Using Clickstream Data to Improve Flash Sales Effectiveness

Victor Martínez-de-Albéniz* Arnau Planas† Stefano Nasini‡

Submitted: April 11, 2018; Revised: June 14, 2019.

Abstract

Flash sales retailers organize online campaigns where products are sold for a short period of time at a deep discount. The demand in these events is very uncertain, but clickstream data can potentially help retailers with detailed information about the shopping process, thereby allowing them to manage such risks. For this purpose, we build a predictive model for shoppers' sequential decisions about visiting a campaign, obtaining product information and placing a purchase, which we validate using a large data set from a leading flash sales firm. The proposed hierarchical approach mirrors the different stages of the shopping funnel and allows for a direct decomposition of its main sources of variation, from customers arrival to products purchase. We identify life-cycle dynamics and heterogeneity across campaigns and products as the main sources of variation: these allow us to provide the best predictions from a statistical standpoint, which outperform machine learning alternatives in out-of-sample accuracy. Our model thus enables flash sales retailers to learn about the performance of new products in a few hours and to update prices so as to better match supply and demand forecast and improve profits. We simulate our forecasting and optimization procedures on several campaigns including thousands of products and show that our model can successfully separate popular and unpopular products and lift revenues significantly.

Key words: Online retail, Shopping funnel, Customer behavior, Responsive pricing.

1 Introduction

Online retailing has become a large sales channel with giant players such as Amazon and continues to grow at double digits (Ben-Shabat et al. 2015). Online retailers have adopted a variety of business models, from owning inventory and selling it online, to being marketplace players where they sell items belonging and stocked by suppliers, and capture a commission in the process. One particular form of

*IESE Business School, University of Navarra, Barcelona, Spain, valbeniz@iese.edu. V. Martínez-de-Albéniz's research was supported by the European Research Council - ref. ERC-2011-StG 283300-REACTOPS and by the Spanish Ministry of Economics and Competitiveness (Ministerio de Economía y Competitividad) - ref. ECO2014-59998-P.

†DatActionS, aplanas@datactions.com.

‡IESEG School of Management (LEM CNRS 9221), Lille/Paris, France, s.nasini@ieseg.fr.

operation that has developed is the *flash sales business model* where the retailer sells a set of items that belong to a third party for a limited time window, at a deep discount. For these third parties, this serves the purpose of *salvaging inventory* that cannot be sold otherwise. This is especially useful for apparel fashion products that could not be sold during the season: flash sales is a way of having a *virtual outlet channel* where merchandize is sold without damaging the brand and cannibalizing the current collection. The largest players in this space are Vipshop, Chinese leader with revenue of USD 11.2 billion in 2017 (Shao 2013, Vipshop 2018); Vente Privée, European leader with a revenue of EUR 3.1 billion (Agnew 2017); and American players such as Rue La La, Zulily or Gilt.

Flash sales players specialize in selling a large variety of goods for a short period of time, with constant changes in the offer. It thus can be seen as an extreme form of e-commerce, where all products are innovative in the sense of Fisher (1997): one would expect these retailers to be responsive in their operations. Unfortunately, most of them do not have flexibility in the quantity of goods they sell. In some cases, the inventory is sitting at the retailer's premises and cannot be replenished; in some others (including the setting that we study here), inventory is owned and located at the third party brand that is willing to run a flash sales campaign. The only degree of freedom is in their pricing, but due to the short time windows, as of today few of them do adjust prices of their products during the selling campaign. At the same time, flash sales retailers operate online and thus are able to know in detail how consumers interact with products displayed on the web. In summary, they face an extremely volatile demand but can potentially use detailed shopping patterns to take better data-driven decisions to improve operations.

The purpose of this paper is to first build an aggregate demand model at the stock-keeping unit (SKU) level, and then suggest re-optimized prices after a short period of time. This can be seen as a pricing version of quick response (Fisher and Raman 1996, Caro and Martínez-de Albéniz 2015). We focus on settings where there are no customers before nor after the flash sales campaign, so our work is not applicable to permanent stores that run limited-time discounts, such as Amazon. Our objectives and focus are shared with Ferreira et al. (2015), who optimize prices at Rue La La, an American flash sales retailer.

Our work takes a different path to achieve the objective. In contrast with Ferreira et al. (2015), we use clickstream data – all actions from customers on the web– which provides a richer and more complex data pool. Consequently, our model does not provide a direct prediction of demand for each product, but considers a broader set of customer decisions: we build a hierarchical model for demand that predicts the different decisions that visitors take during the shopping process. Specifically, we forecast the probabilities that customers advance in the shopping funnel, in three steps: between arrival to the website and visit to a particular campaign; between visit to a campaign and interest in a particular product; and between interest in a product and purchase of that product. This approach efficiently separates campaign-driven effects (e.g., how well-known is a brand) from product consideration issues (e.g., how attractive it is) from actual purchase decisions where price is observed (e.g., cheap vs. expensive). The different layers provide fairly accurate demand predictions with a limited number of parameters, given the high amount of uncertainty in demand. They also separate in each step the general effects related to seasonality, from campaign life-cycle factors (e.g., whether it just started or it is about to finish), from campaign and product effects. Finally and most importantly, because higher layers in our model aggregate information across products, our model provides faster learning (see Bernstein et al. 2018 for a similar idea), which is especially useful when products sell for a short period of time. For instance, even when there are no purchases for a given product yet, there surely are clicks to view the product's information, so it becomes possible to predict a purchase rate, in the same way Bakshi and Peura (2016) predict occurrence of disasters based on the occurrence of minor safety incidents.

A hierarchical perspective like ours thus has several benefits compared to a direct approach that tries to estimate sales without considering funnel variables. First, it provides insights about the demand process, which allow the retailer to improve decisions such as pricing among others. Second, although the statistics literature provides different views on whether aggregate vs. disaggregate forecasts are more accurate (Dunn et al. 1976), our aggregate-to-disaggregate approach improves predictive accuracy, and in particular performs better than direct (disaggregate) machine learning techniques, such as random forests used in Ferreira et al. (2015). Moreover, our approach is able to include large number of fixed effects, so that it is possible to learn quickly from new products. In contrast, it is computationally prohibitive to include categorical variables for most machine learning techniques, so learning needs to be contained to product-specific information, thereby missing higher-level variation and leading to slower learning. As a result, our method is able to improve prediction accuracy by learning faster, provides accurate predictions about the future, and compares favorably against standard machine learning techniques.

Our hierarchical model uses aggregate clickstream metrics, which provide a balance between granularity (down to the SKU level), simplicity (if this was not necessary, machine learning techniques would be an interesting alternative) and tractability (to later apply price optimization). It is validated with real data from a flash sales retailer, using hourly events during 3 years. Hence, besides a modelling contribution, our paper also makes an empirical contribution by documenting (i) the breakdown of uncertainty in the sales of fashion goods in the different shopping process steps; and (ii) the predictable elements of demand, and specifically those related to the life-cycle for short selling windows. One of the key findings from the empirical validation is that product heterogeneity is very strong. In the presence of inventory constraints, such product uncertainty results in many products selling zero units, while many others have a sell-out of 100% and lost sales. It is thus critical to obtain an early indicator of product attractiveness, which can be used to set prices so that supply and demand are more closely matched.

With such early signs of success, we are able to build more accurate demand models and can study the pricing problem for the retailer. We analytically show that under certain conditions there is a unique price that maximizes the retailer expected profit. Optimal prices are increasing in product attractiveness and decreasing in inventory. We then apply price optimization to our data in several campaigns chosen randomly and find that learning from early clicks is able to improve expected profits significantly. The increase is most significant for items for which the predicted sell-out rate is close to

100%, i.e., for those items where the forecast demand is higher than the available supply. Our paper thus provides the methods for prescriptive analytics to improve flash sales retail operations.

While pricing is one actionable way to improve performance, there are other potential improvement directions stemming from our hierarchical model. Indeed, our decomposition of the funnel transitions provide the retailer with valuable information about customer responses to the different elements related to a product. For instance, to evaluate whether the product description and image quality is adequate, a critical metric to follow is the conversion between campaign visits into product clicks, but one should also account for differences across product categories and moments of the day: our model is able to detect the under-performing products for which corrective actions can be applied, e.g., new imaging. As a result, if the retailer wants to increase sales of a particular product, our model will distinguish the effect of campaign design (was the campaign sufficiently attractive? if not, it can be promoted more), the effect of product display (was the the product considered? if not, images or placement can be changed), and the effect of product features (was the price adequate? if not, run our pricing procedure).

The rest of the paper is organized as follows. §2 reviews the relevant literature. §3 describes the context and data. Our demand model is presented in §4 and is estimated with real data in §5. Pricing optimization is studied in §6. §7 concludes the paper. Proofs of our analytical results and supporting tables and figures are contained in the Appendix.

2 Literature Review

This work is connected to several streams of literature, mainly online demand forecasting, including discrete choice models across multiple stages; and quick-response models especially with respect to price optimization.

There is already a relatively large amount of research devoted to the use of online data for demand modelling (see Bucklin and Sismeiro 2009 for a review). Some empirical literature considers longitudinal individual behaviors, see e.g., Moe and Fader (2004a) or Moe and Fader (2004b), but most of the existing work (as ours) focuses on total click or sales data, often disaggregated by product and time period. The main objective in these papers is to model the demand process and in particular to identify predictors that can be used for decision-makers. Moe and Fader (2002) study how pre-launch orders can help predict post-launch orders, using the Bass diffusion model and a Weibull-gamma mixture. More recently, Huang and Van Mieghem (2013) provide a similar analysis. Huang and Van Mieghem (2014) investigate how visits in non-transactional web sites can be used to predict offline sales. Gallino and Moreno (2014) study the interaction of online and offline sales channels. Cui et al. (2018) examine how the information obtained in social media can help on predicting sales using machine learning techniques. Finally, the closest work to this paper is Ferreira et al. (2015): they use machine learning tools to forecast sales and perform price optimization at Rue La La. While the stated goal is the same as ours, the main difference is that they only consider sales metrics (not visits or other layers as we do) within a single campaign. As a result, there is no interaction between campaigns, while we consider the platform's joint evolution with substitution across campaigns, which we find to be very relevant. Their optimal pricing strategy balances demand across products (thus moving demand away from under-supplied items to over-supplied ones), while in our case pricing for each item does not affect substitution patterns much, and thus obeys to a supply-demand matching logic.

While our empirical results are run on clickstream aggregate metrics, we model shopper choices at the individual level. Anderson et al. (1992) or Train (2009) provide excellent reviews of multiple choice models, most of whom rely on the availability of data that match individual characteristics to the products they purchase. Our approach is thus similar to most of the previous economic literature, which estimates product demands from product-level data (Bresnahan 1987, Feenstra and Levinsohn 1995, Berry et al. 1995). We proceed by designing a hierarchical modeling approach, which combines the information available at different points of the online shopping funnel in a sequence of conditional distributions. This relies upon the idea of multiple choices across different stages and upon product-level and aggregate consumer-level data. Hierarchical models are particularly suitable for a fine-grained decomposition of the different sources of data variability, allowing for high flexibility in the inclusion of conditional information (Raudenbush and Bryk 2002). Arora et al. (1998) was one of the first to introduce the hierarchical modeling approach in the context of consumer behavior, opening to a wide range of possibilities for consumer models encoding the information of passing from an option choice to the next by multiple layers of conditional probability. This is related to the concept of marketing funnel in an online context, see e.g., Wiesel et al. (2011) or Liaukonyte et al. (2015). A major difference between our work and these marketing papers is that we focus less on customer behavior (because our focal interest is on aggregate demand on individual products) and more on the detailed, operational management of the products, for which we need to introduce campaign, product and SKU effects that keep changing over time; consequently, our model includes a large number of fixed effects useful for better predictions, as opposed to measuring average effects of simpler covariates.

Some literature exists in the interface between discrete models for sequential decisions and hierarchical modeling (Wooldridge 2010). The majority of these works characterizes individual choices among finite sets of mutually exclusive options (Block et al. 1960). The *elimination by aspects* (Tversky 1972) is one of the most popular frameworks for this purpose. It is based on the sequential elimination of options not having given attributes, until only one option remains. Optimization of more general *consider-then-choose* processes have been analyzed recently by Wang and Sahin (2016), Feldman and Topaloglu (2015) and Aouad et al. (2015). In our context, sequential decisions also take the shopper through a search, but in a well-defined process where purchase can only occur after obtaining more information about the item, which can only happen after a given campaign has been viewed.

