becomes available, the model will count the consumer as choosing the outside option and treat the individual as a new customer when he or she comes back. In our setting, this is not a major concern because the channels and products are not substantially differentiated, unlike iPhones for example. This means it is unlikely customers will wait to buy a product at a specific retailer when the exact same product is available for a comparable price at a competing retailer.

4. Price Elasticity Field Experiment

As noted previously, a challenge that practitioners and academics encounter in identifying the relationship between prices and demand data is the endogeneity that pricing decisions generate. This endogeneity challenge holds when the estimation is done from observational data, where prices are not randomly defined and managers make pricing decisions based on demand information observed at that time. Unfortunately, in most situations, researchers cannot observe this information.

In the past, researchers have used instrumental variables to correct for this bias. The two most commonly used instruments in settings similar to ours are Hausman-style instruments and lagged prices. Although theoretically appealing, these approaches can have significant limitations in practice. First, Hausman-style instruments use the average price of the same product sold in other geographical markets as the instrument for the price in the focal market, as prices in different locations are subject to common firm-level cost shifters but are not correlated with demand in

[5] Specifically, using Durbin-Watson statistics, we test for serial correlation for each product during and two months prior to the experiment period. All Durbin-Watson statistics are above the commonly used lower critical value 1, except for one at 0.984. Thirteen of the fifteen are within the range of 1.5 to 2.5, while 2 means zero autocorrelation (the other two are 0.984 and 1.431). These results suggest serial correlation is not a pressing concern.

the focal market (Hausman 1996, Nevo 2001). The effectiveness of this approach depends on (1) whether the product is sold at multiple locations and (2) sufficient correlation between prices and common cost shifters. These assumptions may not be valid in practice. For example, if product assortment varies across locations, Hausman-style instruments may not be present for all products considered. Another limitation can be present for products with prices that change frequently—daily or sometimes multiple times a day. Cost parameters are only weak instruments in this case, since costs do not change nearly as frequently as prices do.

Similarly, to use lagged prices as an instrument we need sufficient correlation between lagged and current prices, which is often the case (Tereyagoglu et al. 2014 and Villas-Boas and Winer 1999). However, we also need sufficient time-series variations in prices such that lagged and current prices are not highly collinear. For those products with prices that hold relatively constant during the period of study, lagged and current prices are almost always the same. Hence, the validity of lagged prices as instruments is challenged. For products with little historical price variation, it is difficult to estimate product-level price elasticity regardless of what instrument is used. In Section 5, we will illustrate the limitation discussed above by comparing the results from different instrumental variable approaches with our experimental approach.

4.1. Field Experiment Design and Implementation

With the support of our retail partner, we implement a randomized experiment to resolve the endogeneity issue. The field experiment consists of randomly varying prices for the top 15 selling products in the baby-feeding-bottle category during a four-week period. These 15 products represent 7 percent of all the products in the category and capture 54 percent of the unit sales and 55 percent of the revenue for the category during an average week before the test was implemented. Table 1 presents summary statistics of these 15 products before the test implementation. Based on the standard deviations and the means, we note that the average within-product price c.v. (coefficient of variance) is 5.9%, while the highest price c.v. is 13.8% and the lowest is 1.2%.

We focus on the 15 top sellers for two reasons. First, to estimate the impact of the randomized test, we need to focus on products that have a daily sales rate large enough to allow us to obtain statistically significant results. Second, this approach aligns with our ultimate goal of developing a robust methodology to simultaneously maximize total revenue for a group of products. Hence, we focus on the products within this category that have the largest impact on the category revenue.

When estimating the price elasticity of each product, we are concerned about different drivers that potentially can affect customers' sensitivity to prices: the absolute price level and the relative gap between the retailer's price and competitors' prices. The assumption is that price elasticity depends on two main factors: absolute price variations and price variations with respect to the

Table 1 Summary Statistics for the 15 Products in the Field Test

PRODUCT	DAILY SALES (units)				PRICE (¥)				MARKET PRICE (¥)
#	Average	Max.	Min.	SD	Average	Max.	Min.	SD	Average
1	29.0	63.0	0.0	14.5	35.6	40.4	33.9	1.5	40.5
2	15.2	33.0	0.0	8.6	32.2	35.9	31.9	0.7	35.1
3	2.9	15.0	0.0	2.8	40.6	45.0	37.5	2.6	48.6
4	1.8	6.0	0.0	1.4	44.9	48.0	40.0	1.5	48.7
5	3.5	12.0	0.0	2.2	89.0	109.8	80.0	3.9	91.3
6	4.3	13.0	0.0	3.4	88.3	102.0	84.0	2.9	91.2
7	4.3	18.0	0.0	3.6	76.1	82.0	74.3	0.9	79.2
8	2.2	8.0	0.0	2.1	66.2	84.0	61.2	6.6	81.2
9	9.6	27.0	0.0	7.7	91.7	99.0	82.4	6.5	92.8
10	3.7	9.0	0.0	2.5	14.4	16.7	13.9	0.6	20.9
11	17.8	36.0	0.0	7.2	84.9	108.0	78.9	6.3	100.0
12	22.5	65.0	0.0	18.4	86.0	175.0	83.2	11.9	83.9
13	18.2	58.0	0.0	15.4	85.5	109.0	84.0	5.4	84.0
14	3.9	18.0	0.0	4.3	121.5	134.0	107.0	10.2	121.0
15	2.3	27.0	0.0	5.5	130.7	145.7	118.5	11.7	111.0

These summary statistics correspond to the two-month period right before the field test was implemented.

competition. To deal with this situation, we implement the field test by randomly assigning the products being tested to two different groups.

The first group, Group A, randomly changes prices with respect to a baseline price defined by the product's historical price at the retailer, while the second group, Group B, randomly changes prices with respect to the competitors' prices for that product. We focus on the top four competitors that, together with our partner, capture more than eighty percent of the online B2C retail market in China. Once a product is randomly assigned to one of the two groups (variation with respect to the historical baseline or to the competition price), we randomly assign the treatment level for each of the products on each day, depending on the group to which the product belongs.

We define five different treatment levels: high (+10 percent), medium-high (+5 percent), medium (no variation), medium-low (-5 percent), and low(-10 percent). High and low are defined for each product based on its assigned group. For example, if a product belongs to Group A and is randomly assigned "medium-low" treatment, the product price is set 5 percent below the average price for that product over the eight weeks immediate prior to the beginning of the test. If a product belongs to Group B and is randomly assigned "high" treatment, the product's price is set 10 percent above the lowest competitor's price for that product during that day.

Before running the experiment, we analyzed historical data and simulated outcomes we may possibly obtain from a field test based on assumed parameters. A major consideration in this stage, both internally within the research team and when interacting with the retail partner, was the duration of the test. We estimate the minimum price variation and test duration needed for the field test to be effective. With these numbers, we ensure management that we were not asking for unnecessary efforts on their end. Another issue considered at this stage was the need for the company to keep all other factors constant during the testing period, for example, no deviations in promotion strategies. The combination of these factors led to a joint decision of 30-day duration and the price variation range that we considered in the test.

We randomly assign treatment using a fractional factorial design (Mead et al. 2012), taking into account the five different treatment levels. The treatment assigned to a product remained constant during a three-day period. Note that for products in Group B, the treatment is constant but the actual price could vary during the three-day period as competitors may change their prices. We choose to maintain the treatment level during a three-day period to avoid the risk of alerting customers that a test is in place and to minimize the chance of customers' speculative behaviors.

Each day at 9:00 a.m. local time during the four weeks of the test, we monitor competitor prices and adjust our own prices according to the randomized schedule. Additionally, we monitor competitive responses by scraping prices and product availability information from competitors' websites 12 hours after our retail partner changes prices each day. In particular, product availability information is obtained by visiting product detail pages at competitors' websites. When a product is out of stock, consumers will still be able to find it from a retailer's website but will see a label of "Out of Stock" or "Add to Wishlist" on the search results and product details pages.

4.2. Field Test Analysis and Results

The test lasted for a total of 30 consecutive days and the implementation occurred according to plan. As an illustration, Figure EC.1 presents a comparison between units sold and retail prices for two products included in our sample for the period before and after the test was implemented. It is interesting to observe that, while it is very hard to see any pattern between price and demand before the test was implemented (e.g., the top figure for Product 1), after random treatments are introduced, the price elasticity becomes evident by visual inspection (graphs at the bottom of the figure). In addition, the field test allows us to introduce price variation on a number of products that historically present a very stable price pattern (e.g., Product 2 in the figure). It is evident that if prices do not vary, it is almost impossible to estimate their impact on demand.

Of course, beyond the visual examination of the test results, we want to take full advantage of the randomized experiment. In the next section we estimate the customer choice model introduced previously. We then calculate optimal prices that account for, among other controls, the unbiased price elasticity estimates as well as the competition's prices and product availability.

5. Estimation Results

We estimate the demand model using the Maximum Likelihood Estimator which maximizes the overall probability of observing a certain number of purchases for all SKUs on all days during the 30-day test period. The exact formulation is expressed as follows:

$$\max_{\alpha,\beta,\gamma,\lambda} LL(\alpha,\beta,\gamma,\lambda) = \sum_{t=1}^{30}\left[\sum_{j=1}^{J} D_{jrt}\log Pr_{j1t} + \big(M_t - \sum_{j=1}^{J} D_{jt}\big) log\big(1-\sum_{j=1}^{J} Pr_{j1t}\big)\right].$$

Pr_{jrt} is defined in Equation 3.6, where $r = 1$ represents our partner retailer. D_{jt} is the observed sales for product j on day t at our partner retailer, and M_t is the market size on day t.[6]

Table 2 compares estimation results obtained from different types of data, observational vs. experimental, and different types of models, with and without instruments, with and without competition and price comparison. We apply a standard multinomial logit model, to the historical observational data, without instruments (Column 1) and with three instrumental variables (Column 2 to 4) commonly used in the literature: (1) SKU-level costs (including wholesale costs and operating costs); (2) Hausman-style instruments, i.e., average prices in other geographically disparate locations (Hausman 1996); and (3) lagged prices (Villas-Boas and Winer 1999, Tereyagoglu et al. 2014). Note that these models only account for prices and availability of substitutable products at our partner retailer do not yet include competitive information. We then apply the same model to the experimental data without using any instrument (Column 5), while in Column 6 we include competitive prices and product availability in the model, and in Column 7 we further include consumer extent of price comparison across retailers.

5.1. Observational Data vs. Randomized Price Experiment

As noted before, managers may adjust prices based on demand signals unobservable to us and this may cause an upward bias in price elasticity estimates. This is evident in Table 2; the elasticities of 5 of the 15 products turn out to be positive when simply applying a multinomial logit model to historical data without accounting for potential price endogeneity (column 1). Even though positive price elasticity can be found for conspicuous consumption, it would be hard to imagine this is the case for the product category under study (baby-feeding bottles).