Finally, our optimization logic is grounded on the practices of postponement and Quick Response (QR). The original idea of QR has been proposed to learn from early product sales to refine the forecast of demand and hence optimize the quantity to be produced (Fisher and Raman 1996, Iyer and Bergen 1997, Fisher 2009, Caro and Gallien 2010, Gallien et al. 2015). In our context, inventory is given before a campaign, so that price is the only possible operational change. As a result, our problem resembles the pricing problem with an inventory constraint (Van Mieghem and Dada 1999, Federgruen and Heching 1999, Chan et al. 2004). Although there are similarities with dynamic pricing (Bitran et al. 1998, Smith and Achabal 1998, Bitran and Caldentey 2003), in our case the campaign is short and the number of price changes is limited. Cases in which business constraints prevent sellers from conducting extensive experimentation have been studied by Cheung et al. (2017), also in the context of dynamic pricing. Similarly, Ferreira et al. (2018) study Thompson sampling with inventory constraints and a limited number of price changes.

3 Description of the context and data set

The context of study is that of a *virtual outlet* for *flash sales*, an online business model that has only been studied recently (Ferreira et al. 2015). It is quite different from standard e-commerce where a permanent store provides a discount for a period of time and then goes back to the regular price. Instead, a flash sales retailer's main activity is to organize sales events of limited duration that connect two audiences and collect a margin in each transaction. On the one hand, it builds a large user base of individual online consumers, that has access to exclusive deals from famous brands, at deep discounts. On the other hand, it works with many brands, for which it organizes dedicated short campaigns, usually lasting between 3 and 7 days and targeting the firm's large customer base, invited via email and mobile alerts to shop.

We interacted with a leading online flash sales retailer active in several large Western countries. We obtained a comprehensive data set that included all the clicks of customers in four countries over three years (2013-2015), as well as stock availability information. In particular, we had access to visits made to the firm's platform (through web, mobile or app), either to the category or individual products, and purchases, with their respective time stamps (detailed by hour). Given the size and complexity of the data, we focus our study on one single country, although all our results can be replicated in any other market. Table 1 provides an overview of the selected data set in terms of the number of clicks and the units purchased for each year.

Online shopping in the described flash sale retailing context unfolds through a sequence of customer decisions. A customer walks through a sequence of stages from the moment she is informed about a new campaign (notification typically email or app push) until the moment she is able to place a purchase order. The following list summarizes this sequence:

(i) The customer lands on the main page, from which she can access to the different campaigns by clicking to its corresponding image. At this stage only the brand or brands represented in the campaign are displayed, together with the time until the campaign expires.

(ii) Once the customer enters into a given campaign, she visualizes a list of available products. Information about the product is directly visible, through its image and description.

	2013	2014	2015 (Jan. 1 to Dec. 9)
Sum of clicks on campaigns (Arrivals/Entries)	375m	376m	432m
Sum of clicks on products (Interests)	156m	170m	213m
Sum of units purchased (Orders)	4.6m	5.0m	5.6m
Number of campaigns	2,288	2,904	3,747
Number of products	358,470	536,776	1,033,586
Number of SKUs	1,104,539	1,478,311	2,313,009

Table 1: Descriptive statistics of the clickstream data set.

(iii) She can click into an individual product to obtain further details on the product (availability by size for example) or she can add it to her shopping cart for reviewing it later.

(iv) Once the customer has selected all the products that are interesting to her, she can remove some of the selected items and check out to place the order.

After an order has been placed, the items arrive at her home in a few days or weeks. Items can be returned, but at a charge (shipping cost), so the rate of returns at this firm is low.

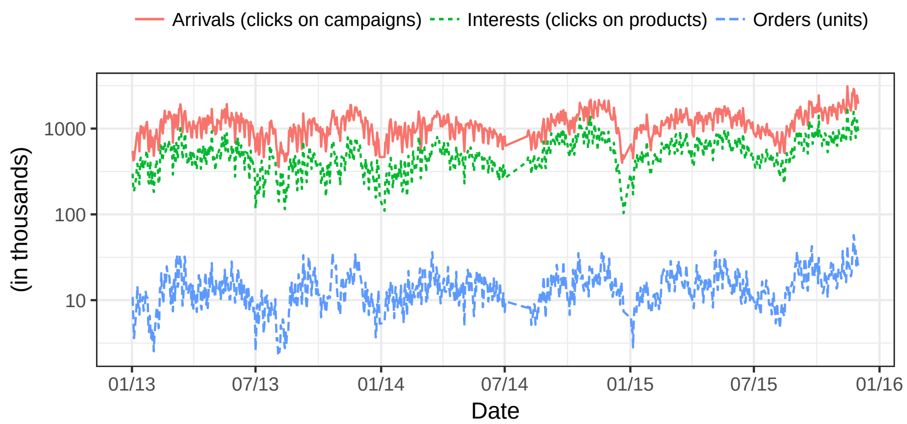


Figure 1: Time dynamics of the total web arrivals and product units sold, per day.

Although the available clickstream data would allow us to model shopping behavior at the customer level, we focus here on aggregate clicks and purchases over all users on an hourly basis (a time window comprising 25,752 hourly observations), for each of the campaigns, products and SKUs. Indeed the focus is to provide aggregate predictions of the demand at the SKU level so as to better match that demand to available supply. There are two main approaches to provide such demand forecasts. We can resort to a direct SKU-level demand forecast exercise, as in Ferreira et al. (2015), using as covariates the features associated with each observation (campaign, product and SKU characteristics). The alternative, that we develop here, is to forecast, in a hierarchical fashion, higher level variables first (arrivals to the platform), and then forecast the conversion probabilities to a lower level (e.g., from arrivals to campaign visits, or from campaign visits to product clicks). While the former has the advantage of simplicity, it has the drawback of missing possible shocks related to higher levels, such as a negative shock on the entire campaign that reduces sales for all the corresponding SKUs. A priori, the use of higher level information should be quite informative: Figure 1 illustrates the strong synchronization between high-level indicators (arrivals to the platform) and low-level ones (units ordered). We reflect this feature in our analytical model as explained next, and compare later in §5.4 direct vs. hierarchical forecast procedures.

4 The AEIO Model

Let us first define the notation for the model. Throughout the paper all vectors will be denoted by boldface characters and assumed to be column vectors. Sets are represented in calligraphic letters:

- $\mathcal{T}$ set of time periods measured in day-hours $T = |\mathcal{T}|$ (e.g., April 20 2015, 18:00).
- $\mathcal{J}$ set of products with $J = |\mathcal{J}|$.
- $\mathcal{C}$ set of campaigns with $C = |\mathcal{C}|$.

Because any product $j \in \mathcal{J}$ belongs to a unique campaign $c \in \mathcal{C}$, we let $c(j)$ be this unique campaign. We also define the relevant time periods for campaigns, calling $t_{\star c}$ the time in which the campaign c starts (the time at which it is available to some customers, e.g., typically premium customers), t_{0c} the time in which the campaign c open to all (after which most of the traffic enters), and $t_{\dagger c}$ the time in which the campaign c ends. We call the period between $t_{\star c}$ and t_{0c} the *pre-opening*, and the period between t_{0c} and $t_{\dagger c}$ the *regular campaign*. About 25% of campaigns had a pre-opening, but these percentage increased in 2015 and it is now common to run a pre-opening. We also note by $\mathcal{C}(t)$ the set of campaigns active at time t, by $\mathcal{J}(c)$ the set of products of the campaign c, and by $\mathcal{J}(c, t)$ the set of products belonging to the campaign c that are available at time t, i.e., not sold out.

We model a collection of dependent variables that determine the aggregate shopping behavior of the firm's customers, and several independent variables, which are used as covariates to account for known sources of variation. The dependent variables are:

- A_t Arrivals to the web page at time $t \in \mathcal{T}$, i.e., total number of customer clicks.
- E_{ct} Entries into a given campaign $c \in \mathcal{C}$ out of all clicks at time $t \in \mathcal{T}$. Note that $A_t = \sum_{c \in \mathcal{C}(t)} E_{ct}$. Indeed, we do not have total platform traffic, so we construct it from campaign clicks, and hence ignore possible landings to the platform that do not visit any campaign.
- I_{jt} Interests in a product, measured by the number of add-to-cart or product views for product $j \in \mathcal{J}$ at time $t \in \mathcal{T}$. When a customer both views and adds to cart an item, we count it only once. This variable is used by the retailer and is considered to be more robust than only considering add-to-carts, or only product views.
- O_{jt} Ordered units of product $j \in \mathcal{J}$ sold at time $t \in \mathcal{T}$.

Exogenous variables that influence A_t, E_{ct}, I_{jt} and O_{jt} are denoted $\boldsymbol{X}_t^A, \boldsymbol{X}_{ct}^E, \boldsymbol{X}_{jt}^I, \boldsymbol{X}_{jt}^O$, and the model parameters associated with them $\boldsymbol{\theta}^A, \boldsymbol{\theta}^E, \boldsymbol{\theta}^I, \boldsymbol{\theta}^O$ respectively. They generally contain time-dependent properties (weekday-hour, which capture the impact of push email marketing events that usually happen at a certain time of the day, week of the year, or time trend; note that we cannot include the number of active campaigns because it has high collinearity with the time trend), campaign properties (attributes related to visible information including average list price which is a proxy for brand strength), product properties (visible attributes such as product family, list price, number of sizes available, etc.) and interactions (hours since a campaign opened, day of the campaign; effects related to the time constraint put on the shoppers). These features are often used in the literature and capture the elements that drive the shopping process, i.e., intrinsic shopping seasonality, tangible product features (Ferreira et al. 2015) and appetite for novelty (Caro et al. 2014). We verify that they indeed are very significant in explaining the variation in sales performance. Tables 9-11 in the Appendix provide specific details on these covariates.

Given the complexity of the shopping process (very high number of campaigns and products, which change all the time), we design a hierarchical model with different aggregation levels, which we call *layers*. Our hierarchical model consists of four layers of information, one for each of the dependent variables. For ease of notation we are going to call them layers A, E, I and O, respectively. We first model the arrival process of customers to the platform, and then the branching process of those arrivals into entries into campaigns, which then generate interest on products and finally results in realized sales of the SKUs related to those products. In other words, we model the shopping funnel where events in each stage happen independently of other stages. The hierarchical structure of this sequence of decisions can be captured by four sub-models, which are based on a collection of conditionally independent distributions.

The first layer A accounts for the number of arrivals, i.e., clicks produced in the platform. For each time $t \in \mathcal{T}$, a Poisson-Gamma model is adopted to mirror the count process generating the observed number of clicks A_t. This is a mixture of distributions, which was first derived by Greenwood and Yule (1920) to account for over-dispersion commonly observed in discrete or count data. Ehrenberg (1972) first proposed the usage of Poisson-Gamma distribution as a model for consumer behavior, which resulted in a number of subsequent contributions on this vain (e.g., Poisson-Gamma models applied to grocery store trips in Frisbie 1980, Weibull-gamma models in online retailing in Moe and Fader 2002). Its popularity is due to the closed-form expression of the marginal distribution, which is known to be negative binomial, and for its flexibility for modeling data over-dispersion. For regression purposes, this distribution can be specified in terms of its mean and variance, which are then related to the explanatory variables X_t^A:

$$E[A_t] = e^{\boldsymbol{\theta}^A \boldsymbol{X}_t^A} \text{ and } Var[A_t] = E[A_t] + \frac{E[A_t]^2}{\alpha}, \tag{1}$$

The parameter α provides the required extra degree of freedom to allows for over-dispersion when $\alpha < \infty$ (indeed the empirical distribution of the number of arrivals to the platform supports the presence of a substantial over-dispersion). Note that we tested other model specifications such as lognormal or Poisson, but the negative binomial was found to be the one that performed best (it resulted in the highest AIC).