The power of this instrumental variable approach critically depends on (1) the availability of a valid instrument; and (2) whether the instrument itself captures sufficient variations from the original variable to avoid insignificant estimates arising from weak instruments. In our context, we find the performance of some commonly used price instruments suffer these concerns.

First, when SKU-level costs are used as an instrument in our setting (Column 2), 6 out of the 15 elasticities turned out to be positive and all estimates are much noisier (as shown by large standard errors) than those in our base analysis without instruments (Column 1). This is because

[6] We test two approaches to approximate market size similar to what is commonly done in the literature (see Berry et al. 1995, for example). In the first approach we assume market size is constant, and equals the maximum of daily category web traffic during the period under study. In the second approach, the market size is allowed to vary by day, and is proportional to category web traffic observed at our retail partner. The estimation results are not sensitive to which approach we use. The results shown are based on the first approach.

Table 2 Estimates of Price Elasticities for Feeding Bottle SKUs

	Historical Data				Randomized Price Experiment		
	(1) No Comp Info	(2) No Comp Info IV Cost	(3) No Comp Info IV Hausman	(4) No Comp Info IV Lagged Price	(5) No Comp Info	(6) w/ Comp Info Ind. Demand Shocks	(7) w/ Comp Info Corr. Demand Shocks
Product Specific Price Elasticity (β_j)							
Product 1	0.974	-4.969	0.060	0.755	-5.496***	-5.438***	-1.675***
	(0.624)	(5.259)	(0.704)	(0.664)	(0.939)	(0.234)	(0.157)
Product 2	-2.451***	-10.453	2.328	-2.288***	-2.390**	-2.267**	-0.367***
	(0.519)	(6.639)	(2.157)	(0.549)	(1.129)	(1.004)	(0.014)
Product 3	-5.436***	7.190	-5.200**	-8.066***	-7.109***	-4.572***	-6.773***
	(1.043)	(13.222)	(2.561)	(1.399)	(1.715)	(0.636)	(0.643)
Product 4	-4.398**	-5.068	-1.511	-2.648	2.280	-0.006	-0.004
	(1.8216)	(5.0866)	(6.7204)	(2.1419)	(2.0565)	(0.025)	(0.001)
Product 5	-1.848**	-0.525	-3.193	-1.161	0.096	-0.004***	-0.953***
	(0.906)	(5.678)	(2.381)	(1.228)	(2.468)	(0.003)	(0.103)
Product 6	-1.410	-10.999	-1.779	-0.796	-4.028*	-2.367***	-1.054
	(1.213)	(10.129)	(2.797)	(1.261)	(2.332)	(0.523)	(0.727)
Product 7	-0.214	2.076	-3.764	-0.252	-1.534	-2.399***	-0.540
	(0.649)	(9.952)	(2.666)	(0.714)	(1.415)	(0.472)	(0.340)
Product 8	3.709	10.121***	N/A	5.768**	-4.222***	-2.964***	-1.164**
	(2.422)	(2.467)	N/A	(2.402)	(1.406)	(0.336)	(0.535)
Product 9	-3.713***	-11.545***	-5.500***	-3.587***	-4.450***	-4.296***	-1.118***
	(0.699)	(2.020)	(1.269)	(0.712)	(0.872)	(0.293)	(0.159)
Product 10	2.244	-0.684	-0.363	3.001*	-4.400	-4.845***	-4.149***
	(1.506)	(10.976)	(2.475)	(1.640)	(3.381)	(0.551)	(1.343)
Product 11	-8.225***	-9.944***	-7.994***	-8.279***	-4.324***	-4.058***	-0.504***
	(0.810)	(1.090)	(0.817)	(0.846)	(0.959)	(1.872)	(0.093)
Product 12	1.487**	2.151	-4.919***	-0.795	-17.149***	-3.049***	-2.187***
	(0.654)	(5.002)	(1.606)	(0.611)	(5.917)	(0.077)	(0.045)
Product 13	0.812	-21.701***	-26.739***	1.505	-26.061***	-4.515***	-11.281***
	(1.252)	(7.294)	(5.726)	(1.435)	(3.075)	(0.284)	(0.518)
Product 14	-3.632***	7.548	N/A	-3.438***	-2.110*	-2.525***	-0.922***
	(0.546)	(13.609)	N/A	(0.519)	(1.151)	(0.209)	(0.167)
Product 15	-2.914***	10.459	N/A	-2.821***	-3.460***	-3.848***	-1.142***
	(0.759)	(16.980)	N/A	(0.703)	(0.709)	(0.197)	(0.151)
Retailer Preferences							
Retail Partner						baseline	baseline
Competitor 1						0.260	0.217**
						(3.163)	(0.096)
Competitor 2						0.366	0.017
						(4.225)	(0.119)
Competitor 3						-3.750	-1.836***
						(10.683)	(0.045)
Competitor 4						-3.887***	-2.464***
						(1.457)	(0.010)
Extent of Price Comparison (λ)							0.7911***
							(0.0203)
Product Intercepts	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-Week & Holiday	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Days	67	67	67	67	30	30	30
# Purchases	7690	7690	7298	7580	3742	3742	3742
Pseudo R2	0.6696	0.6671	0.6749	0.6668	0.5834	0.7107	0.7164
MAD (daily)	0.439	0.436	0.496	0.449	0.451	0.387	0.377
Log Likelihood	-35973.844	-36650.168	-29056.259	-36681.365	-17281.855	-17053.73	-17036.01

Column 1-4 display estimates using historical data without and with instruments. Column 5-7 display estimates from the 30-day randomized price experiment. Column 2 uses cost as the instrument to price. Column 3 uses the Hausman style instrument: average prices of the same product sold at other geographical locations by the same company. Because certain products are not sold at either of the other locations or off shelf for most of the time during the study period (available for less than one week during two months), price elasticity for these products are not available. Column 4 uses one-day lagged price as instrument. In Columns 1-6, $elasticity_{jA} = \beta_j(1 - s_{jA})$, *and* $\approx \beta_j$ *if* s_{jA} *is small. In Column 7,* $elasticity_{jA} = \frac{\beta_j}{1-\lambda}(1 - s_{jA}) - \frac{\lambda}{1-\lambda}\beta_j(s_{jA} + ns_{jA})$, *where* ns_{jA} *denotes the share of product j offered by Retailer A within the product nest, i.e., the set of product j's available at all retailers.*
*** $p < 0.01$, ** $p < 0,05$, * $p < 0.1$.

costs do not change nearly as fast as prices in this context (i.e., daily for some products) and for this reason, it turns out to be a weak instrument. Second, even though Hausman style instrument, average prices in other geographical locations (Column 3), has the best performance of all three instruments when available, it is subject to limitations as well. The instrument does not exist for some SKUs, because these SKUs are not offered at other locations or are off shelf for most of the time during the study period (available for less than a week during two months). Lastly, results from using lagged prices as instruments are almost identical to the results from using original prices. This is driven by the high correlation between prices and lagged prices. Recall that for some products, prices hardly varied during the period under study.

We now turn to the results based on the randomized price experiments presented in Column 5. We note that 13 of the 15 price elasticities are negative, and 10 of the 15 products present a lower price elasticity compared to the one obtained when using historical data. It is worthwhile noting that this is achieved with only half the sample size (30 days vs. 67 days). This analysis suggests that with the randomized experiment, we are able to correct, to a certain extent, the positive biases caused by price endogeneity.[7] In addition, the results not only show improvement over the historical data without instruments but also over historical data with commonly used instruments.

5.2. With vs. Without Competitor Information

We next focus our attention on how our proposed method, which accounts for competitor prices and availability (Column 6), performs relative to the standard multinomial logit model (Columns 1-4). It is important to emphasize that in our setting, price randomization per se is no guarantee for exogeneity. This is because if competitors respond by adjusting their prices based on the randomized prices, not accounting for competitors' actions would introduce correlation between our prices and the unobserved demand shocks through correlation with competitor prices. If competitor prices and our prices are positively correlated, then there will be a positive bias in the estimation of price sensitivities, β_j, unless we account for competitors' prices appropriately. To illustrate this, suppose competitors follow our prices closely and decrease their prices as we decrease ours; as a result, we will receive a lower demand shock. This generates a positive correlation between our price and our demand and hence biases price elasticity estimates upward.

To explore this issue, we study whether it is the case that, during the period of our experiment, the competitors react to our randomized changes by changing prices accordingly. We first note the percentage of time that daily prices changed from the previous day: 38.6% (our retail partner), 1.9% (Competitor 1), 16.3% (Competitor 2), 19.5% (Competitor 3), 11.1% (Competitor 4). Table EC.1 further shows the frequencies of competitor price responses within 12 and 24 hours after our price change. We see that competitors did not seem to follow closely our random price changes. The most responsive competitor, Competitor 2, followed only 11 of our 91 price increases and 9 of our 75 price decreases. This pattern is consistent with that in the pre-test period.

This result also indicates that during the one-month test period, the experiment successfully introduced random variations not only in our own price levels but also in the *relative* price levels

[7] For products with low sales levels such as Product 4 which is sold only 1 or 2 units per day, the estimated price coefficients tend to be insignificant. This indicates that more price variations are required to estimate price elasticities for slow moving products.

to competitors. Using the full choice model, we observe a significant improvement in the model fit: the R-square increased from 0.5848 (Column 5) to 0.7107 (Column 6), a 21.5 percent increase.[8]

5.3. Extent of Price Comparison

Finally, we compare the last two columns in Table 2. Column 7 introduces an additional parameter, λ, that captures the extent to which consumers conduct price comparisons. Recall that $\lambda = 0$ means utility shocks for the same products across retailers are independent, in which case columns 6 and 7 should be identical. The larger the value of λ, the higher the correlation of utility shocks for the same product across retailers and the more intense the price comparison is. The estimate of λ equals 0.7911 and significant with 99% confidence, which suggests a non-negligible correlation in utility shocks.[9] In other words, consumers engage in extensive price comparisons across competitors. The high intensity of price comparison indicates that retailers need to follow competitors' price adjustments closely to stay competitive in the market. The impact of either overpricing or underpricing can be significant. In fact, despite the small difference in overall model fit, columns 6 and 7 will suggest very different responses when competitors change prices, as we later discuss.

We now turn our attention to the estimate of competitor significance, i.e., α_r. Our retail partner is the baseline. Competitor 1 admits the highest preference, followed by Competitor 2, our retail partner, and Competitors 3 and 4 respectively. This is consistent with management knowledge of the market.[10] Except for Competitor 2, all estimates are statistically significant (as compared to the baseline). All these retailers sell directly to customers. We do not account for prices charged by small individual sellers who sell through online platforms, because no single seller sells enough to be considered a major competitor and these small sellers either do not engage in dynamic pricing or merely follow prices.

We examine whether it is likely that the statistical non-significance of some α_r can be caused by low stockout probabilities. We summarize the stockouts percentages during the experiment period for all retailers in Table EC.2, together with the final estimates of the competitor significance (α_r). The table also presents within product price variation, between product price variation,

[8] We also conduct an out-of-sample test to make sure we are not over-fitting the data. We divide the 30-day test window into two periods: first 20 days in sample and the remaining 10 days out of sample. The out-of-sample R-squares are almost identical to the R-squares presented in the table, indicating no evidence of over-fitting. The in-sample estimates are consistent as well. We present the full sample estimates and use them as inputs for subsequent optimization analysis because they are less noisy due to a larger sample size.

[9] Based on numerical results from Koppelman and Wen (2000), a 0.79 similarity score implies a correlation of 0.91 to 0.95.

[10] The estimated percentage market shares of our retail partner, Competitor 1, Competitor 2, Competitor 3, and Competitor 4 in the online retail market (B2C and C2C combined) in 2013 are 1.52%, 14.0%, 2.34%, 1.2%, and 3.39%, respectively. Source omitted to protect identity of the retail partner and the key competitors. Competitor 1 is commonly perceived as the market leader in many categories, while Competitor 4 mostly specializes in home appliances and consumer electronics.

and percentage of time daily price changes from the previous day during the period of analysis. Indeed, there are two retailers, Competitor 2 and 3, which did not experience any stockouts during the experiment period. This does not necessarily lead to non-significant estimates of competitor importance (α_r), though. For example, Competitor 3's estimate is significant. This is likely because price variations at competitors would also lend additional moment conditions for identification, as we discussed in the identification section. For Competitor 2, which has zero stockouts and low within-product price variation during the experiment period, the estimated competitor significance (α_r) seems non-significant. However, the point estimate is indeed very close to zero (compared to the other points estimates), which suggests that it is likely that the competitor 2 could be indeed close to our partner retailer. Nevertheless, it would be better if competitors experienced more stockouts in the month during our experiment.

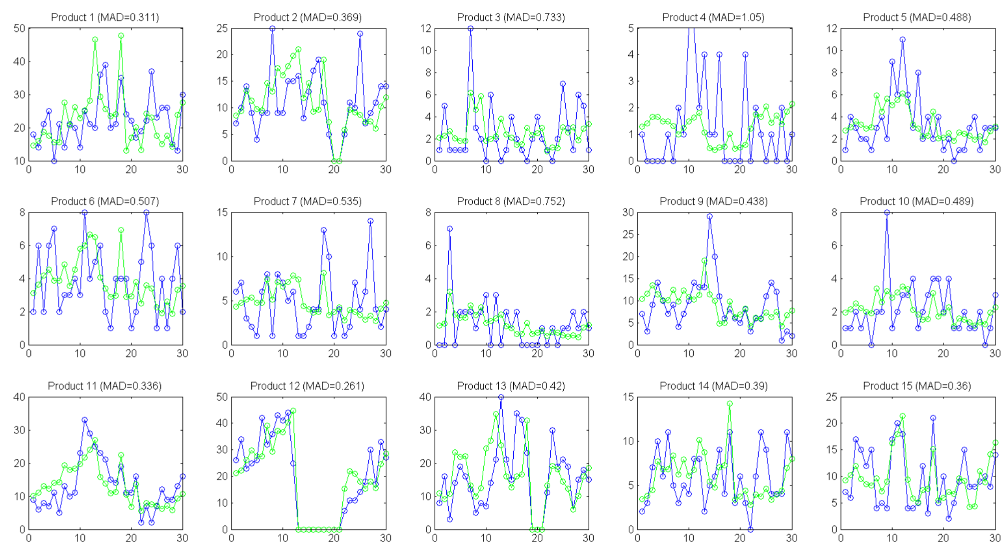


Figure 1 Model Goodness of Fit

Figure 1 shows the goodness of fit graphically. It plots the predicted daily sales (in green) against observed daily sales (in blue) based on our estimation results in Column 7. The average daily mean absolute deviation (MAD) is 0.377. Note there is a negative correlation of -0.697 (sig= 0.0041) between model goodness of fit by product, as measured by MAD, with average daily sales. In other words, the model better predicts demand for fast-moving products than for slow-moving products.

Based on estimated model parameters, we calculate sample own and cross-price elasticities as shown in Table 3. Note that own price elasticity differs by product due to differences in price sensitivities measured by β_j. Also note that our demand is most sensitive to prices of Competitors 1 and 2 but not to prices of Competitors 3 and 4. This is because Competitors 3 and 4 are relatively small players in this category, as revealed by the estimated retailer preference α_r. However, retailer preference α_r alone does not completely explain the difference in cross-price elasticity across retailers. Note that sometimes our demand is more sensitive to prices of Competitor 2 than those of Competitor 1, even though Competitor 2 is less preferred than Competitor 1 ($\alpha_2 < \alpha_1$). This is because for some products, Competitor 2 actually charges a lower price than Competitor 1 and thus increases its competitiveness in the market. In sum, which competitors to follow closely not only depends on the average retailer preference but also competitors' actual price levels for the SKU being considered.

Comparing results obtained from typical reduced-form linear demand models and our results, we note that, thanks to the randomized price experiments, we are able to obtain accurate measures of price elasticities with simple linear demand models. However, even though we may be able to randomize our own prices, we cannot randomize competitors' prices. Therefore, estimating the effect of competitors' prices on our demand non-parametrically with a reduced-form regression model can be challenging, especially with a short sample period (i.e., 30 days). This is why modeling consumers' choices of competing retailers becomes appealing. See EC.3 in the electronic companion for a detailed discussion.

Table 3 Own and Cross-Price Elasticities

Product #	Own	Competitor 1	Competitor 2	Competitor 3	Competitor 4
1	-5.5378	-1.2071	-2.8775	-0.0055	-0.0001
2	-1.7681	-0.7598	-0.6386	-0.0012	0.0000
3	-5.4942	-0.0018	-0.0095	-0.0120	-0.0001
4	-0.0046	-0.0093	-0.0069	0.0000	0.0000
5	-1.5826	-0.4744	-0.7552	-0.0013	0.0000
6	-2.5504	-0.7253	-1.2292	-0.0020	-0.0001
7	-0.9213	-0.4088	-0.3209	-0.0006	0.0000
8	-3.6766	-1.8118	-1.0456	-0.0068	0.0000
9	-3.4141	-0.8532	-1.7617	-0.0023	-0.0001
10	-1.8954	-0.0883	-0.0164	-0.0069	0.0000
11	-2.4377	-0.9699	-0.9174	-0.0023	-0.0001
12	-8.2826	-1.5770	-4.9116	-0.0064	0.0000
13	-23.6245	-0.0152	-14.2382	-0.0138	-0.0022
14	-3.3974	-1.6779	-0.9875	-0.0051	-0.0001
15	-4.1404	-1.3791	-1.6345	-0.0094	-0.0001

We note as well that estimated price elasticities indeed vary across products. Given that we have only 15 SKUs, a small sample size, regressing estimated price elasticities on product features will not yield statistically significant results. However, we make the following interesting observation on the difference of price elasticities of domestic and imported brands. Among the 15 SKUs, we have five different brands: Pigeon (four SKUs, domestic), Love (one SKU, domestic), NUK (six SKUs, origin Germany), Avent (three SKUs, origin U.K.), and Brown (one SKU, origin U.S.). We find that the three SKUs with the most elastic demand are all domestic products (all Pigeon, price levels of ¥84, ¥84, and ¥36). We also note that the three SKUs with least elastic demand are all imported brands (two NUK and one Avent, with price levels of ¥90, ¥76, and ¥44). That is, consumers are less elastic to imported brands than to domestic brands, given similar prices. There can be several possible explanations. Imported brands are typically perceived as high quality in China (Ozer et al. 2014), particularly for baby-and-mom products given the recent scandals of tainted infant milk in China.[11] The observation of lower elasticity on prices of imported brands can therefore be explained by higher expected product quality reducing price sensitivity or less-price-sensitive consumers being drawn to higher quality products such as imported brands (Erdem et al. 2002). It may also be explained by the fact that more channels sell domestic brands than imported brands, as imported brands are typically distributed only through online channels.[12] Products offered at more venues exhibit higher price elasticity (Nelson 1974).

In the next section, we use the estimates obtained from the choice model presented in Column 7 of Table 2 to determine the best response prices for each product.

6. Best-Response Pricing

Based on the estimates obtained from the choice model and taking into account competitor prices and product availability, we optimize prices for our retail partner to maximize total revenue for the category subject to several constraints imposed by the retailer. The constraints include a lower bound on average category margin, lower and upper bounds on individual product margins, and manufacturer price restrictions. Changes in recommended prices come from four different sources: (1) changes in costs, (2) changes in the retailer's own product availability, (3) changes in competitor prices, and (4) changes in competitor product availability. On a particular day t, we solve the following constrained nonlinear optimization problem:

$$\begin{aligned}
\max_{p_{1t},p_{2t},\ldots,p_{Jt}} \quad & \sum_{j=1}^{J} p_{tj} s_{jt}(p_{jt}, z_{jt}; p_{-jt}, z_{-jt}; \mathbf{p_{ct}}, \mathbf{z_{ct}}; \alpha, \beta, \gamma, \lambda) \\
s.t. \quad & \frac{\sum_{j=1}^{J} (p_{jt} - c_{jt}) s_{jt}(p_{jt}, z_{jt}; p_{-jt}, z_{-jt}; \mathbf{p_{ct}}, \mathbf{z_{ct}}; \alpha, \beta, \gamma, \lambda)}{\sum_{j=1}^{J} p_{jt} s_{jt}(p_{jt}, z_{jt}; p_{-jt}, z_{-jt}; \mathbf{p_{ct}}, \mathbf{z_{ct}}; \alpha, \beta, \gamma, \lambda)} \geq \textit{margin target}, \\
& \textit{margin } LB_j \leq \frac{p_{jt} - c_{jt}}{p_{jt}} \leq \textit{margin } UB_j, \forall j, \\
& LB_j \leq p_{jt} \leq UB_j, \forall j,
\end{aligned}$$

[11] Food Safety Is Crucial in China Deal for Baby Milk. *The New York Times.* August 27, 2014.

[12] According to Nielson China eCommerce Report in September 2013, imported feeding bottle brands are mostly distributed online rather than off-line. For example, among the top brands Pigeon, Brown, NUK, and Avent for the online channels, the only one that made the top 10 list for off-line channels is the domestic brand Pigeon.

where $s_{jt}(p_{jt}, z_{jt}; p_{-jt}, z_{-jt}; \mathbf{p_{ct}}, \mathbf{z_{ct}}; \alpha, \beta, \gamma, \lambda)$ denote the market share of product j at our retail partner, as a function of its own price and availability p_{jt}, z_{jt}, prices and availability of the other substitutable products at our retail partner p_{-jt}, z_{-jt}, prices and availability of all products at competitors $\mathbf{p_{ct}}, \mathbf{z_{ct}}$, and all the estimated parameters $\alpha, \beta, \gamma, \lambda$. The market share can be calculated using Equation 3.6.