The second layer E explains how total arrivals are distributed across the different campaigns. In the data, we observed that launching a new campaign reduces the visits to other campaigns, so our model needs to internalize substitution effects across campaigns. So we use a choice model, where a visitor picks one among $\mathcal{C}(t)$ available campaigns (if one individual picks more than one campaign, then we consider that it generates several arrivals; given the large number of individuals, choices can still be regarded as independent). As long as the probability of each click is regarded as statistically independent of the rest of clicks, visits to campaigns follow a multinomial distribution. Specifically, the probability of this conversion is

$$e_{ct} := \frac{e^{\boldsymbol{\theta}^E \boldsymbol{X}^E_{ct}}}{\sum_{c' \in \mathcal{C}(t)} e^{\boldsymbol{\theta}^E \boldsymbol{X}^E_{c't}}}, \qquad \text{for } c \in \mathcal{C}(t),\ t \in \mathcal{T}. \tag{2}$$

Note that we do not include an outside option here because $A_t = \sum_{c \in \mathcal{C}(t)} E_{ct}$. The expected number of visits to campaign c at t is equal to $\mathbb{E}[E_{ct}] = e_{ct} A_t$ and the expected visit share is e_{ct}. This model with substitution fits the data better than a model without substitution (where E_{ct} is modeled as a binomial distribution with a branching probability that depends on $\boldsymbol{X}^E_{ct}$ and A_t tries).

The third layer I describes the conversion mechanism of the clicks to each campaign into interest in specific products within the campaign. Such interest is measured by either visits to the product information page or clicks to add the product into the shopping cart (a pre-selection step before confirming the purchase). This variable was suggested by the analysts of the company and was found to be more robust than only visits to the product information page or only pre-selection before purchase clicks. We model the conversion step by a binomial branching. Note that we also tested models with substitution but this required a significantly more involved estimation effort and had very similar performance to the proposed binomial branching, from a prediction standpoint. The probability of this conversion is

$$i_{jt} = \frac{e^{\boldsymbol{\theta}^I \boldsymbol{X}^I_{jt}}}{1 + e^{\boldsymbol{\theta}^I \boldsymbol{X}^I_{jt}}}, \qquad \text{for } j \in \mathcal{J}(c),\ c \in \mathcal{C}(t),\ t \in \mathcal{T}. \tag{3}$$

The fourth and last layer O encodes the final conversion mechanism from product interest into actual product purchases. Again a binomial distribution is a natural representation of such probability as long as each purchase is regarded as independent from others, which is reasonable given the large number of individuals. The probability of this conversion is

$$o_{jt} = \frac{e^{\boldsymbol{\theta}^O \boldsymbol{X}^O_{jt}}}{1 + e^{\boldsymbol{\theta}^O \boldsymbol{X}^O_{jt}}}, \qquad \text{for } j \in \mathcal{J}(c,t),\ c \in \mathcal{C}(t),\ t \in \mathcal{T}. \tag{4}$$

5 Empirical Results

In this section, we estimate the model described in §4 and in particular highlight the value of early information to predict future behavior.

5.1 Estimation and Cross-Validation Approach

The numerical results presented in this section are based on maximum likelihood estimation. Indeed, with our specification, the log-likelihood is concave in the parameters $\boldsymbol{\theta}^A, \boldsymbol{\theta}^E, \boldsymbol{\theta}^I, \boldsymbol{\theta}^O$ because of known properties of generalized linear models (McFadden 1974).

We estimate the model using a subset of the data (the training set) and then validate it by generating out-of-sample forecasts that are compared to actual realizations. Since our data has observations between January 1, 2013 and December 9, 2015, we opted to split the data in two, obtaining a training set with data from January 1, 2013 to September 30, 2015 (we will call this set Σ_0), while the out-of-sample forecasts are performed in the data between October 1, 2015 and December 9, 2015 (we will call this set Σ_1). Choosing different time partitions led to similar results. For layers I and O, due to the large amount of products, we opted for taking a random sample (denoted by Σ_p) of 10,000 products: hence, the training set, was the data in $\Sigma_p \cap \Sigma_0$, and the out-of-sample results were considered in the entire Σ_1, i.e., no sampling in the validation set.

We consider three model specifications with increasing levels of complexity for layers E, I and O:

- *Baseline models* $E1$, $I1$ and $O1$ involving life-cycle variables, i.e., related to the age of the campaign, but no other campaign or product features (i.e., no heterogeneity).

- *Complete models* $E2$, $I2$ and $O2$ involving explanatory variables for campaign (in $E2$) or product (in $I2$ and $O2$) attractiveness captured by attributes, besides the life-cycle variables already present in $E1$, $I1$ and $O1$; note that these attributes are selected from tangible characteristics such as category types, variety or price. See Tables 9-11 in the Appendix for more details.

- *Saturated models* $E3$, $I3$ and $O3$, involving campaign (in $E3$) or product (in $I3$ and $O3$) labels (dummy variables) instead of their attributes, besides the life-cycle variables already present in $E1$, $I1$ and $O1$;

For the models without campaign or product fixed effects (baseline and complete models), Σ_0 provided estimators for all the covariates, hence we could provide a forecast out-of-sample, as reported in §5.3. However, for the models with campaign or product fixed effects, we observed the appearance of new features in Σ_1 that were not present in Σ_0, i.e., new campaign, product and SKU identifiers. This forced us to adjust the structure of the out-of-sample validation as shown in §5.3. Namely, there was no adjustment for layer A (training with Σ_0 and validation with Σ_1). But in layers E, I and O, out-of-sample forecasts were created by using the estimates from Σ_0, together with an estimate of the fixed effects just from the pre-opening (see §5.3 for more details). That is, predictions for a new campaign or product combine information about general life-cycle drivers and individual characteristics estimated from a few hours before the campaign opens.

5.2 Estimated Parameters

We report here the results for each layer and for the entire model (unconditional quantity of units sold).

Arrivals. We used a negative binomial specification of (1). This choice is supported by the significant over-dispersion in the model, as the estimated value of α is 5.80. McFadden's pseudo-R^2 is equal to 0.81 (24,017 observations; 220 variables; null deviance 133,955 vs. residual deviance 24,721; AIC 518,954). Given that we are considering hourly observations in our model, this high accuracy indicates that it is possible to generate high quality daily and weekly visit forecasts.

It is worth noting that the coefficient for the linear time trend (measured in years) is equal to 0.0996 (p-value below 10^{-15}), which means that traffic is growing at 10% per year. Figure 2 reports the estimated effects for seasonality. As we can observe, there are strong variations during the week and during the year. In particular, as expected arrivals severely drop at the night and peak during the morning (most campaigns open at 7:00 am).

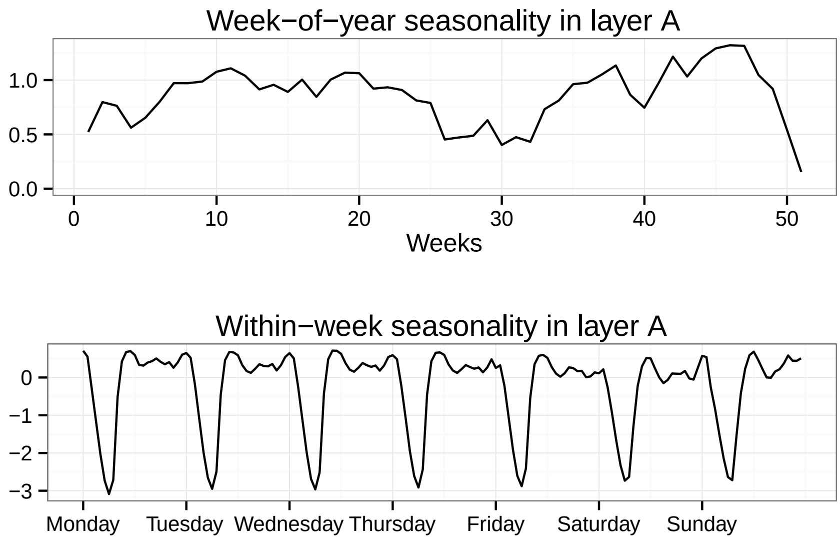


Figure 2: Estimated seasonality effects in model A, for (1).

Entries into a campaign. The branching of arrivals into campaigns is based on the conditional multinomial model (2). As previously mentioned, the substitution structure is necessary to account for the fact that, when a new campaign is launched, the number of visits to older campaigns drops significantly.

We identified two sets of variables that explained the differences of visits across campaigns. The first one includes campaign life-cycle factors. Namely, the amount of clicks to a campaign varies strongly during its duration, peaking at the moment of launch t_{0c} and decreasing afterwards. The second set of variables includes campaign characteristics, which are associated to visible information for the shopper, such as whether the products come from an expensive brand, or whether the campaign includes a single brand or multiple ones, see Table 10 for more details. They are important because we find very successful campaigns next to others much less popular. Note that we do not add seasonality variables because, by construction, the sum of campaign shares is equal to one and thus does not vary over time.

The three model specifications at layer E are: model $E1$ (baseline model), where we only included life-cycle variables but no other campaign feature (i.e., no heterogeneity); model $E2$ (complete model), where we added variables for campaign attractiveness captured by attributes (including number of products, average price, whether it had more than one brand included, whether the campaign had a pre-opening, etc.); and model $E3$ (saturated model), where, besides life-cycle variables, campaign attractiveness was included through a campaign dummy instead of attributes. Models $E1, E2$ and $E3$ are compared to a null model $E0$ without life-cycle nor characteristics, i.e., one where all campaigns have the same attractiveness (i.e., e_{ct} is one over the total number of active campaigns). This comparison is measured by McFadden's pseudo-R^2 and AIC. Table 2 contains the results of the estimation.

		Model		
		E1	E2	E3
Life-cycle	In pre-opening	-0.66***	-1.73***	-1.69***
	$t - t\_0c$ (hours)	-0.09***	-0.09***	-0.09***
	$t - t\_{\star}c$ (hours)	0.07***	0.04***	0.03***
	Day 1	1.46***	0.64***	0.68***
	Day 2	0.82***	0.11***	0.14***
	Day 3	0.48***	-0.05***	-0.03***
	Last day	-2.77***	-2.88***	-3.14***
Campaign		None	Duration FE	Campaign FE
	With pre-opening		0.89***	
	Log of average of campaign list prices		-0.11***	
	Single brand		0.21***	
	Time trend (years)		0.19***	
	Number of observations	1,151,746	1,151,746	1,151,746
	Number of variables	7	17	7,953
	Log-likelihood	-3,736,469,200	-3,591,148,588	-3,399,893,797
	Pseudo-R^2	44.6%	63.7%	88.7%
	AIC	7,472,938,415	7,182,297,209	6,799,803,498

Table 2: Estimates of model (2), for the three specifications $E1, E2$ and $E3$. P-values: *** ($< 0.1\%$), ** ($< 1\%$), * ($< 5\%$).

As we can see in the table, model $E1$ performs much better than a model without covariates, as its AIC decreases and pseudo-R^2 increases to 0.446. In other words, life-cycle factors are of primary importance in the prediction of campaign clicks E_{ct}. We see that traffic during the life of the campaign varies strongly: it decays over time, and in particular the first day has the highest traffic ($e^{1.46} = 4.3$ times higher). The first hours of a campaign also exhibit strong decay (-9% per hour since start, up to a maximum of 12h). In further refinements of our models, we saw that this decline was only present for campaigns without pre-opening, while there was a slight increase in traffic for those campaigns with a pre-opening: the lack of decay seems to occur because the more impatient customers are also those that visit during the pre-opening. We also see that attractiveness is significantly lower during the pre-opening ($e^{-0.66} = 0.51$ of the attractiveness during the regular campaign), which reflects that the pre-opening is only available to a subset of customers. These findings remain similar across models $E2$ and $E3$.

While model $E1$ is much better than $E0$, at the same time there is still a large amount of residual uncertainty. This uncertainty can be further reduced by adding campaign characteristics as covariates. Specifically, model $E2$ decreases AIC and increases pseudo-R^2 to 0.637, i.e., clears about a third of the uncertainty in model $E1$. We find that conversion increases for lower-price brands and focused campaigns. Note that at this point only visible characteristics from the campaign choice stage are included in the model. Specifically, we include average list prices as a proxy for brand price positioning, but miss other relevant campaign features such as variety, actual prices or brand perception.

Campaign heterogeneity is not fully captured unless campaign dummies are present, in model $E3$: again AIC decreases strongly and leads to a pseudo-R^2 of 0.887. This improvement occurs at the expense of a large number of coefficients (7,953 instead of 17). Interestingly, we find that campaign dummies may vary by orders of magnitude (which amount to visits to campaign easily varying by a factor of 1,000); this variation can not be captured in model $E2$ via campaign attributes, which explains why $E3$ performs much better.