Hanson and Martin (1996) show that logit profit functions are not concave in prices. One can easily show that our constrained optimization problem is also non-concave in prices. Therefore, in order to find the optimal solution, we seed 10 different initial values using the MATLAB optimization procedure for constrained optimization and choose the one that yields the highest revenue. The optimization procedure used to choose the optimal prices for a given day takes approximately 30 seconds. We correlated recommended SKU margins with SKU-specific price elasticities and found the correlation to be 52 to 57% on different days, suggesting that our algorithm indeed recommends a higher price (relative to cost) when the product's demand is less price elastic (note less elastic means higher values), which is intuitive.

Table 4 Best-Response Prices under Different Models

	Competitor 1 Price		**Suggested Own Price Ind. Demand Shocks**		**Suggested Own Price Corr. Demand Shocks**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Product #	**Day 1**	**Day 2**	**Day 1**	**Day 2**	**Day 1**	**Day 2**
1	42.8	42.8	38.1	37.4	39.1	39.2
2	36.8	36.8	38.6	38.4	37.5	36.8
3	48.8	48.8	46.8	47.8	45.0	45.0
4	48.8	48.8	49.8	48.3	48.3	48.5
5	108.0	108.0	87.7	87.8	87.5	86.3
6	108.0	108.0	87.4	87.3	89.3	87.9
7	79.0	79.0	81.7	81.5	82.6	82.6
8	85.0	85.0	72.3	72.3	74.4	74.2
9	108.0	108.0	84.4	84.4	87.9	86.6
10	18.0	18.0	17.0	16.7	18.9	18.9
11	108.0	108.0	100.6	100.6	101.5	101.5
12	98.0	98.0	79.1	80.3	81.0	81.4
13	98.0	98.0	80.9	81.3	83.2	83.4
14	**105.0**	**95.0 (↓10.0)**	**98.8**	**97.8 (↓1.0)**	**103.1**	**97.5 (↓5.6)**
15	**109.0**	**119.0 (↑10.0)**	**104.7**	**105.9 (↑1.2)**	**99.3**	**103.2 (↑3.9)**

In this table, we illustrate our reactions in a real scenario where Competitor 1 reduces the price for product 14 by ¥10, while raising the price for product 15 by ¥10 from day 1 to day 2. All other competitors' prices remain unchanged. We demonstrate that using the demand model with independent demand shocks versus the demand model with correlated demand shocks yields very different price reactions. With independent demand shocks, the algorithm suggests only a small change in our retail partner's price in response: a decrease of only ¥1.0 for product 14 and an increase of ¥1.2 for product 15. With correlated demand shocks, however, the algorithm suggests more sizable changes in prices: a decrease of ¥5.6 for product 14 and an increase of ¥3.9 for product 15.

We then demonstrate the importance of considering correlated demand shocks within a product nest. Table 4 shows how two models, with independent and correlated demand shocks across retailers, lead to different responses when competitors change prices. Note that in this table, inputs to the optimization model (costs, margins and bounds) are altered for confidentiality reasons. However, the structure of the optimization itself and demand parameters previously estimated

$(\alpha, \beta, \gamma, \lambda)$ are intact. With independent demand shocks, prices are driven primarily by costs and price elasticities that are product specific. When competitors change prices, lost demand will be allocated to all available alternatives—this means all products at all retailers—proportionally based on their current market shares. Since there is a large number of options, the spillover effect to the same product at a different retailer is very small and for this reason the suggested price change is very small (e.g., when competitors' prices change by ¥10, we only change the price of the same product by ¥1 as shown in Column 4 in Table 4).

Substitution patterns are key to responsive pricing. At one extreme, if customers do not substitute across retailers, there is no need to follow competitors' prices. At the other extreme, if customers always compare prices, one should almost always follow competitors' price changes. From our model's estimation, consumers exhibit a strong price comparison behavior ($\lambda = 0.7911$). This explains why under such a model, prices are more responsive to competitor price changes, as shown in Column 6 in Table 4; our suggested prices change by a similar order of magnitude when competitors change their prices (e.g., when competitor prices change by ¥10 the estimated prices change by ¥4 to ¥6). We do not match fully the price change of ¥10 because we already price lower than competitors prior to this change.[13]

In sum, a model that fails to capture the extent to which consumers compare prices across retailers will lead to insufficient response to competitor price changes. This is true even when, like in our setting, we are able to capture unbiased price elasticities with a randomized price experiment.

In order to gauge the potential revenue impact of our proposed best-response pricing strategy, we conducted an out-of-sample test on the historical data—the same period which we used to compare the estimation results in Table 2. We find that, had the retailer used this best-response pricing algorithm to adjust prices in response to price changes in the market instead, they would have obtained 7 percent higher category revenue while holding gross margin on a par with the existing practice. The revenue improvement is estimated as follows. First, we obtain the optimal prices by solving the best-response price optimization for the historical period. Second, we obtain daily total unique web traffic to the feeding bottle category and use it as a proxy for the market size of each day. Note that this measure of market size accounts for both those customers who purchased as well as those who visited but did not purchase. Lastly, for each customer in the market, we predict her probabilities of purchasing each product on each day given the new prices calculated in the first step. We then obtain expected demand for each product on each day multiplying these probabilities with the market size obtained in the second step.

[13] We also note that under the model with independent demand shocks, the price correlations between our retail partner and Competitors 1, 2, 3, and 4: 0.519, 0.312, 0.346, and 0.337 respectively. While with correlated demand shocks, the price correlations are 0.960, 0.405, 0.589, and 0.373. This indicates that with correlated demand shocks, competitors' prices are more correlated.

Finally, we would like to note that when determining the best prices, our methodology only factors into account the current prices charged by competitors, but not incorporate competitors' potential responses to our price changes (Li et al. 2016).

7. Testing Best-Response Pricing with a Controlled Experiment

We test the performance of our best-response pricing algorithm with a field experiment at our partner retailer, in which prices are changed daily according to the outcome of our model.

7.1. Experiment Design and Implementation

Note that the objective of our pricing algorithm is to maximize revenue for the category. A specific product's price changes will lead not only to revenue changes for that product but also for other products with similar features due to substitution. For this reason, a valid experiment design requires minimal substitution between treatment and control groups; otherwise, the control group will be contaminated due to the spillover effect. Instead of matching products on their main attributes, we identify one existing attribute that allows a clean separation between market segments: bottles designed for ranges of babies' ages. Each feeding bottle is designed for a specific age group because babies of different ages require different nipple sizes, nipple shapes, and bottle volume. There is hardly any substitution between feeding bottles that are designed for a particular age group. Within our 15 feeding bottles, we identify 9 bottles that are designed specifically for babies ranging from birth to six months old, and the remaining 6 bottles are designed for babies seven months and older.

Although we believe it is highly unlikely that the pricing algorithm could work for one group of bottles and not for the other, to ensure the difference between the two age groups does not drive the result of our pricing algorithm, we rotate the implementation of the treatment between these two groups. Table 5 shows the design of the experiment, which lasts for a total of five weeks. Note that in the last week, we let the experiment return to a state where neither of the two groups receives treatment, to further confirm that we are not capturing an overall time-trend effect.

Table 5 Experiment Design: Test Best Response Pricing

	Group 1 (Baby age: 0-6 months)	Group 2 (Baby age: 7 months and above)
Week 0	Control	Control
Week 1	Treatment	Control
Week 2	Control	Treatment
Week 3	Treatment	Treatment
Week 4	Control	Control

Control: current pricing practice. Treatment: best-response pricing algorithm.

Typically, one can use the Difference-in-Differences (DiD) estimator to compute the treatment effect in such settings. The critical identifying assumption underlying the DiD estimator is the existence of a parallel trend. That is, the two groups would otherwise follow the same trend in the absence of the treatment. However, in our context particularly, because the control and treatment groups are not similar products matched according to product attributes but rather dissimilar products to minimize across group spillover, it is more likely the parallel trend assumption may not hold. It is possible that in a short period of time, the demand for the two groups of products actually follows different trends due to such reasons as seasonality of births. Indeed, we find evidence from data prior to the experiment that products in Group 1 exhibit a slight upward trend in demand, while products in Group 2 exhibit a slight downward trend in demand, and these trends exist across all regions that the company operates in. When the parallel trend assumption is violated, the DiD estimator will be biased (Wooldridge 2010).

To seek for a solution to this problem, we note that our retail partner operates in multiple regions of the country. These regions are geographically separated, 800 to 1,000 miles apart. Each region has its own warehouse, logistics team, and management personnel. The population within each region is very dense; thus, the demand is sufficiently large to allow for dedicated shipping in which a warehouse ships only to customers within its own region. This means each region can be viewed as a separate market since there is no demand or supply substitution across these markets. The retailer's website requires customers to choose their locations before being able to browse any content. Once a customer places an order, the individual's shipping address will be verified again to avoid any cross-selling or overlap between transactions across regions. This is a common practice among all major retailers in the country.

This modus operandi gives us an advantage in our estimation approach since we are able to introduce an additional comparison group. We implement the pricing algorithm in only one of the regions, region A, leaving the other regions as controls. We match region A with the other regions in the country by looking at online shopping traffic, sales volume, demographics (population density and income distribution), and, in particular, margin rates, one of the key constraints of our optimization procedure. After the process, we identify two regions, B and C, that match region A.

To ensure the validity of regions B and C as controls, we confirm none of their pricing managers or product managers are aware of the concurrent experiment conducted in region A. In region A, where the experiment is implemented, it is not feasible to keep managers entirely unaware of the experiment because we need their cooperation to be able to adjust prices. However, we exercise extra care not to alter pricing managers' regular decisions so that controls in Table 5 are valid. To do this, we communicate to the managers that they should make pricing decisions as they normally do. Specifically, instead of framing the experiment as a test of a potentially superior pricing algorithm (i.e., a pilot or an implementation), we communicate to the team that the test is a new randomized pricing experiment. Furthermore, when a group of products receives treatment, we ask a designated person to update the product prices instead of sending the recommended prices to the pricing team, in an effort to avoid non-compliance or biasing their decisions.

7.2. Triple-Differences Estimator

The experiment treatment we describe allows us to adopt a triple-differences estimator to measure the impact of the proposed pricing methodology. The triple differences come from the comparisons of periods when the treatment is applied and when it is not; the two age groups; and regions A, B, and C. Note that the key distinction of this estimator from the more commonly used difference-in-differences (DiD) estimator is the presence of a third comparison group (i.e., across regions).

If the null hypothesis is that without treatment the ratio of Group 1 revenue (or Group 2 revenue) at region A over Group 1 revenue (or Group 2 revenue) at region B is constant, then we could use the following regression to examine the effect of treatment, where the treatment alternates between Groups 1 and 2.

$$\begin{aligned}\ln(Rev_{gdm}) = {} & \alpha_0 + \alpha_1 \, Week\ Dummy_{gdm} + \alpha_2 \, Group1_{gdm} + \alpha_3 \, Treatment_{gdm} + \\ & \alpha_4 Region\ Dummy_{gdm} + \alpha_5 Day\ of\ Week_{gdm} + \alpha_6 Margin_{gdm} + \alpha_7 Traffic_{gdm} + \\ & \alpha_8 Region \times Margin_{gdm} + \alpha_9 Region_{gdm} \times Traffic_{gdm} + \varepsilon_{gdm},\end{aligned}$$

where subscript g denotes group, d denotes date, and m denotes geographical region. For instance, Rev_{gdA} denotes group g's revenue on day d at region A. $Treatment_{gd} = 1$ for Group 1 in weeks 1 and 3 and Group 2 in weeks 2 and 3 in region A; otherwise, it equals zero. The coefficient of interest is α_3, which can be interpreted as the percentage of revenue changes due to the treatment. Table 6 shows the estimation results.[14] The revenue increases for the treated category ranges from 10.9 percent to 12.4 percent, depending on control variables included in the regression. Note that this revenue improvement is achieved while holding gross margin on a par with the existing practice through the margin constraint. Furthermore, we obtained similar estimates when we compared the test and control regions during only periods when both or neither age group of feeding bottles is tested. The analysis is shown in the electronic companion EC.4.

To summarize, our proposed model is capable of growing revenue because (1) it measures price elasticity accurately, which allows the retailer to charge a category revenue-maximizing margin for each product, and (2) it measures cross-price elasticity accurately, which allows the retailer to respond to competition only when necessary instead of attempting to always match all competitors' price changes.

[14] A version of this Table reporting all covariates can be found in Table EC.3

Table 6 Revenue Impact of Best Response Pricing

ln(daily revenue)	(1) w/o daily margin and traffic	(2) w/ daily margin	(3) w/ daily margin & traffic
Treatment (α_3)	0.109*	0.112*	0.124*
	(0.063)	(0.064)	(0.064)
Daily Margin	no	no	yes
Daily traffic	no	yes	yes
Day of Week Dummy	yes	yes	yes
Month Dummy	yes	yes	yes
Const	7.976***	8.017***	7.922***
# obs	432	432	432
# treatment	38	38	38
R-sq	0.732	0.733	0.737

*** $p < 0.01$, ** $p < 0,05$, * $p < 0.1$. *Products and dates were dropped when the algorithm was wrongly or not implemented.*

8. Conclusion

To charge the right price, one first needs to obtain an unbiased measure of price elasticity. This often is challenging when relying simply on historical sales and price data because prices are very likely to be correlated with unobserved demand shocks that are accessible to pricing managers but not to researchers. In a setting where prices change rapidly, such as online retailing, the task becomes even more difficult because the commonly used instruments do not change as fast or are subject to other limitations. We show that a randomized price experiment is an effective way to address this concern. However, since it is costly to run field experiments, it is crucial to design the experiment in such a way that it will induce a sufficient number of random variations in both absolute and relative (relative to competitors' prices) terms within a reasonable price range and time frame.

Accurate measure of price elasticity alone is not sufficient for price prescriptions, particularly in a dynamic competitive setting. Levels of price elasticity only suggest the products on which to charge higher or lower margins; however, they do not completely answer how to respond to competitor price changes. Accurate response to competitor price changes depends most critically on consumer engagement in price comparison across retailers. Moreover, responses should be differentiated based on the significance of the competitor in the marketplace.

While competitors' price and product availability data can be obtained by monitoring competitor websites, the absence of competitors' sales information poses a significant challenge to estimate a full consumer-choice model. We show that our and competitors' stockouts can be used as a valid identification strategy to achieve this objective because they provide a temporary variation to the consumer choice set.

We want to emphasize that field studies present a set of challenges different from those arising from either conducting laboratory experiments or relying on observational data. Field studies involve generating desired data in a way that minimizes interference from other parallel business activities with compatible or competing interests that could contaminate the result of the experiment ex-post. For instance, framing the experiment and communicating it to stakeholders are particularly important for the validity of the control group.

Based on estimates of the proposed consumer-choice model, we show that a best-response pricing algorithm that takes into account consumer behavior, competitor actions, and supply parameters demonstrates significant revenue improvement—11 percent for the product category under study. Such improvement is not specific to this one category in particular. For example, we also obtained 19 percent revenue improvement in the kettle category in a follow-up study (see Table EC.4).

Finally, with ever-expanding product spaces and competitor entries and exits, market conditions change rapidly for online retailers. Hence, we suggest retailers test demand responses periodically to keep up with the evolving market and implement an effective dynamic pricing strategy.

Acknowledgments

The authors would like to thank the retail partner for close collaboration on this project, and conference participants at the 2014 Consortium for Operational Excellence in Retailing at The Wharton School and 2014 DIIE Collaborative Academic/Practitioner Workshop on Operational Innovation at London Business School for helpful comments. The authors would also like to thank INFORMS Revenue Management and Pricing Society for its recognition of our work through the Practice Award. The financial support of the Wharton Global Initiatives program and Fishman-Davidson Center at the Wharton School of Business is gratefully acknowledged.

References

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–90.

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

Cohen, Maxime, Ngai-Hang Leung, Kiran Panchamgam, Georgia Perakis, Anthony Smith. 2016. The impact of linear optimization on promotion planning. Forthcoming.

Conlon, Christopher T., Julie Holland Mortimer. 2013. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics* **5**(4) 1–30.

Erdem, T., J. Swait, J. Louviere. 2002. The impact of brand credibility on cosnumer price sensitivity. *Intern. J. of Research in Marketing* **19** 1–19.

Farias, Vivek F., Srikanth Jagabathula, Devavrat Shah. 2013. A nonparametric approach to modeling choice with limited data. *Management Science* **59**(2) 305–322.

Ferreira, Kris Johnson, Bin Hong Alex Lee, David Simchi-Levi. 2016. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* **18**(1) 69–88.

Gaur, Vishal, Marshall L. Fisher. 2005. In-store experiments to determine the impact of price on sales. *Production and Operations Management* **14**(4) 377–387.

Guadagni, Peter M., John D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* **2**(3) 203–238.

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

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

Harsha, Pavithra, Shivaram Subramanian, Joline Uichanco. 2016. Omni-channel revenue management through integrated pricing and fulfillment planning. Working Paper, Ross School of Business, University of Michigan.

Hausman, Jerry. 1996. Valuation of new goods under perfect and imperfect competition. Timothy Bresnaha, Robert Gordon, eds., *The Economics of New Goods*. University of Chicago Press, Chicago, IL.

Honhon, Dorothe, Vishal Gaur, Sridhar Seshadri. 2010. Assortment planning and inventory decisions under stockout-based substitution. *Operations Research* **58**(5) 1364–1379.

Jagabathula, Srikanth, Paat Rusmevichientong. 2016. A nonparametric joint assortment and price choice model. Forthcoming.

Kök, A. Gurhan, Yi Xu. 2011. Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science* **57**(9) pp. 1546–1563.

Koppelman, F., C.H. Wen. 2000. The paired combinatiorial logit model: properties, estimation and application. *Transportation Research Part B* **34** 75–89.

Lal, Rajiv, Ram Rao. 1997. Supermarket competition: The case of every day low pricing. *Marketing Science* **16**(1) 60–80.

Li, Jun, Serguei Netessine, Sergei Koulayev. 2016. Price to compete ... with many: How to identify price competition in high dimensional space. Working Paper, Ross School of Business, University of Michigan.

Lu, Y., , A. Musalem, M. Olivares, A. Schilkrut. 2013. Measuring the effect of waiting time on customer purchases. *Management Science* **59**(8) 1743–1763.

Mead, R., S. G. Gilmour, A. Mead. 2012. *Statistical Principles for the Design of Experiments*. First edition ed. Cambridge University Press, Cambridge, United Kingdom.

Nelson, P. 1974. Advertising as information. *Journal of Political Economy* **81** 729–954.

Nevo, Aviv. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.

Newman, Jeffrey P, Mark E Ferguson, Laurie A Garrow. 2013. Estimating gev models with censored data. *Transportation Research Part B: Methodological* **58** 170–184.

Ozer, Ozalp, Yanchong Zheng, Yufei Ren. 2014. Trust, trustworthiness, and information sharing in supply chains bridging china and the united states. *Management Science* **60**(10) 2435–2460.

Perdikaki, O., S. Kesavan, J.M. Swaminathan. 2012. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management* **14**(1) 145–162.

Tellis, G.J., G. Gaeth. 1990. Best value, price-seeking and price aversion: The impact of information and learning on consumer choices. *Journal of Marketing* **54** 34–45.

Tereyagoglu, Necati, Peter Fader, Senthil K Veeraraghavan. 2014. Multi-attribute loss aversion and reference dependence: Evidence from the performing arts industry. *Available at SSRN 2499265* .

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, United Kingdom.

Villas-Boas, J. Miguel, Russell S. Winer. 1999. Endogeneity in brand choice models. *Management Science* **45**(10) 1324–1338.

Vulcano, Gustavo, Garrett van Ryzin, Richard Ratliff. 2012. Estimating primary demand for substitutable products from sales transaction data. *Operations Research* **60**(2) 313–334.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Second edition ed. The MIT Press, Cambridge, Massachusetts.