Interest in a product. The conversion from campaign clicks to product visit or add-to-cart is based on the conditional binomial model (3). Because conversion propensity may vary over time, so we include seasonality variables in our specifications. Again we constructed three model specifications: model $I1$ (baseline model) included only seasonality and life-cycle variables, model $I2$ (complete model) seasonality, life-cycle variables and product attributes, while model $I3$ (saturated model) had seasonality, life-cycle variables and product dummies. These are compared to a model $I0$ without seasonality, product dynamics nor heterogeneity. We estimated the model parameters $\boldsymbol{\theta^I}$ using a random subset of the original product-hour observations from the training set, which resulted in over 1.6 million observations. The estimates of the binomial model are presented in Table 3, together with estimation details.

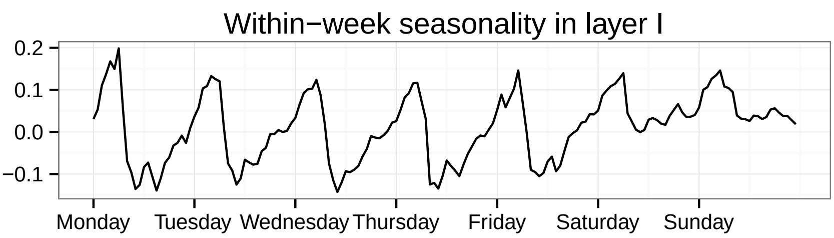


Figure 3: Estimated seasonality effects in specification $I3$, for (3).

		Model		
		I1	I2	I3
Life-cycle	In pre-opening	0.11***	0.10***	0.14***
	$t - t_{0c}$ (hours)	-0.01***	-0.01***	-0.01***
	$t - t_{\star c}$ (hours)	-0.01***	0.003***	-0.001***
	Day 1	0.14***	-0.02***	0.06***
	Day 2	0.08***	-0.04***	0.04***
	Day 3	0.05***	-0.07***	0.01***
	Last day	-0.27***	-0.44***	-0.45***
Product		None	Product category FE	Product FE
	With pre-opening		0.12***	
	Log of average of campaign list prices		-0.38***	
	Log of number of campaign products		-0.84***	
	Log of product list price		0.27***	
	Log of number of product sizes		0.05***	
	Ratio of price to list price		0.08***	
	Ratio of price to average campaign price		0.08***	
	Number of observations	1,696,331	1,696,331	1,696,331
	Number of variables	175	322	9,680
	Log-likelihood	-21,591,350	-20,872,368	-19,930,067
	Pseudo-R^2	0.4%	24.0%	55.0%
	AIC	43,183,051	41,745,379	39,879,493

Table 3: Estimates of model (3), for the three specifications $I1, I2$ and $I3$. P-values: *** ($< 0.1\%$), ** ($< 1\%$), * ($< 5\%$).

In contrast with model (2), the conversion propensity from campaign visits to product interest has strong within-week variations, with peaks of conversion occurring during the evening and at night, see Figure 3. Indeed, this seems to be the time where shoppers have more time to retrieve product information, while they may be busy at work in the middle of the day. Also, life-cycle is much weaker when it comes to product interest: while the pseudo-R^2 of model $E1$ was quite high, that of model $I1$ is 0.004, much lower. This indicates that there is again a certain life-cycle in the conversion of campaign visits into product interest, but much less pronounced compared to the life-cycle of the campaign. Specifically, conversion tends to be higher during the pre-opening (if there is one) and slightly decay over time, although reaching very low conversions towards the end of the campaign. Given that i_{jt} can be interpreted as the likelihood of putting product j into the consumer's consideration set, this can be interpreted as shoppers having larger consideration sets in the beginning of a campaign, where they seek information; later in the campaign consideration sets become smaller as customers may have received information about products earlier in the campaign, and may be more selective.

The predictions of models $I2$ and $I3$ are significantly more accurate, with lower AIC and pseudo-R^2 of 0.240 and 0.550. We hence find significant heterogeneity across products in the conversion between campaign visits and product interest, even within the same campaign. In model $I2$, the likelihood to seek more information about a particular product increases in campaigns with lower list prices and reduced variety (hence lower competition for any product within the campaign). In addition, products with higher list prices and higher actual prices (actual price to list price, and actual price to average campaign price), and increased product complexity (hence more sizes to choose from) trigger more information requests: these goods are also those with less discounts and with more uncertainty about fit, thus generating more need for inspection from consumers; however, as seen in the next paragraph, although interest is higher, conversion into a purchase is lower. Again, heterogeneity across products is only captured fully by the use of product dummies in $I3$.

Orders of a product. The final conversion from product visit or add-to-cart into actual purchases follows the conditional binomial model (4). We again estimated the model parameters $\boldsymbol{\theta^O}$ using all the SKUs related to the random subset that was selected in the previous layer I (cf. Table 3), i.e., we included all the SKUs (i.e., combination of product and size) for the products for which interest was expressed. To avoid censoring due to availability problems, we only considered observations during the periods where there was positive stock of an SKU, and this is the reason why we do not include inventory as a covariate in the model. As before, we build model $O1$ with seasonality and life-cycle variables, $O2$ with seasonality, life-cycle variables and product attributes and $O2$ with seasonality, life-cycle variables and product dummies, which are compared to $O0$ without seasonality, product dynamics nor heterogeneity. Table 4 describes the results of the estimation.

As with submodel (3), we find that model $O1$ only provides a pseudo-R^2 of 0.010. This means that its predictive power is low even if seasonality and life-cycles in purchase conversion may be significant. In this case, the variation is opposite to that of clicks and interest, namely it increases over the lifetime of the campaign. This means that early in the campaign, customers visit the site and

		Model		
		O1	O2	O3
Life-cycle	In pre-opening	0.07***	0.07***	0.11***
	$t - t_{0c}$ (hours)	-0.02***	-0.01***	-0.01***
	$t - t_{\star c}$ (hours)	-0.05***	-0.04***	-0.04***
	Day 1	-0.45***	-0.40***	-0.36***
	Day 2	-0.16***	-0.12***	-0.12***
	Day 3	-0.12***	-0.10***	-0.08***
	Last day	0.29***	0.34***	0.52***
SKU		None	Product category FE	Product FE
	With pre-opening		0.22***	
	Log of average of campaign list prices		-0.21***	
	Log of number of campaign products		-0.03***	
	Log of product list price		-0.52***	
	Log of number of product sizes		-0.82***	
	Ratio of price to list price		-1.27***	
	Ratio of price to average campaign price		-0.22***	
	Number of observations	3,914,891	3,914,891	3,914,891
	Number of variables	175	322	9,680
	Log-likelihood	-453,566	-418,615	-392,245
	Pseudo-R^2	1.0%	13.1%	22.2%
	AIC	907,481	837,873	803,850

Table 4: Estimates of model (4), for the three specifications $O1, O2$ and $O3$. P-values: *** ($< 0.1\%$), ** ($< 1\%$), * ($< 5\%$).

obtain information about products, but they make a purchase with lower probability, compared to later in the campaign. In particular, the likelihood of making a purchase strongly increases between the first and second day, and makes a very significant jump up of more than 30% in the last day of a campaign (although it is also true that the amount of product interest also reduces considerably during the last day). There are also strong within-week variations, as shown in Figure 4, with peaks in the middle of the day. In other words, while product interest and browsing are weakest during the day, decisions about purchasing are highest at that time. This suggests that shoppers retrieve information during evening and night, but they place their orders during the day.

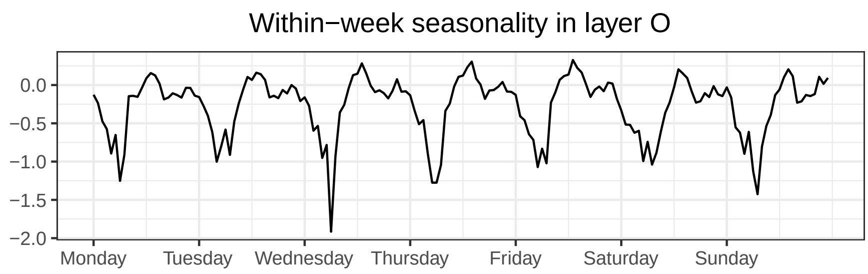


Figure 4: Estimated seasonality effects in specification $O3$, for (4).

Again we find that product heterogeneity is strong and can only be fully captured by models $O2$ and $O3$, with lower AIC and a pseudo-R^2 of 0.131 and 0.222. In model $O2$, we see that purchase conversion is driven by both campaign effects and product features. Specifically, campaigns with lower list prices and less variety tend to have higher conversions. Furthermore, product features matter even more and in the opposite direction to that of layer I: items with lower list prices, higher discounts (compared to the list price and the average price within the campaign) and fewer available sizes have higher conversions, due to higher value and reduced fit risks to consumers. In model $O3$ we see that again including individual product information allows us to increase the goodness-of-fit. One could think that 0.222 is still a poor pseudo-R^2. However, one should keep in mind that O_{jt} is a process made of zeros almost always, and ones sometimes. As a result, it is difficult to improve the fit any further.

Finally, we examined the correlation between product dummies in $O3$ and those from model $I3$: it is small and slightly negative (equal to -0.12) which suggests that the two layers indeed capture two different shopping moments that are unrelated to each other (putting the product into the consideration set and then making a choice). More importantly, this implies that to generate accurate predictions, we need to create separate models for interest and purchase.

5.3 Model Validation

The results presented so far indicate that while it is possible to obtain fairly accurate predictions of total visits A_t, the amount of uncertainty becomes very high at the campaign, product and SKU levels. There are three reasons for this. First, conversion averages are very small because they turn aggregated metrics into lower-level ones - for example from campaigns to products, and from products to SKUs. As a result, the resulting variables are highly volatile. Second, there is a large amount of variation in the behaviors too, and this is captured in our model through heterogeneity in campaigns and products. Third, the process reflects the simultaneous effect of several sources of variability, which cannot be easily reduced to a few model covariates. Table **??** summarizes the breakdown of uncertainty in each of the stages.

		Layer			
		A	E	I	O
	1		44.6%	0.4%	1.0%
Model	2	81%	67.6%	24.0%	13.1%
	3		88.7%	55.1%	22.2%

Table 5: In-sample model uncertainty reduction for each of the stages, measured in pseudo-R^2. The comparison of AIC provides similar insights.

We validate here the quality of the models, through the analysis of in- and out-of-sample predictions. For this purpose, we show the comparison between the hour-level predictions of the model $\hat{Z}$ and their realizations Z, where Z is the hour-level value of $A_t, E_{ct}, I_{jt}, O_{jt}$, in Figures 8-11 in the Appendix. We separate this illustration in two, within sample on the left plots (the training sample set between January 1, 2013 and September 30, 2015) for models A, $E1, I1$ and $O1$ and $E2, I2, O2$, and out-of-sample on the right plots (the validation sample set between October 1, 2015 and December 9, 2015). To complement these figures, we add histograms in the background to inform about the cumulative distribution of observations over the different predicted values. We see in the figures that the prediction accuracy of the models is similar in- and out-of-sample, which suggests that it is adequate for forecasting purposes. However, the in-sample error in models $E1, I1$ and $O1$ is significantly higher than that in models $E2, I2$ and $O2$, respectively.

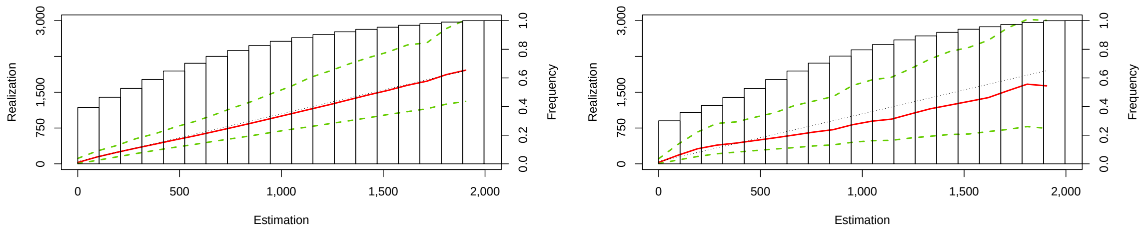
Further, we tested the performance in-sample of models $E3, I3$ and $O3$ (shown below in Figures 5-7) and found that it provided a considerable improvement. The improved performance of models $E3, I3$ and $O3$ compared to $E2, I2$ and $O2$ highlights that there is significant heterogeneity across campaigns and products. In particular, we have introduced campaign fixed effects in the second layer E and product fixed effects in the third and fourth layers I and O, because campaign and product features were not able to capture the high degree of variation: indeed success of fashion goods cannot be predicted a priori (Christopher et al. 2004).

Unfortunately, we cannot use such fixed effects for out-of-sample forecasting, e.g., to predict what will sell from the products launched next week. Given that these dummies are very informative, we here design an alternative approach so that early information about campaigns and products can be used to improve forecasts. We use a well-established idea from Fisher and Raman (1996) or Moe and Fader (2002): by observing early "sales" – any information that suggests that a campaign or a product will be successful – we can improve forecasts significantly. Specifically, we observe the behavior of a certain campaign and product over their first hours, and we estimate the campaign and product dummies, while keeping all other estimators fixed. In the case when there is no information, we simply use as dummy the average of the dummies for those products in the sample that also had no information in the early hours (this is equivalent to regressing the missing dummies against the observed dummies, using lack of observation as an explanatory covariate). For example, when there are no orders in these early hours, we set the product dummy equal to the average of the product dummies for all the products in the sample that did not have receive any early order. After the dummies have been computed, we re-forecast the different indicators in our model with these updated informative dummies.

As suggested by the flash sales retailer, we use pre-opening information (the number of visits and sales of the products during the time periods that were only available to premium users, before it is open to all). Even if users arriving in the pre-opening are structurally different from those that arrive later (because the pre-opening may be restricted to premium customers only), we expect that preferences over the different products should exhibit similarities, and hence this can provide valuable information to detect the success of a campaign or a product during the main campaign. One concern to implement this idea is that, since the number of products is very large, most products sell very small amounts during the first hours of a campaign. Fortunately, even in that case, the layered structure of our model is able to detect higher-than-average clicks on the campaign (E_{ct}) or on the product (I_{jt}), so this information will increase the likelihood of higher-than-average future purchases.

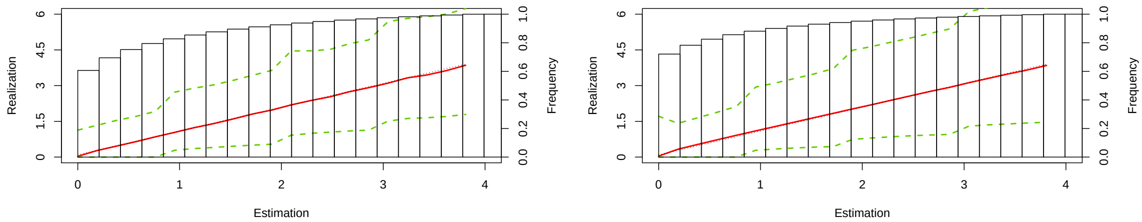
Our method is tested in the period after October 1, 2015, which contains 652 campaigns, 146,479 products and 349,881 SKUs, with a pre-opening. Some of them have extremely low traffic, which could indicate that the pre-opening did not occur in reality, so we filter them out and obtain a final sample of 592 campaigns, 106,527 products and 329,218 SKUs. Figures 5-7 show the results of this re-optimization procedure. We observe that the approach slightly overestimates the prediction of campaign entries, especially for high-traffic campaigns, and underestimates those with low-traffic. Upon closer inspection, this is due to the fact that, to produce a prediction, campaigns that will be launched after the estimation date are unknown (in the sense that their attractiveness is not estimated) and are thus attributed the average attractiveness (mean of the fixed effect for concurrent campaigns), which differs from the true attractiveness. Note that this divergence is not present in model E2, which produces a different attractiveness for each campaign based on attribute differences. In contrast, our procedure delivers excellent results for layers I and O.

To confirm the qualitative insights from the discussion above, we report the distribution of absolute percentage errors (APE) of the item-level prediction. Specifically, for each campaign (in layer E), product (in layer I) and SKU (in layer O), we respectively compute the predicted number of entries, conditional on arrivals (which are an input to the prediction); the predicted number of interests, conditional on campaign entries; and the predicted number of orders, conditional on product interests. We then sum these predictions over all the hours in which the items were sold, and obtain a total prediction $\hat{Z}$, which combines conversion probabilities given by our model with actual number of potential arrivals, entries and interests respectively. We compare $\hat{Z}$ to the actual realization Z, and



(a) Specification E3, in sample (b) Specification E3, out of sample

Figure 5: Comparison between campaign clicks $\hat{E}_{ct}$ from model (2) and actual E_{ct}, conditional on knowledge of A_t.



(a) Specification I3, in sample (b) Specification I3, out of sample

Figure 6: Comparison between product visits or add-to-cart $\hat{I}_{jt}$ from model (3) and actual I_{jt}, conditional on knowledge of $E_{c(j)t}$.

measure the absolute percentage error $|Z/\hat{Z} - 1|$. Given that some forecasts might be very close to zero, there might be extreme underestimations, so the use of mean absolute percentage error (MAPE) is highly affected by those outliers and thus could be misleading. For this reason, we also report the median APE, as well as different quantile levels (indeed there is a large number of SKUs for which there are no sales, thus generating an APE of one, which turns out be the median for layer O across the three models).

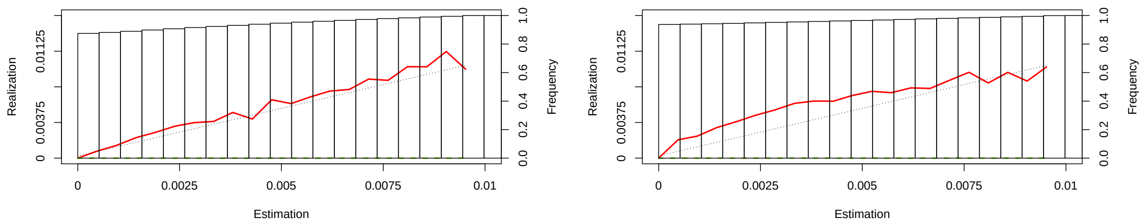
Table 6 contains the results, both in- and out-of-sample. We see that indeed the more sophisticated model 3 performs best compared to models 1 and 2. The difference is more pronounced for layers I and O, with lower errors (lower MAPE for layers I and O, of 0.409 and 1.155 vs. 0.612 and 1.218 for model 2 and 0.806 and 1.505 for model 1 respectively), while qualitatively similar for layer E, as pointed out above (MAPE of 0.401 for model 3 vs. 0.416 for model 2).

5.4 Comparison with other forecast approaches

To further assess the prediction accuracy provided by our AEIO hierarchical model, we compare in this section the overall sales prediction provided by our model, i.e., the expected number of units sold of a particular SKU, against the actual sales figure. We also consider alternative methods used in the literature, specifically the random forest approach developed in Ferreira et al. (2015), which has

			MAPE	Median = Q50%	Q80%	Q90%	Q95%	Q99%
Layer E	In-sample	AEIO - model 1	0.577	0.397	0.840	1.071	1.683	3.841
		AEIO - model 2	0.478	0.351	0.704	0.972	1.322	2.415
		AEIO - model 3	0.006	0.000	0.000	0.000	0.000	0.158
	Out-of-sample	AEIO - model 1	0.558	0.339	0.882	1.082	1.564	3.703
		AEIO - model 2	0.416	0.354	0.570	0.815	0.976	1.991
		AEIO - model 3	0.401	0.256	0.528	0.904	0.993	2.710
Layer I	In-sample	AEIO - model 1	0.931	0.667	1.000	1.508	2.758	6.424
		AEIO - model 2	0.860	0.565	1.000	1.396	2.439	7.014
		AEIO - model 3	0.209	0.000	0.001	1.000	1.000	2.776
	Out-of-sample	AEIO - model 1	0.806	0.598	0.903	1.519	2.498	5.387
		AEIO - model 2	0.612	0.497	0.785	0.970	1.607	3.472
		AEIO - model 3	0.409	0.264	0.562	0.829	1.254	2.842
Layer O	In-sample	AEIO - model 1	1.748	1.000	1.032	2.998	5.819	18.206
		AEIO - model 2	1.453	1.000	1.000	2.089	3.898	13.802
		AEIO - model 3	0.968	1.000	1.000	1.000	2.000	6.500
	Out-of-sample	AEIO - model 1	1.505	1.000	1.000	2.166	4.629	14.215
		AEIO - model 2	1.218	1.000	1.000	1.385	2.730	8.512
		AEIO - model 3	1.155	1.000	1.000	1.115	2.524	8.501

Table 6: In- and out-of-sample forecast accuracy of each layer's conversion probability prediction vs. actual conversion value (layer E on top, layer I in the middle, layer O at the bottom). MAPE is the mean absolute percentage error and Qx% represents the x% quantile of the distribution of absolute percentage error.



(a) Specification O3, in sample

(b) Specification O3, out of sample

Figure 7: Comparison between product purchases $\hat{O}_{jt}$ from model (4) and actual O_{jt}, conditional on knowledge of I_{jt}.

been highlighted as one of the most accurate machine learning methods for forecasting purposes (Cui et al. 2018). The difference between our approach and this latter technique is that random forest is a *direct* forecasting method, in the sense that it uses campaign, product and SKU characteristics for the prediction, but not the realization of traffic variables (arrivals to the platform, entries into campaigns, and interest in products). Hence, it has the advantage of optimizing the model to focus on the variables of interest – orders received – but at the same time misses any relevant information contained in higher layers. In contrast our model estimates conversion across layers, but has access to potentially useful information on these higher layers.

Our implementation of random forest was the following. As Ferreira et al. (2015), we consider a forest of 100 trees, with a maximum depth of 5 layers, and, to ensure a fair comparison, the same covariates that were included in model 2: the age of the campaign, in hours, the hour of the day, the week day, the average campaign price, the number of products within the campaign, the product price, the discount percentage, the ratio of product price to average campaign price (to account for price substitution in the same way as Ferreira et al. 2015). We trained the random forest model with the same in-sample set from before (sample of 10,000 products before September 30, 2015), and generated out-of-sample predictions afterwards on the validation set used to build Table 6.

We thus generate a prediction for each hour and SKU, following the different models, which we then aggregate into a single SKU prediction: the expected number of units sold for each SKU. Because APE penalize predictions that are very close to zero (e.g., if the sales prediction is 0.01, when the actual sale is zero, then APE is one; and when the actual sale is one, then APE is one hundred), we select the 228,666 SKUs for which all forecasting methods provide a prediction of at least 0.5 units sold for the entire sales window (results are similar when the threshold is above 0.1). For these, we report the in- and out-of-sample performance for each of our three models against random forest in Table 7. The table reports forecast accuracy at the SKU level. We observe that the random forest benchmark provides a very satisfactory forecasting approach. It is superior to our model 2, with the same information. This is true both in-sample and out-of-sample, with a similar performance (MAPE of 1.273 vs 1.438 in sample, and MAPE of 1.053 vs. 1.127 out of sample). In contrast, random forest is less competitive that our model 3, which has accounts for individual SKU fixed effects. The difference is very large in sample (MAPE of 1.273 vs. 0.545 for our model 3) because our model 3 is able to account for individual characteristics. Out of sample the difference is less pronounced but still sizeable (MAPE of 0.974 vs. 1.053), which indicates that indeed the use of individual item characteristics from multiple layers, even from a few hours of observation, can significantly improve demand forecasts.

		MAPE	Median APE = Q50%	Q80%	Q90%	Q95%	Q99%
In-sample	AEIO - model 1	2.147	0.911	2.411	4.943	8.201	21.360
	AEIO - model 2	1.438	0.763	1.595	3.233	5.474	12.306
	AEIO - model 3	0.545	0.399	1.000	1.000	1.404	2.532
	Random forest	1.273	0.705	1.317	2.755	4.662	10.692
Out-of-sample	AEIO - model 1	1.837	1.000	1.279	3.441	6.572	19.159
	AEIO - model 2	1.127	1.000	1.000	1.170	2.584	8.455
	AEIO - model 3	0.974	1.000	1.000	1.000	1.791	6.092
	Random forest	1.053	1.000	1.000	1.194	2.536	7.469

Table 7: In- and out-of-sample forecast accuracy of the sales prediction vs. actual realization, for our three models and the random forest benchmark. MAPE is the mean absolute percentage error and Qx% represents the x% quantile of the distribution of absolute percentage error.

6 Price Optimization

Our four-layered model for demand highlights that most uncertainty in product sales comes from the high degree of uncertainty in campaign popularity (captured by a campaign parameter) and product popularity (captured by two product parameters, one for interest and another for purchase). This is problematic for the retailer because operational decisions take place at the SKU level: indeed it reserves inventory and sets prices individually for each item. Even worse, these decisions are taken before any of these uncertainties are cleared. One of the consequences is that sell-out rates (actual sales divided by reserved inventory) are highly concentrated around zero or 100%, as also noted by Ferreira et al. (2015).