Electronic Companion for Competition-Based Dynamic Pricing in Online Retailing: A Methodology Validated with Field Experiments

EC.1. Supplemental Figures and Tables

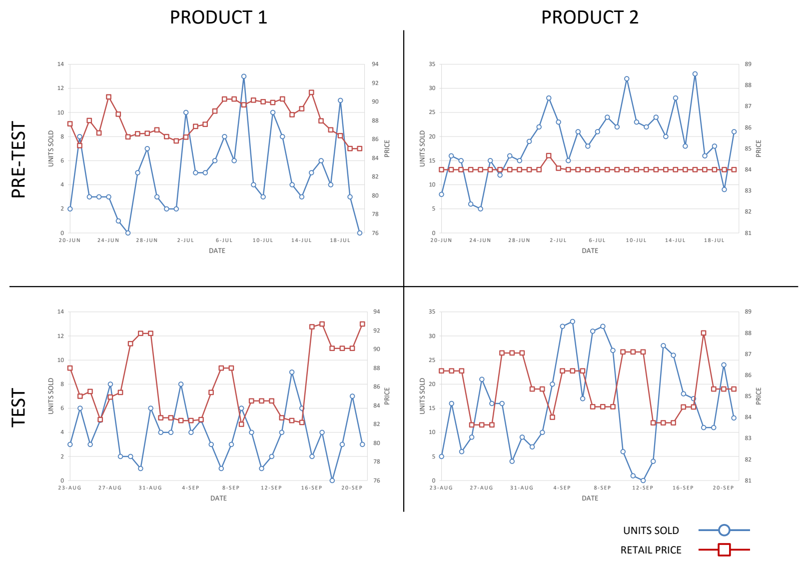


Figure EC.1 Example of Test Implementation Results — Price-Sales Relationship

Figure EC.2 Example of Test Implementation Results — Price-Market Price Relationship

Table EC.1 Competitor Price Responses to the Randomized Price Experiment

	12-Hour Price Response													
	Competitor 1			**Competitor 2**			**Competitor 3**			**Competitor 4**			
Our Price	↓	–	↑	↓	–	↑	↓	–	↑	↓	–	↑	**Total**
↓	1	74	0	5	66	4	3	68	4	6	67	2	75
–	4	275	4	15	258	10	19	256	8	8	263	12	283
↑	0	91	0	4	80	7	7	81	3	6	84	1	91
Total	5	440	4	24	404	21	29	405	15	20	414	15	449

	24-Hour Price Response													
	Competitor 1			**Competitor 2**			**Competitor 3**			**Competitor 4**			
Our Price	↓	–	↑	↓	–	↑	↓	–	↑	↓	–	↑	**Total**
↓	1	72	2	9	59	7	3	61	11	7	66	2	75
–	4	273	5	20	243	19	28	232	22	15	254	13	282
↑	0	90	1	5	75	11	11	71	9	9	80	2	91
Total	5	440	4	24	404	21	29	405	15	20	414	15	449

↓: Price decreases. ↑: Price increases. –: No price changes.

Table EC.2 Price Variations, Stockouts and the Estimated Competitor Significance

	(1) Average Price	(2) Within Product Price Variation	(3) Between Product Price Variation	(4) Price Change	(5) Stockout	(6) Competitor Significance (α_r)
Partner Retailer	¥73.94	¥4.99	¥31.03	38.6%	3.8%	baseline
Competitor 1	¥85.09	¥3.88	¥33.68	1.9%	2.2%	0.217**
Competitor 2	¥78.47	¥1.97	¥31.15	16.3%	0%	0.017
Competitor 3	¥78.70	¥11.04	¥30.85	19.5%	0%	-1.836***
Competitor 4	¥92.77	¥15.36	¥30.90	11.1%	17.9%	-2.464***

Table EC.3 Revenue Impact of Best Response Pricing – Feeding Bottle

ln(daily revenue)	(1) w/o daily margin and traffic	(2) w/ daily margin	(3) w/ daily margin & traffic
Treatment (α_3)	0.109*	0.112*	0.124*
	(0.063)	(0.064)	(0.064)
Group 1 Dummy	-0.118***	-0.114***	-0.117***
	(0.033)	(0.037)	(0.037)
No test week 1	Baseline		
Test week 1	-0.284***	-0.283***	-0.283***
	(0.067)	(0.067)	(0.068)
Test week 2	-0.112	-0.114	-0.124*
	(0.073)	(0.074)	(0.075)
No test week 2	0.267***	0.270***	0.246***
	(0.051)	(0.051)	(0.052)
Test week 3	0.138**	0.136**	0.100
	(0.059)	(0.062)	(0.065)
Region B	-0.656***	-0.779***	-0.977*
	(0.043)	(0.139)	(0.586)
Region C	-1.257***	-1.282***	-2.026***
	(0.047)	(0.126)	(0.726)
Daily margin		-0.539	-0.363
		(1.023)	(1.057)
Region B X Daily margin		1.683	1.504
		(1.875)	(1.921)
Region C X Daily margin		0.314	0.402
		(1.682)	(1.643)
ln(Daily traffic)			0.020
			(0.105)
Region B X ln(Daily traffic)			0.062
			(0.139)
Region C X ln(Daily traffic)			0.216
			(0.178)
day of week dummy	yes	yes	yes
month dummy	yes	yes	yes
const	7.976***	8.017***	7.922***
# obs	432	432	432
# treatment	38	38	38
R-sq	0.732	0.733	0.737

**** $p < 0.01$, ** $p < 0,05$, * $p < 0.1$. Products and dates were dropped when the algorithm was wrongly or not implemented.*

Table EC.4 Revenue Impact of Best Response Pricing – Kettle

ln(daily revenue)	(1) w/o daily margin and traffic	(2) w/ daily margin	(3) w/ daily margin & traffic
Our Algorithm	0.190	0.227*	0.241*
	(0.116)	(0.123)	(0.134)
After	0.155	0.151	0.135
	(0.135)	(0.136)	(0.147)
Region B	-1.161***	-1.170***	0.397
	(0.082)	(0.088)	(0.724)
Region C	-1.324***	-1.288***	-0.680
	(0.088)	(0.090)	(0.688)
Daily margin		-1.451	-0.548
		(1.401)	(1.243)
Region B X Daily margin		0.412	-0.225
		(1.746)	(1.608)
Region C X Daily margin		2.672	2.078
		(1.967)	(1.884)
ln(Daily traffic)			0.588***
			(0.154)
Region B X ln(Daily traffic)			-0.345
			(0.228)
Region C X ln(Daily traffic)			0.076
			(0.211)
day of week dummy	yes	yes	yes
month dummy	yes	yes	yes
const	8.664***	8.649***	6.422***
	(0.078)	(0.080)	(0.569)
# obs	183	183	183
# treatment	26	26	26
R-sq	0.725	0.730	0.764

This experiment lasted for four weeks. Eighteen SKUs among the top thirty SKUs (by sales units) in this category participated in the experiment. The other SKUs did not participate because the retailer had little discretion in changing their prices due to manufacturer price protections. During the four weeks of experiment, we applied our algorithm to all eighteen SKUs in region A, while using the other two regions, B and C, as controls.

EC.2. Stockout and Price Variations as Sources of Identification: A Simulation Study

We simulate a simple choice scenario with two retailers, R and C, each of whom offers two products, 1 and 2, with the following parameters,

- product-specific intercepts $\alpha_1 = 0.5, \alpha_2 = 1.5$,
- product-specific price elasticity $\beta_1 = -0.5, \beta_2 = -1$,
- retailer-specific intercept $\alpha_R = 0, \alpha_C = 0.2$,
- extent of price comparison $\lambda = 0.7$,
- mean utility for the outside option equals 0.

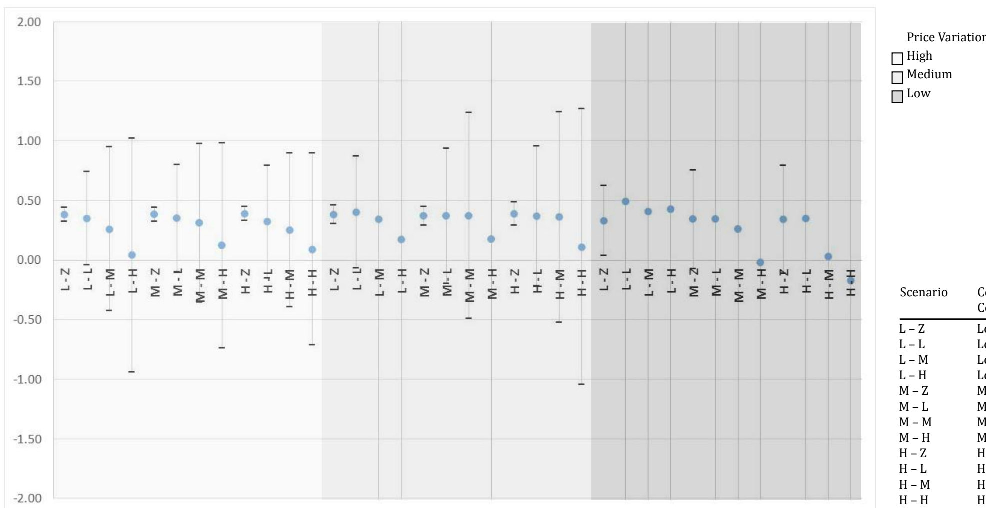
We simulate (1) three levels of price variation: low, medium and high, i.e., coefficient of variance $= 0.1, 0.5$, and, 1, respectively; (2) three levels of price correlation: low, medium and high, $\rho = 0, 0.5$, and 0.8, respectively; and (3) four levels of stockout probability: zero, low, medium and high, $0\%, 5\%, 10\%$, and 20%, respectively. For each of the 3·3·4=36 scenarios, we conduct 100 simulation rounds. In each round, we simulate decisions from 2,000 customers during a 30-day period. Note that we only observe retailer R's sales, not retailer C's sales. When estimating the model, we seed the estimation with 10 different random initial values.

Without considering changes in availability, price variations across products and retailers are the sole source of variation for identification. If variation of prices is low or if a competitor's price is highly correlated with a retailer's own price (i.e., relative price variation is low), ignoring changes in availability will lead to biased and imprecise estimates. We make two observations from the results shown in Table EC.5.

First, when price variation decreases, correlation of competitors' prices increases, or stockout probability increases, the average estimate from the 100 rounds of simulation deviates significantly from the true parameters. For example, with low price variation, high competitor price correlation and high stockout probability, the estimated $\hat{\alpha}_1 = -0.17$ (true value $\alpha_1 = 0.5$), $\hat{\alpha}_C = 0.05$ (true value $\alpha_C = 0.2$). However, with stockout information, the estimates are much closer to the true values, $\hat{\alpha}_1 = 0.44$ (true value $\alpha_1 = 0.5$) and $\hat{\alpha}_C = 0.24$ (true value $\alpha_C = 0.2$).