What can the retailer do to improve its decisions? In our case, campaign "assortment" is fixed (this is the outcome of a lengthy negotiation with the brands) and product inventory is an input (this is left-over inventory from the brands). Thus the only relevant decision that the retailer can change is price. We describe in this section how a flash sales retailer can use price optimization to maximize its profits. We view this exercise as a proof-of-concept to extract value of the better forecast accuracy associated with the AEIO model. In particular, we ignore potential strategic customer behavior derived from price optimization, and hence a full validation would require a field experiment.

It is worth noting that Ferreira et al. (2015) take a similar approach. There, demand modelling does not contain individual product features but instead uses regression trees based on product characteristics; this does not capture the large amount of uncertainty at the product level. This means that our saturated models are able to provide more accurate predictions at the product level, as shown in §5.4. For this purpose, we consider the effect of price variation on the demand. Although the retailer does not currently change prices during the campaign, we can estimate the potential improvement of doing so. It turns out that, since the end of our conversations with the retailer, the firm started modifying prices during the campaign, at most once, at the start of the regular sales period, i.e., at the pre-campaign. The reason for this is that customer perception is minimally affected yet profit improvement remains significant: only a small fraction of customers has accessed the campaign and thus received actual information about products or placed orders; in case a customer has placed an order, our model would suggest price to go up, so it would not generally trigger a complaint.

Since arrivals to the website and visits to campaigns occur without knowledge of the price, we assume that A_t and E_{ct} are independent of price. In contrast, I_{jt} and O_{jt} can be influenced by price, which we denote as p_{jt} for product j at time t. From models (2) - (4), we know that Q_{it} follows a binomial distribution with A_t tries and probability of success $\beta_{jt} = e_{c(j)t} i_{jt} o_{jt}$, which is a function of p_{jt}. Since we focus on a concrete product j, we drop the sub-index j in what follows.

We assume that the retailer posts a starting price $p_t = p^{pre}$ for $t \in [t_\star, t_0)$. This initial price is not subject to optimization, and, consistently with the platform's actual practices, follows a general discount negotiated with the brand, uniformly applied to all products. After observing the demand during the pre-campaign, as described in §5.3, the retailer is then able to change the price to $p_t = p$ for $t \in [t_0, t_\dagger]$. Following the structure of our model, we have that total (potential) demand for $t \in [t_0, t_\dagger]$ can be written as $\sum_{t=t_0}^{t_\dagger} \text{Binom}(\beta_t, A_t)$. This is a function of p to the extent that it influences β_t. Letting K be the total available inventory at t_0, expected sales are $\mathbb{E} \min \Big(K, \sum_{t=t_0}^{t_\dagger} \text{Binom}\big(\beta_t(p), A_t\big)\Big)$. Assuming without loss of generality that the cost of the product is zero (otherwise p can denote the product's unit revenue minus unit cost), the retailer's optimization problem can thus be written as $\max_p \pi(p)$ where

$$\pi(p) := p \cdot \mathbb{E}\left[\min\left(K, \sum_{t=t_0}^{t_\dagger} \text{Binom}\Big(\beta_t(p), A_t\Big)\right)\right]. \tag{5}$$

Note that maximizing (5) is a pricing problem for the newsvendor and is well understood, see e.g., Petruzzi and Dada (1999). It turns out that when β_t is stationary and well-behaved, profit is unimodal and the optimal price can be determined by local search.

Proposition 1. *When $\beta_t(p)$ is independent of t, let $p(\beta)$ be the unique price such that $\beta_t(p) = p$ and $q(\beta) = \mathbb{E} \min \Big(K, \sum_{t=t_0}^{t_\dagger} Binom\big(\beta_t(p), A_t\big)\Big)$.*

(a) When

$$\frac{p''}{p'} - 2\frac{p'}{p} \geq 0 \text{ at all points where } -\frac{p'}{p} \leq \frac{1}{\beta}, \tag{6}$$

there exists a unique β^* that maximizes the retailer's profit, for any level of inventory and arrival distribution. This maximum is uniquely characterized by

$$-\frac{p'}{p} = \frac{q'}{q}. \tag{7}$$

(b) Condition (6) *is satisfied when conversion* β *is a logistic function of* p*, i.e.,* $\beta(p) = \frac{e^{v-\gamma p}}{1+e^{v-\gamma p}}$.

The proposition thus establishes regularity conditions on the price-conversion relationship to guarantee a single local optimum of the profit function. Note that these conditions are independent of the level of inventory and the arrival stream distribution A_t. They require that p is not too convex in the range where its relative decrease $-\beta p'/p$ is small, which typically coincides with β being large enough. This condition is satisfied for the functional forms used in our empirical study. Note that these results become difficult to extend when $\beta_t(p)$ varies over time (which is the case in the data), but fortunately the profit function seems to be quasi-concave in our proposed model, as tested in our numerical experiments.

Furthermore, the pricing optimality condition (7) relates price sensitivity $-p'/p \geq 0$ with the relative increase in sales q'/q obtained by a marginal increase in conversion. Interestingly, inventory K only appears in the term q'/q: as a result, when q'/q increases with K, which was verified numerically in the specifications we used, the optimal conversion increases and therefore the optimal price decreases with K. For example, when $\beta(p) = \frac{ve^{-\gamma p}}{1+ve^{-\gamma p}}$ and v is small enough, $p^* \approx 1/\gamma$ and expected revenue is equal to ve^{-1}/γ. In contrast, when v is high, inventory limitations drive the optimal price up so that items are still sold with high probability but also at higher unit margins. Hence, obtaining better demand forecast through early clicks will allow the retailer to increase prices for items that are expected to sell out, and to reduce them for the rest.

After establishing analytically the pricing optimality condition, we can reoptimize numerically prices using our model. In that case $\beta_t = e_t \frac{e^{v_i-\gamma_i p}}{1+e^{v_i-\gamma_i p}} \frac{e^{v_o-\gamma_o p}}{1+e^{v_o-\gamma_o p}}$. As discussed in the introduction, because there is no price variation within a campaign at the moment (and hence in our data), we cannot estimate price sensitivities γ_i and γ_o in models I3 and O3 respectively. Indeed lack of price variation in an item makes price sensitivity unidentifiable when product dummies are used. To be able to estimate it, one should introduce some random price variations (as opposed to endogenous price variation as in Fisher et al. 2017) and introduce price as a covariate in our model 3 to obtain product type or campaign price sensitivities. However, this is possible in models I2 and O2: we thus use price sensitivity $\gamma_i = -0.08/LISTP - 0.08/AVGP$ where $LISTP$ is the list price of product j and $AVGP$ is the average actual price of campaign $c(j)$ (negative because higher prices trigger higher requests for information, see Table 3); and $\gamma_o = 1.27/LISTP + 0.22/AVGP$ (see Table 4).

To illustrate this procedure, we select 10 random campaigns run after October 1, 2015, all of them offering a pre-opening. We replicated sampling several times and obtained similar results. These campaigns include 4,512 SKUs. We estimated models E3, I3 and O3 during the pre-opening to generate campaign and product dummies. Using price sensitivity γ_i and γ_o described above, we proceed to optimize price individually for each campaign, taking into account actual arrivals A_t and the actual inventory level available for the SKU. In order to avoid unrealistic price changes, we limit the possible price variation to -/+ 100% of the current price. This implies that even in the highest price increase the product will still be heavily discounted in comparison with the list price.

Table 8 shows the results of price optimization. We separate items by the ratio of forecast demand with price p^{pre} to available inventory, in six bins for predicted sell-out. We then report for each bin the actual sell-out and stock-out rates, and the average suggested variation in price and revenue.

Predicted sell-out	<10%	10-30%	30-50%	50-70%	70-90%	>90%	All
Number of SKUs	963	1,145	633	405	377	989	4,512
% of SKUs	21.3%	25.4%	14.0%	9.0%	8.4%	21.9%	100.0%
Average actual sell-out	5.2%	13.0%	19.5%	27.8%	36.6%	56.3%	25.0%
Actual stock-out rate	0.0%	0.8%	2.8%	8.6%	18.0%	43.7%	12.5%
Average suggested price variation	58.4%	61.5%	62.4%	65.0%	71.7%	93.8%	69.2%
Average revenue improvement	12.3%	13.4%	14.3%	16.1%	25.0%	72.8%	27.5%

Table 8: Segmentation of the 4,512 SKUs by predicted sell-out rate.

From the third and fourth rows of the table, we see that our approach tends to slightly overestimate sell-out (because for instance predictions of $> 90\%$ result in a value of 56%), especially for high predictions. On the other hand, it provides an excellent classification of SKUs: the items for which we predict lower sell-out rates indeed sell on average less and rarely stock-out; in contrast, those with higher sell-out rates sell strongly, e.g., when predicted sell-out is higher than 90%, then actual sell-out rate is highest at 56% and there is a stock-out event 44% of the time. This again means that early clicks are quite informative to separate the items that will have higher demand.

Furthermore, our optimization procedure, which is applied for each item independently, suggests that prices should be generally increased. More importantly, items with lower predicted sell-out should increase their price at the expense of a reduction in sell-out rate, thereby improving profit by $12-16\%$ despite suggested price increases of $58-65\%$. In contrast, items with higher predicted sell-out should increase price more (+94% on average, which suggests that for a large majority the limit of +100% price variation becomes active). Because these items were inventory-constrained, the decrease in demand does not reduce expected sales much, leading to a predicted revenue improvement of 73%. This implies that revenue improvement is generally quite significant, especially for items likely to sell-out, for which predicted revenue improvement is highest. This also means that, in case impact on consumers has to be minimized, it is advisable for the retailer to focus on the items with high predicted sell-out (only 22% of the SKUs), for which a steep price increase should be applied and which results in the strongest profit improvement. For instance, in Table 8, if all SKUs have their prices modified as suggested, total revenue would increase by 27%, but when the intervention only applies to the SKUs with predicted sell-out above 90%, total revenue would increase by 18% already.

In summary, this exercise illustrates the value of our model: early clicks provide a more precise forecast at the SKU level, and specifically allow us to separate items that are underdemanded from those that will sell out completely. As a result, after a few hours of operation, our model can guide price reoptimization to increase profits. Our simulation suggests that such price updates can generate significant improvements in profits by better matching supply and demand.

7 Conclusion

In this paper, we have developed a demand model for flash sales, which is a growing business model for online retailers where products are sold for a short time window at heavy discounts. The model incorporates four different layers that capture the interactions of visitors with campaigns first, and products second, from acquisition of information to purchase. The first layer captures the number of visitors arriving to the online platform and the three others represent the conversion from one stage to the next, conditionally on previous layers. Our framework allows to separate effects of time variations (seasonality and life-cycle patterns) from item characteristics (campaigns and products). Despite its conceptual complexity, the model can be easily estimated using standard likelihood maximization.

We illustrate the use of our model with an application to a large data set of hourly observations over 3 years, with millions of clicks and products. We find strong evidence of life-cycle variations, with an intense browsing behavior at the beginning of a campaign but a stronger purchase behavior at the end of it. We also see high degree of heterogeneity across campaigns and products, which implies that models with individual fixed effects tend to perform much better from a goodness-of-fit perspective. In particular, we develop a method to learn about new products performance by observing a few hours of activity. These "early clicks" can then be used to improve demand forecasts. This improvement can be particularly valuable to the retailer because it can adjust prices to better match supply and demand. A simulation with several campaigns chosen randomly shows that, according to our model, such price reoptimization can deliver significant increases in revenues. The increase in revenue is most significant for items that are predicted to stock out with high likelihood.

Our hierarchical perspective can thus help retail operations by providing insight about customer behavior inside the shopping funnel, by improving forecast accuracy and by allowing the retailer to learn about demand faster than direct demand models that do not have access to updated information obtained from higher-level layers.

Our study reveals further opportunities for improving the flash sales business model. Our detailed data set allows to empirically uncover rich patterns on search and purchase behavior. Instead of focusing on product performance as done here, it is possible to examine individual responses to several aspects of online operations, such as product display and visibility, inventory availability and fulfillment lead-times. These responses can be mapped into many dimensions too, such as customer visit frequency and depth (number of clicks), purchase amounts or returns. In addition to this empirical research agenda, another direction of future research is to develop analytical models that consider customer browsing and purchasing, in the line of Wang and Sahin (2016) or Bernstein et al. (2018).

Acknowledgements

We are grateful to the feedback a Senior Editor and two reviewers, which helped us improve the paper.