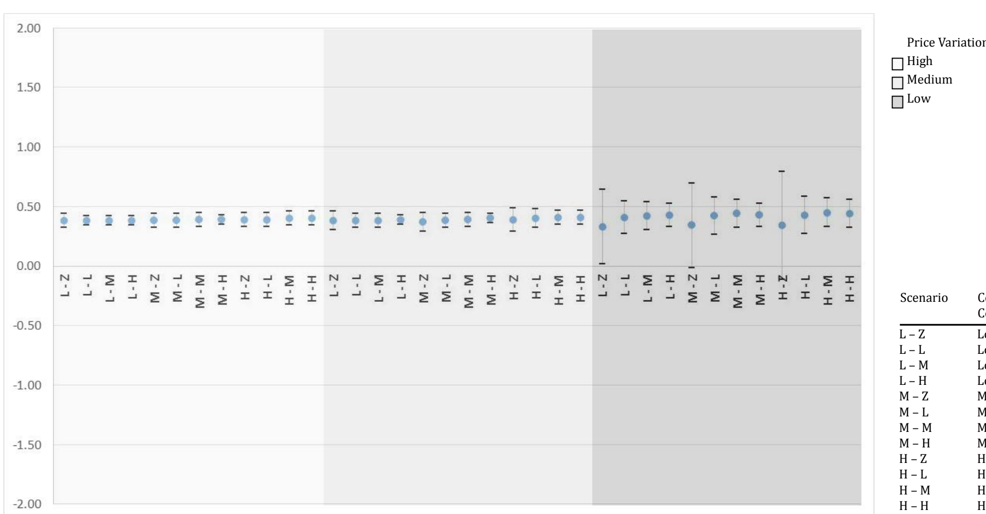
Second, when accounting for stockouts, the estimates not only are closer to the true values, they also are more precise. One can see the estimates become much less precise as a whole as variation of prices decreases, correlation of competitors' prices increases, and stockout probability increases. The standard deviation of the 100 estimates increases sharply, from 0.03 to 10.70 in Column 1, for example. More importantly, even when there is *zero* stockout, low price variation and high price correlation lead to imprecise estimates as well. For example, 100 rounds of simulation yield a standard deviation of 0.23 with low price variation and high competitor-price correlation, almost eight times as large as the standard deviation of 0.03 with high price variation and low competitor-price correlation in Column 1. For better visualization, Figure EC.3 illustrate the estimates of $\hat{\alpha}_1$ graphically, including the confidence interval of the estimates, where the benefits of including the stockout information become evident.

Ignoring Stock Out



Scenario	Comp. Price Corr.	Stockout Probability
L – Z	Low	Zero
L – L	Low	Low
L – M	Low	Medium
L – H	Low	High
M – Z	Medium	Zero
M – L	Medium	Low
M – M	Medium	Medium
M – H	Medium	High
H – Z	High	Zero
H – L	High	Low
H – M	High	Medium
H – H	High	High

Accounting for Stock Out



Scenario	Comp. Price Corr.	Stockout Probability
L – Z	Low	Zero
L – L	Low	Low
L – M	Low	Medium
L – H	Low	High
M – Z	Medium	Zero
M – L	Medium	Low
M – M	Medium	Medium
M – H	Medium	High
H – Z	High	Zero
H – L	High	Low
H – M	High	Medium
H – H	High	High

Figure EC.3 Estimation Results Based on 100 Simulations per Scenario

Table EC.5 Estimation Results Based on 100 Simulations per Choice Scenario

Price Variation	Comp Price Correlation	Stockout Probability	Ignore Stockouts $\alpha_1 = 0.5$ (1)	$\alpha_2 = 1.5$ (2)	$\beta_1 = -0.5$ (3)	$\beta_2 = -1$ (4)	$\alpha_C = 0.2$ (5)	$\lambda = 0.7$ (6)	Account for Stockouts $\alpha_1 = 0.5$ (7)	$\alpha_2 = 1.5$ (8)	$\beta_1 = -0.5$ (9)	$\beta_2 = -1$ (10)	$\alpha_C = 0.2$ (11)	$\lambda = 0.7$ (12)
High	Low	Zero	0.38 (0.03)	1.46 (0.02)	-0.57 (0.01)	-1.11 (0.01)	0.21 (0.01)	0.71 (0.01)	0.38 (0.03)	1.46 (0.02)	-0.57 (0.01)	-1.11 (0.01)	0.21 (0.01)	0.71 (0.01)
High	Low	Low	0.35 (0.20)	1.37 (0.21)	-0.59 (0.11)	-1.07 (0.11)	0.22 (0.07)	0.62 (0.08)	0.38 (0.02)	1.45 (0.02)	-0.57 (0.01)	-1.11 (0.01)	0.21 (0.01)	0.71 (0.01)
High	Low	Medium	0.26 (0.35)	1.20 (0.36)	-0.59 (0.14)	-0.98 (0.16)	0.17 (0.13)	0.56 (0.11)	0.38 (0.02)	1.45 (0.02)	-0.57 (0.01)	-1.12 (0.01)	0.21 (0.01)	0.71 (0.01)
High	Low	High	0.04 (0.50)	0.89 (0.51)	-0.62 (0.17)	-0.88 (0.21)	0.13 (0.25)	0.45 (0.14)	0.38 (0.02)	1.46 (0.02)	-0.57 (0.01)	-1.12 (0.02)	0.21 (0.01)	0.70 (0.01)
High	Medium	Zero	0.38 (0.03)	1.46 (0.03)	-0.58 (0.01)	-1.13 (0.01)	0.22 (0.01)	0.70 (0.01)	0.38 (0.03)	1.46 (0.03)	-0.58 (0.01)	-1.13 (0.01)	0.22 (0.01)	0.70 (0.01)
High	Medium	Low	0.35 (0.23)	1.40 (0.25)	-0.60 (0.10)	-1.11 (0.10)	0.22 (0.08)	0.63 (0.08)	0.38 (0.03)	1.46 (0.03)	-0.58 (0.01)	-1.13 (0.02)	0.22 (0.01)	0.70 (0.01)
High	Medium	Medium	0.31 (0.34)	1.28 (0.36)	-0.62 (0.12)	-1.08 (0.14)	0.21 (0.13)	0.55 (0.11)	0.39 (0.03)	1.47 (0.02)	-0.58 (0.01)	-1.13 (0.01)	0.22 (0.01)	0.70 (0.01)
High	Medium	High	0.12 (0.44)	0.97 (0.47)	-0.62 (0.14)	-0.97 (0.18)	0.19 (0.23)	0.43 (0.15)	0.39 (0.02)	1.47 (0.03)	-0.58 (0.01)	-1.13 (0.02)	0.23 (0.01)	0.70 (0.01)
High	High	Zero	0.39 (0.03)	1.49 (0.03)	-0.59 (0.01)	-1.15 (0.02)	0.23 (0.01)	0.69 (0.01)	0.39 (0.03)	1.49 (0.03)	-0.59 (0.01)	-1.15 (0.02)	0.23 (0.01)	0.69 (0.01)
High	High	Low	0.32 (0.24)	1.42 (0.27)	-0.61 (0.09)	-1.15 (0.09)	0.23 (0.07)	0.63 (0.08)	0.39 (0.03)	1.49 (0.03)	-0.59 (0.01)	-1.15 (0.02)	0.23 (0.01)	0.69 (0.01)
High	High	Medium	0.25 (0.33)	1.34 (0.37)	-0.61 (0.12)	-1.13 (0.12)	0.23 (0.12)	0.57 (0.11)	0.40 (0.03)	1.49 (0.03)	-0.59 (0.01)	-1.15 (0.02)	0.23 (0.01)	0.69 (0.01)
High	High	High	0.09 (0.41)	1.06 (0.45)	-0.60 (0.14)	-1.08 (0.16)	0.22 (0.19)	0.47 (0.16)	0.40 (0.03)	1.50 (0.03)	-0.59 (0.01)	-1.15 (0.02)	0.23 (0.01)	0.69 (0.01)
Medium	Low	Zero	0.38 (0.04)	1.50 (0.03)	-0.58 (0.01)	-1.14 (0.02)	0.21 (0.01)	0.71 (0.01)	0.38 (0.04)	1.50 (0.03)	-0.58 (0.01)	-1.14 (0.02)	0.21 (0.01)	0.71 (0.01)
Medium	Low	Low	0.40 (0.24)	1.51 (0.25)	-0.61 (0.12)	-1.12 (0.11)	0.25 (0.07)	0.64 (0.06)	0.38 (0.03)	1.50 (0.03)	-0.58 (0.01)	-1.15 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Low	Medium	0.34 (2.71)	1.41 (2.68)	-0.63 (0.16)	-1.11 (0.17)	0.23 (0.11)	0.59 (0.09)	0.38 (0.03)	1.51 (0.03)	-0.58 (0.01)	-1.15 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Low	High	0.17 (2.59)	1.17 (2.53)	-0.63 (0.23)	-1.06 (0.27)	0.22 (0.22)	0.49 (0.14)	0.39 (0.02)	1.52 (0.03)	-0.58 (0.01)	-1.15 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Medium	Zero	0.37 (0.04)	1.51 (0.04)	-0.59 (0.01)	-1.16 (0.02)	0.22 (0.01)	0.70 (0.01)	0.37 (0.04)	1.51 (0.04)	-0.59 (0.01)	-1.16 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Medium	Low	0.37 (0.29)	1.50 (0.31)	-0.60 (0.11)	-1.16 (0.09)	0.23 (0.07)	0.64 (0.06)	0.38 (0.03)	1.52 (0.04)	-0.59 (0.01)	-1.16 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Medium	Medium	0.37 (0.44)	1.45 (0.45)	-0.61 (0.15)	-1.16 (0.14)	0.25 (0.12)	0.59 (0.09)	0.39 (0.03)	1.53 (0.03)	-0.59 (0.01)	-1.16 (0.02)	0.22 (0.01)	0.70 (0.01)
Medium	Medium	High	0.17 (2.08)	1.22 (2.07)	-0.61 (0.20)	-1.13 (0.22)	0.23 (0.21)	0.50 (0.13)	0.40 (0.02)	1.53 (0.03)	-0.59 (0.01)	-1.16 (0.02)	0.23 (0.01)	0.69 (0.01)
Medium	High	Zero	0.39 (0.05)	1.54 (0.06)	-0.60 (0.01)	-1.18 (0.02)	0.23 (0.01)	0.69 (0.01)	0.39 (0.05)	1.54 (0.06)	-0.60 (0.01)	-1.18 (0.02)	0.23 (0.01)	0.69 (0.01)
Medium	High	Low	0.37 (0.30)	1.54 (0.33)	-0.60 (0.10)	-1.19 (0.09)	0.24 (0.07)	0.65 (0.06)	0.40 (0.04)	1.56 (0.05)	-0.60 (0.01)	-1.19 (0.02)	0.23 (0.01)	0.69 (0.01)
Medium	High	Medium	0.36 (0.45)	1.48 (0.49)	-0.60 (0.14)	-1.18 (0.13)	0.25 (0.12)	0.61 (0.09)	0.41 (0.03)	1.56 (0.04)	-0.60 (0.01)	-1.18 (0.02)	0.23 (0.01)	0.69 (0.01)
Medium	High	High	0.11 (0.59)	1.25 (0.65)	-0.58 (0.19)	-1.16 (0.18)	0.20 (0.21)	0.55 (0.14)	0.41 (0.03)	1.56 (0.04)	-0.60 (0.01)	-1.18 (0.02)	0.23 (0.01)	0.69 (0.01)
Low	Low	Zero	0.33 (0.15)	1.55 (0.17)	-0.60 (0.04)	-1.21 (0.07)	0.22 (0.03)	0.69 (0.02)	0.33 (0.16)	1.55 (0.17)	-0.60 (0.04)	-1.21 (0.07)	0.21 (0.03)	0.69 (0.02)
Low	Low	Low	0.49 (3.35)	1.92 (3.35)	-0.58 (0.29)	-1.31 (0.40)	0.23 (0.19)	0.64 (0.12)	0.41 (0.07)	1.64 (0.12)	-0.60 (0.04)	-1.22 (0.06)	0.24 (0.02)	0.69 (0.02)
Low	Low	Medium	0.41 (8.92)	1.72 (8.89)	-0.54 (0.45)	-1.26 (0.52)	0.19 (0.28)	0.65 (0.18)	0.42 (0.06)	1.65 (0.10)	-0.61 (0.04)	-1.22 (0.05)	0.24 (0.02)	0.68 (0.02)
Low	Low	High	0.43 (7.59)	1.89 (12.16)	-0.53 (0.77)	-1.38 (0.82)	0.07 (0.45)	0.63 (0.32)	0.43 (0.05)	1.64 (0.08)	-0.61 (0.04)	-1.22 (0.04)	0.24 (0.01)	0.68 (0.01)
Low	Medium	Zero	0.34 (0.21)	1.56 (0.24)	-0.61 (0.04)	-1.22 (0.06)	0.22 (0.05)	0.69 (0.02)	0.34 (0.18)	1.55 (0.20)	-0.61 (0.04)	-1.23 (0.06)	0.22 (0.04)	0.69 (0.02)
Low	Medium	Low	0.34 (7.41)	1.67 (7.45)	-0.57 (0.28)	-1.22 (0.34)	0.24 (0.22)	0.66 (0.12)	0.42 (0.08)	1.65 (0.12)	-0.61 (0.04)	-1.23 (0.06)	0.24 (0.02)	0.68 (0.02)
Low	Medium	Medium	0.26 (11.12)	1.65 (11.05)	-0.54 (0.47)	-1.23 (0.45)	0.18 (0.28)	0.66 (0.18)	0.44 (0.06)	1.68 (0.09)	-0.62 (0.04)	-1.23 (0.05)	0.24 (0.02)	0.68 (0.02)
Low	Medium	High	-0.02 (7.52)	1.29 (7.47)	-0.56 (0.80)	-1.29 (0.62)	0.07 (6.20)	0.64 (0.44)	0.43 (0.05)	1.66 (0.07)	-0.62 (0.04)	-1.22 (0.04)	0.24 (0.01)	0.68 (0.01)
Low	High	Zero	0.34 (0.23)	1.55 (0.26)	-0.61 (0.05)	-1.23 (0.06)	0.22 (0.05)	0.68 (0.02)	0.34 (0.23)	1.56 (0.27)	-0.61 (0.05)	-1.24 (0.06)	0.22 (0.05)	0.68 (0.02)
Low	High	Low	0.35 (6.74)	1.68 (6.77)	-0.55 (0.34)	-1.22 (0.33)	0.24 (0.25)	0.67 (0.14)	0.43 (0.08)	1.66 (0.11)	-0.61 (0.05)	-1.24 (0.06)	0.24 (0.02)	0.68 (0.02)
Low	High	Medium	0.03 (7.45)	1.38 (7.49)	-0.56 (0.56)	-1.25 (0.44)	0.17 (0.31)	0.70 (0.22)	0.45 (0.06)	1.68 (0.09)	-0.62 (0.05)	-1.24 (0.05)	0.24 (0.02)	0.68 (0.02)
Low	High	High	-0.17 (10.70)	1.26 (10.60)	-0.53 (0.91)	-1.25 (0.60)	0.05 (2.27)	0.64 (2.96)	0.44 (0.06)	1.66 (0.07)	-0.62 (0.05)	-1.23 (0.04)	0.24 (0.02)	0.68 (0.01)

Note: average estimates from all 100 simulations are listed outside parentheses, while standard deviations are inside parentheses.

EC.3. Results from Reduced-Form Regressions using the Experiment Data

Besides the consumer choice model, we also estimated the reduced-form demand model where demand is linear in prices (own price and competitors' prices) and axillary covariates (product fixed effects, day-of-week and holiday effects). Specially, we estimated the following demand models with increasing complexity,

$$\begin{aligned}
d_{jt} &= \alpha_j + \beta p_{jt} + X_t\gamma + \epsilon_{jt}, \\
d_{jt} &= \alpha_j + \beta p_{jt} + \sum_{r=2}^{J} \beta_r p_{jrt} + X_t\gamma + \epsilon_{jt}, \\
d_{jt} &= \alpha_j + \beta p_{jt} + \sum_{r=2}^{J} \beta_r p_{jrt} + \delta O_{jt} + \sum_{r=2}^{J} \beta_r O_{jrt} + X_t\gamma + \epsilon_{jt},
\end{aligned}$$

where d_{jt} denotes the observed demand of SKU j on day t. α_j is the product fixed effect for SKU j. p_{jt} denotes our retail partner's price of SKU j on day t. $p_{jrt}, r = 2, 3, , J$ denote the prices of SKU j charged by competitors. O_{jt} denotes the out-of-stock status of our retail partner, and $O_{jrt}, r = 2, 3, , J$ denote the out-of-stock statuses of all competitors. X_t include day-of-week effects and the holiday effects. The estimation results are shown in Table EC.6.

We made the following observations based on the estimation results: 1) thanks to our randomized price experiments, the coefficient estimate of self-price is significantly negative across all models; 2) however, the coefficient estimates of competitor prices are insignificant and sometimes take the wrong sign. Therefore, we cannot reliably draw inferences on how firms should respond to competitors' price changes. The reason is that reduced-form models typically need more variations in prices to identify both self- and cross-price elasticities separately. However, with the limited duration of the experiment (30 days), the variations might not be sufficient. We were able to successfully estimate the self-price elasticity because we randomized our partner's prices; however, it is impossible to do so for the competitors' prices. Due to this reason, we would want to rely on structural models to help us simultaneously identify competitor-price elasticities. Of course, this comes at the expense of making certain assumptions absent in the reduced-form model, e.g., in the underlying utility function, consumers have the same sensitivity level to prices charged by different retailers. Note that this is an implicit assumption employed by almost all consumer choice models. It does not imply that competitor-price elasticity is the same as self-price elasticity, because the ultimate measure of elasticity is also affected by the estimated significance of the competitor.

We also conducted additional analyses on each SKU separately, and also analyzed pooled data allowing all price coefficients to be product specific. The resulted estimates are even noisier (with wrong signs and larger standard errors), especially with competitor price coefficients. It further indicates that many of these estimates from a reduced-form regression model cannot be used reliably to determine best price responses to competitor price changes, because in order to do so with a reduced-form model, one would need more variations from a much larger data sample.

Table EC.6 Estimation Results from Linear Demand Models

	(1)	(2)	(3)
Self price	-0.308***	-0.334***	-0.321***
	(0.058)	(0.059)	(0.050)
Competitor 1's price		-0.020	0.060
		(0.023)	(0.045)
Competitor 2's price		0.049	0.131
		(0.151)	(0.128)
Competitor 3's price		0.104***	0.035
		(0.026)	(0.023)
Competitor 4's price		-0.022**	0.017
		(0.010)	(0.017)
Self out of stock			-19.847***
			(1.520)
Competitor 1 out-of-stock			7.504*
			(4.129)
Competitor 2 out-of-stock			NA
Competitor 3 out-of-stock			NA
Competitor 4 out-of-stock			0.848
			(2.111)
Product fixed effects	Yes	Yes	Yes
Day of Week and Holidays	Yes	Yes	Yes
# days	30	30	30
# products	15	15	15
# obs	450	450	450
adj. R-sq	0.538	0.557	0.684
MAD (daily)	0.460	0.461	0.400

EC.4. Revenue Impact of Best Response Pricing – Alternative Analyses

Besides the triple difference estimator, we also conducted additional analyses where we only used the differences across regions and across time, not between the control and treatment groups, because the parallel trend assumption between the control and treatment groups may be violated, as we discussed in Section 7. To do so, we use data during the test weeks when both groups are tested or neither is tested. We compared region A (the test region) with the benchmark regions B and C, before and after the treatment is applied. Specifically,

$$\begin{aligned}\ln(Rev_{dm}) = {} & \alpha_0 + \alpha_1\, Week\ Dummy_{dm} + \alpha_2 Region\ Dummy_{dm} + \alpha_3\, Treatment_{dm} + \alpha_4 Day\ of\ Week_{dm} + \\ & \alpha_5 Margin_{dm} + \alpha_6\, Traffic_{dm} + \alpha_7 Region \times Margin_{dm} + \alpha_8 Region_{dm} \times Traffic_{dm} + \varepsilon_{dm},\end{aligned}$$

where d denotes date and m denotes geographical region. The unit of analysis is day-region, that is, Rev_{dm} represents the total revenue of all fifteen bottles from both age groups on day d in region

m. The coefficient of interest is α_3, which can be interpreted as the percentage of revenue changes due to the treatment.

We obtain consistent estimates of the treatment effect as in the triple difference estimator, where our algorithm resulted in a 15.6% increase in revenue when bottles of both baby groups received the treatment, as shown in Column 3 in Table EC.7. We note, however, the statistical significance is in general lower compared to the results shown in Table 6. It is primarily because the sample size is smaller (the analyses are restricted to weeks only when both or neither age groups received the treatment, reducing the sample size by nearly half).

Table EC.7 Revenue Impact of Best Response Pricing – Alternative Analyses

	(1)	(2)	(3)
ln(daily revenue)	**w/o daily margin and traffic**	**w/ daily margin**	**w/ daily margin & traffic**
Our Algorithm	0.065	0.077	0.156*
	(0.081)	(0.081)	(0.090)
test week 3	0.154**	0.208***	0.134*
	(0.064)	(0.062)	(0.073)
Region B	-0.633***	-0.625***	-1.508**
	(0.055)	(0.149)	(0.738)
Region C	-1.253***	-1.273***	-2.567***
	(0.061)	(0.126)	(0.857)
daily margin		-2.592***	-2.561***
		(0.867)	(0.882)
Region B X daily margin		-0.172	-0.165
		(1.973)	(1.975)
Region C X daily margin		0.277	0.860
		(1.510)	(1.610)
ln(daily traffic)			-0.060
			(0.112)
Region B X ln(daily traffic)			0.230
			(0.181)
Region C X ln(daily traffic)			0.347
			(0.214)
day of week dummy	yes	yes	yes
month dummy	yes	yes	yes
const	7.994***	8.138***	8.402***
# obs	234	234	234
# treatment	24	24	24
R-sq	0.748	0.7578	0.7644