References

Agnew, H. 2017. Jacques-Antoine Granjon has grand designs for Vente-Privee. *Financial Times* June 6:online.

Anderson, S. P., A. De Palma, and J.-F. Thisse. 1992. *Discrete choice theory of product differentiation*. MIT Press.

Aouad, A., V. F. Farias, and R. Levi. 2015. Assortment optimization under consider-then-rank choice models. Working paper, MIT.

Arora, N., G. M. Allenby, and J. L. Ginter. 1998. A hierarchical Bayes model of primary and secondary demand. *Marketing Science* 17 (1): 29–44.

Bakshi, N., and H. Peura. 2016. Reporting (or Not) of Precursors to Disasters. Working paper, SSRN.

Ben-Shabat, H., P. Nilforoushan, C. Yuen, and M. Moriarty. 2015, April. AT Kearney - The 2015 Global Retail E-Commerce Index - Global Retail E-Commerce Keeps On Clicking. https://www.atkearney.com/documents/10192/5691153/Global+Retail+E-Commerce+Keeps+On+Clicking.pdf.

Bernstein, F., S. Modaresi, and D. Sauré. 2018. A dynamic clustering approach to data-driven assortment personalization. *Management Science* Forthcoming:NA.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*:841–890.

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

Bitran, G., R. Caldentey, and S. Mondschein. 1998. Coordinating clearance markdown sales of seasonal products in retail chains. *Operations research* 46 (5): 609–624.

Block, H. D., J. Marschak et al. 1960. Random orderings and stochastic theories of responses. *Contributions to probability and statistics* 2:97–132.

Bresnahan, T. F. 1987. Competition and collusion in the American automobile industry: The 1955 price war. *The Journal of Industrial Economics*:457–482.

Bucklin, R. E., and C. Sismeiro. 2009. Click here for Internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing* 23 (1): 35–48.

Caro, F., and J. Gallien. 2010. Inventory Management of a Fast-Fashion Retail Network. *Operations Research* 58 (2): 257–273.

Caro, F., and V. Martínez-de Albéniz. 2015. Fast fashion: Business model overview and research opportunities. In *Retail Supply Chain Management: Quantitative Models and Empirical Studies, 2nd Edition*, ed. N. Agrawal and S. A. Smith, 237–264. Springer, New York.

Caro, F., V. Martínez-de Albéniz, and P. Rusmevichientong. 2014. The Assortment Packing Problem: Multiperiod Assortment Planning for Short-Lived Products. *Management Science* 60 (11): 2701–2721.

Chan, L. M., Z. M. Shen, D. Simchi-Levi, and J. L. Swann. 2004. Coordination of pricing and inventory decisions: A survey and classification. In *Handbook of quantitative supply chain analysis*, 335–392. Springer.

Cheung, W. C., D. Simchi-Levi, and H. Wang. 2017. Dynamic pricing and demand learning with limited price experimentation. *Operations Research* 65 (6): 1722–1731.

Christopher, M., R. Lowson, and H. Peck. 2004. Creating Agile Supply Chains in the Fashion Industry. *International Journal of Retail and Distribution Management* 32 (8): 367–376.

Cui, R., S. Gallino, A. Moreno, and D. J. Zhang. 2018. The operational value of social media information. *Production and Operations Management* 27 (10): 1749–1769.

Dunn, D., W. Williams, and T. DeChaine. 1976. Aggregate versus subaggregate models in local area forecasting. *Journal of the American Statistical Association* 71 (353): 68–71.

Ehrenberg, A. S. C. 1972. *Repeat-buying; theory and applications.* North-Holland Pub. Co.

Eliashberg, J., Q. Hegie, J. Ho, D. Huisman, S. J. Miller, S. Swami, C. B. Weinberg, and B. Wierenga. 2009. Demand-driven scheduling of movies in a multiplex. *International Journal of Research in Marketing* 26 (2): 75–88.

Federgruen, A., and A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Operations Research* 47 (3): 454–475.

Feenstra, R. C., and J. A. Levinsohn. 1995. Estimating markups and market conduct with multidimensional product attributes. *The review of economic studies* 62 (1): 19–52.

Feldman, J., and H. Topaloglu. 2015. Assortment optimization under the multinomial logit model with nested consideration sets. Working paper, Cornell University.

Ferreira, K. J., B. H. A. Lee, and D. Simchi-Levi. 2015. Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 18 (1): 69–88.

Ferreira, K. J., D. Simchi-Levi, and H. Wang. 2018. Online network revenue management using thompson sampling. *Operations Research* 66 (6): 1586–1602.

Fisher, M. 2009. OR FORUM–Rocket Science Retailing: The 2006 Philip McCord Morse Lecture. *Operations Research* 57 (3): 527–540.

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

Fisher, M. L. 1997. What is the right supply chain for your product? *Harvard Business Review* 75:105–117.

Fisher, M. L., and A. Raman. 1996. Reducing the Cost of Demand Uncertainty Through Accurate Response to Early Sales. *Operations Research* 44 (1): 87–99.

Frisbie, G. A. 1980. Ehrenberg's Negative Binomial Model Applied to Grocery Store Trips. *Journal of Marketing Research* 17 (3): 385–390.

Gallien, J., A. J. Mersereau, A. Garro, A. D. Mora, and M. N. Vidal. 2015. Initial shipment decisions for new products at Zara. *Operations Research* 63 (2): 269–286.

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

Gallino, S., and A. Moreno. 2015. The Value of Fit Information in Online Retail: Evidence from a Randomized Field Experiment. *Manufacturing & Service Operations Management* Forthcoming:NA.

Greenwood, M., and G. U. Yule. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *Journal of the Royal statistical society* 83 (2): 255–279.

Huang, T., and J. A. Van Mieghem. 2013. The promise of strategic customer behavior: On the value of click tracking. *Production and Operations Management* 22 (3): 489–502.

Huang, T., and J. A. Van Mieghem. 2014. Clickstream data and inventory management: Model and empirical analysis. *Production and Operations Management* 23 (3): 333–347.

Iyer, A. V., and M. E. Bergen. 1997. Quick Response in Manufacturer-Retailer Channels. *Management Science* 43 (4): 559–570.

Liaukonyte, J., T. Teixeira, and K. C. Wilbur. 2015. Television advertising and online shopping. *Marketing Science* 34 (3): 311–330.

Martínez-de-Albéniz, V., and A. Valdivia. 2016. The Impact of Exhibition Scheduling on Museum Attendance. *Manufacturing & Service Operations Management* Forthcoming:NA.

McFadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, ed. P. Zarembka. Academic Press, New York.

Moe, W. W., and P. S. Fader. 2002. Using advance purchase orders to forecast new product sales. *Marketing Science* 21 (3): 347–364.

Moe, W. W., and P. S. Fader. 2004a. Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing* 18 (1): 5–19.

Moe, W. W., and P. S. Fader. 2004b. Dynamic conversion behavior at e-commerce sites. *Management Science* 50 (3): 326–335.

Petruzzi, N. C., and M. Dada. 1999. Pricing and the newsvendor problem: A review with extensions. *Operations Research* 47 (2): 183–194.

Raudenbush, S. W., and A. S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*, Volume 1. Sage.

Shao, H. 2013. Meet Vipshop, The Highest Valued Chinese E-Commerce Stock. *Forbes Asia* November 26:online.

Smith, S., and D. Achabal. 1998. Clearance Pricing and Inventory Policies for Retail Chains. *Management Science* 44 (3): 285–300.

Train, K. 2009. *Discrete choice methods with simulation*. Cambridge university press.

Tversky, A. 1972. Choice by elimination. *Journal of mathematical psychology* 9 (4): 341–367.

Van Mieghem, J. A., and M. Dada. 1999. Price versus production postponement: Capacity and competition. *Management Science* 45 (12): 1639–1649.

Vipshop 2018, February. Vipshop Reports Unaudited Fourth Quarter and Full Year 2017 Financial Results. http://ir.vip.com/phoenix.zhtml?c=250900&p=irol-newsArticle&ID=2331936.

Wang, R., and O. Sahin. 2016. The Impact of Consumer Search Cost on Assortment Planning and Pricing. *Management Science* Forthcoming:NA.

Wiesel, T., K. Pauwels, and J. Arts. 2011. Practice Prize Paper-Marketing's Profit Impact: Quantifying Online and Off-line Funnel Progression. *Marketing Science* 30 (4): 604–611.

Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Appendix: Proofs and Additional Information

Proposition 1

Proof. For every realization of uncertainty ε (which includes A and any other model-related uncertainty such as product attractiveness) and every price p, we define $\beta(p,\varepsilon)$ as the conversion. Letting $\phi(\beta) = \mathbb{E}\min\{K, \sum_{i=1}^{A} X_i\}$ where X_i is a Bernoulli variable with probability β, we have

$$q = \mathbb{E}_\varepsilon \phi\Big(\beta(p,\varepsilon)\Big) = \mathbb{E}_\varepsilon\left[\sum_{i=1}^{\infty} \frac{A!}{i!(A-i)!}\min\{K,i\}\beta^i(1-\beta)^{A-i}\right].$$

As a result,

$$\begin{aligned}
\frac{dq}{dp} &= \mathbb{E}_\varepsilon \frac{\partial\beta}{\partial p}\left[\sum_{i=1}^{\infty}\frac{A!}{i!(A-i)!}\min\{K,i\}\left(i\beta^{i-1}(1-\beta)^{A-i} - (A-i)\beta^i(1-\beta)^{A-1-i}\right)\right] \\
&= \mathbb{E}_\varepsilon \frac{\partial\beta}{\partial p}\left[A\sum_{i=1}^{\infty}\frac{(A-1)!}{(i-1)!(A-i)!}\left(\min\{K,i\}-\min\{K,i-1\}\right)\beta^{i-1}(1-\beta)^{A-i}\right] \\
&= \mathbb{E}_\varepsilon \frac{\partial\beta}{\partial p}\left[A\sum_{i=0}^{\infty}\frac{(A-1)!}{i!(A-1-i)!}1_{i\le K-1}\beta^i(1-\beta)^{A-1-i}\right] \\
&= \mathbb{E}_\varepsilon \frac{\partial\beta}{\partial p}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right] \\
&\le 0,
\end{aligned}$$

because $\partial\beta/\partial p \le 0$. Also,

$$\begin{aligned}
\frac{d^2q}{dp^2} &= \mathbb{E}_\varepsilon\left(\frac{\partial\beta}{\partial p}\right)^2\left[A\sum_{i=0}^{\infty}\frac{(A-1)!}{i!(A-1-i)!}1_{i\le K-1}\left(i\beta^{i-1}(1-\beta)^{A-1-i}-(A-1-i)\beta^i(1-\beta)^{A-2-i}\right)\right] \\
&\quad +\mathbb{E}_\varepsilon\frac{\partial^2\beta}{\partial p^2}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right] \\
&= \mathbb{E}_\varepsilon\left(\frac{\partial\beta}{\partial p}\right)^2\left[A(A-1)\sum_{i=1}^{\infty}\frac{(A-2)!}{(i-1)!(A-1-i)!}\left(1_{i\le K-1}-1_{i-1\le K-1}\right)\beta^{i-1}(1-\beta)^{A-1-i}\right] \\
&\quad +\mathbb{E}_\varepsilon\frac{\partial^2\beta}{\partial p^2}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right] \\
&= -\mathbb{E}_\varepsilon\left(\frac{\partial\beta}{\partial p}\right)^2\left[A(A-1)\sum_{i=0}^{\infty}\frac{(A-2)!}{i!(A-2-i)!}1_{i=K-1}\beta^i(1-\beta)^{A-2-i}\right] \\
&\quad +\mathbb{E}_\varepsilon\frac{\partial^2\beta}{\partial p^2}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right] \\
&= -\mathbb{E}_\varepsilon\left(\frac{\partial\beta}{\partial p}\right)^2\left[A(A-1)1_{\sum_{i=1}^{A-2}X_i=K-1}\right] \\
&\quad +\mathbb{E}_\varepsilon\frac{\partial^2\beta}{\partial p^2}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right] \\
&\le \mathbb{E}_\varepsilon\frac{\partial^2\beta}{\partial p^2}\left[A1_{\sum_{i=1}^{A-1}X_i\le K-1}\right].
\end{aligned}$$

Note that $q \ge \mathbb{E}_\varepsilon[\beta A1_{\sum_{i=1}^{A}X_i\le K}] \ge \mathbb{E}_\varepsilon[\beta A1_{\sum_{i=1}^{A-1}X_i\le K-1}]$, regardless of the functional form of A and the value of K. At a price that maximizes $\pi = pq$, $\pi' = 0 = q + p\cdot dq/dp$, which implies

$$\mathbb{E}_\varepsilon\left(\beta + p\frac{\partial\beta}{\partial p}\right)[A1_{\sum_{i=1}^{A-1}X_i\le K-1}] \ge 0$$

π is quasi-concave when, in all such p, $\pi'' = 2dq/dp + p \cdot d^2q/dp^2$. Noting that

$$\pi'' \leq \mathbb{E}_\varepsilon \left(2\frac{\partial \beta}{\partial p} + p\frac{\partial^2 \beta}{\partial p^2} \right) [A1_{\sum_{i=1}^{A-1} X_i \leq K-1}]$$

it is sufficient that for all weights $w_\varepsilon \geq 0$ such that $\sum_\varepsilon \left(\beta + p\frac{\partial \beta}{\partial p}\right) w_\varepsilon \geq 0$, $\left(2\frac{\partial \beta}{\partial p} + p\frac{\partial^2 \beta}{\partial p^2}\right) w_\varepsilon \leq 0$, i.e., part (a).

For part (b), consider a logistic conversion where $\beta(p,\varepsilon) = \frac{e^{\varepsilon-\gamma p}}{1+e^{\varepsilon-\gamma p}}$. Without loss of generality, we can prove the property with the case $\gamma = 1$. In this case,

$$\frac{\partial \beta}{\partial p} = -\frac{e^{\varepsilon-p}}{(1+e^{\varepsilon-p})^2} \text{ and } \frac{\partial^2 \beta}{\partial p^2} = \frac{e^{\varepsilon-p}}{(1+e^{\varepsilon-p})^2} - \frac{2(e^{\varepsilon-p})^2}{(1+e^{\varepsilon-p})^3}.$$

Since condition (6) involves ratios of the derivatives of p, we can without loss of generality consider only β such that $p(\beta) > 0$. The points for which $-p'/p \leq 1/\beta$ satisfy $1/(1-\beta) \leq -\ln(\beta) + \ln(1-\beta) + \ln(v)$, i.e., $\beta \geq \beta_0$ where β_0 is the unique solution to $1/(1-\beta) = -\ln(\beta) + \ln(1-\beta) + \ln(v)$. The condition $\frac{p''}{p'} - 2\frac{p'}{p} \geq 0$ becomes

$$\frac{1}{\beta} - \frac{1}{1-\beta} \leq \frac{2(\frac{1}{\beta} + \frac{1}{1-\beta})}{-\ln(\beta) + \ln(1-\beta) + \ln(v)}.$$

This is satisfied in two cases: when $\beta \geq \beta_0 \geq 1/2$; or when $\beta_0 \leq \beta < 1/2$ and

$$0 \geq -\ln(\beta) + \ln(1-\beta) + \ln(v) - \frac{2(\frac{1}{\beta} + \frac{1}{1-\beta})}{\frac{1}{\beta} - \frac{1}{1-\beta}} = -\ln(\beta) + \ln(1-\beta) + \ln(v) - \frac{2}{1-2\beta},$$

which is true at $\beta = \beta_0$ because $\frac{1}{1-\beta_0} - \frac{2}{1-2\beta_0} \leq 0$ when $\beta_0 < 1/2$; and as β increases the variation of the right-hand side is $-\frac{1}{\beta} - \frac{1}{1-\beta} - \frac{4}{(1-2\beta)^2} \leq 0$. This completes the proof of part (b). □

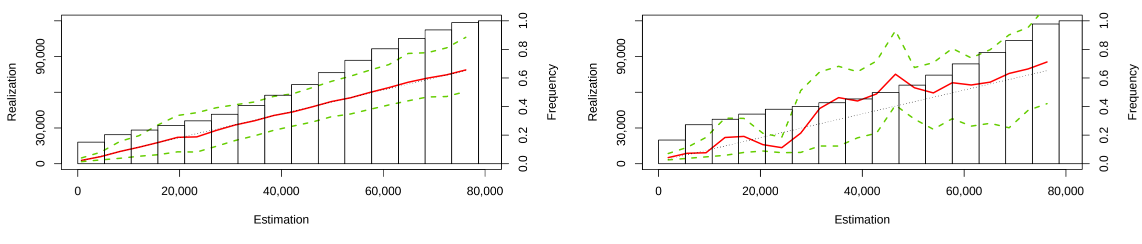
AEIO Model Covariates and Validation

Covariate	Description
$DAYHOUR_t^{dh}$	Dummy for day-hour $d = Monday, \ldots, Sunday$, $h = 0, \ldots, 23$: captures during-week seasonality
$WEEK_t^w$	Dummy for week $w = 1, \ldots, 52$: captures seasonality over the year
$TIME_t$	Number of days since the beginning of the data: captures growth of total traffic during the time

Table 9: Detail of covariates $\boldsymbol{X}_t^A$ corresponding to layer A, characterizing the arrival distribution (1).

Covariate	Description
$AGE_{ct} = t - t_{0c}$	Hours since campaign started: captures decay of interest (Caro et al. 2014)
$DAYS^d_{ct}$	Dummy for campaign being in days $d = 1, 2, 3$ since start, in its last day $d = last$ or during pre-campaign $d = pre$: captures opening and ending effects (Eliashberg et al. 2009, Martínez-de-Albéniz and Valdivia 2016)
$WITHPREOP_c$	Dummy for whether campaign c had a pre-opening: captures heterogeneity in campaign type
$LOGAVGLISTP_c$	Log of average of list price of all products in campaign c: captures brand strength
$SINGLEBRAND_c$	Binary variable of whether campaign c only carries one brand: captures campaign focus
$TIME_c$	Day number since the beginning of the data of the opening date of campaign c: captures increase in campaign attractiveness over time

Table 10: Detail of covariates $\boldsymbol{X}^E_{ct}$ corresponding to layer E, characterizing the conversion probability (2). Note that only active campaigns are included at time t.



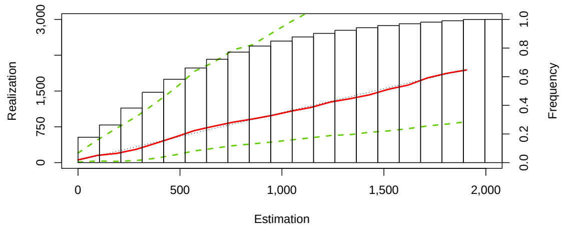
(a) In sample

(b) Out of sample

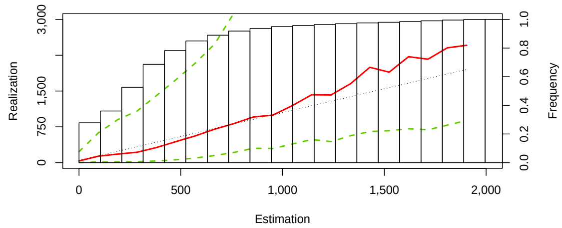
Figure 8: Comparison between predicted arrivals $\hat{A}_t$ from model (1) and actual arrivals A_t, using the negative binomial specification. In each panel, the sample mean (red solid line) is compared against a 100% accuracy (black dotted line). The dashed lines represent the mean plus and minus the standard deviation.

Covariate	Description
$AGE_{c(j)t} = t - t_{0c(j)}$	Hours since campaign $c(j)$ to which product j belongs started: captures decay of interest (Caro et al. 2014)
$DAYS^{d}_{c(j)t}$	Dummy for campaign being in days $d = 1, 2, 3$ since start, in its last day $d = last$ or during pre-campaign $d = pre$: captures opening and ending effects (Eliashberg et al. 2009, Martínez-de-Albéniz and Valdivia 2016)
$WITHPREOP_{c(j)}$	Dummy for whether campaign $c(j)$ had a pre-opening: captures heterogeneity in campaign type
$LOGAVGLISTP_{c(j)}$	Log of average of list price of all products in campaign c: captures brand strength
$LOGPRODUCTS_{c(j)}$	Log of the number of products in campaign $c(j)$: captures substitution options to product j
$PRODTYPE_j$,	Categorical variable describing the category of product j: captures variations across categories
$LOGLISTP_j$	Log of list price of product j: captures quality perception
$LOGSIZES_j$	Log of the number of sizes within product j: captures potential difficulties on finding the right size (Gallino and Moreno 2015)
$RATIOPtoLISTP_j$	Ratio of actual price of product j to list price of product j: captures sensitivity to discount
$RATIOPtoAVGP_j$	Ratio of actual price of product j to average actual price among all products of campaign $c(j)$: captures reference effects from campaign pricing (Ferreira et al. 2015)

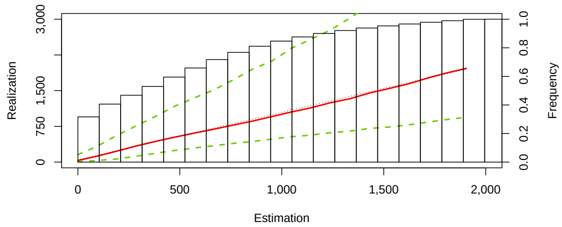
Table 11: Detail of covariates $\boldsymbol{X}^I_{jt} = \boldsymbol{X}^O_{jt}$ corresponding to layers I and O, characterizing the conversion probabilities (3)-(4). Note that only active campaigns are included at time t. For layer O, we note that observations where inventory is stocked-out are removed (they cannot be clicked on the website), so there is no need to control for stock availability.



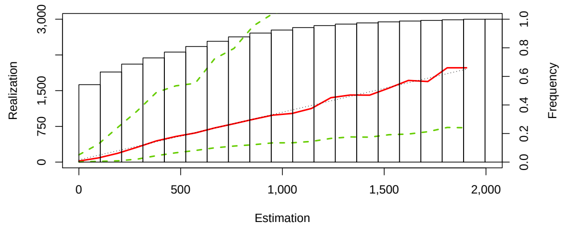
(a) Specification E1, in sample



(b) Specification E1, out of sample

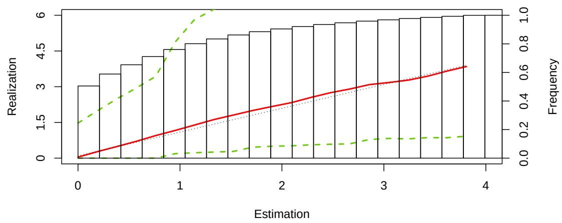


(c) Specification E2, in sample

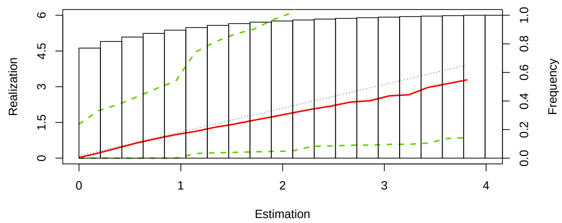


(d) Specification E2, out of sample

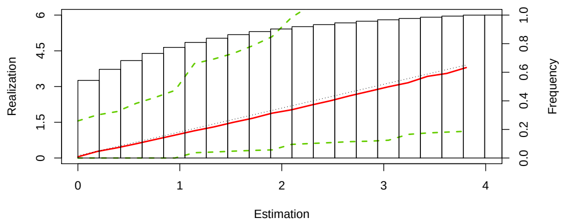
Figure 9: Comparison between campaign clicks $\hat{E}_{ct}$ from model (2) and actual E_{ct}, conditional on knowledge of A_t.



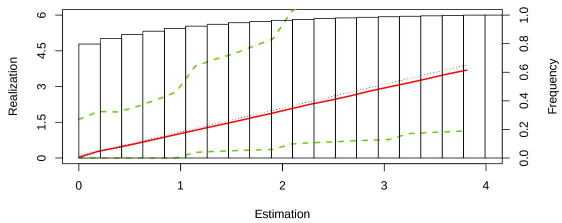
(a) Specification I1, in sample



(b) Specification I1, out of sample

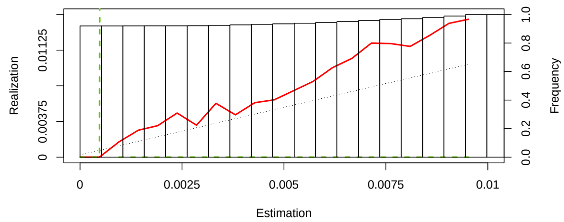


(c) Specification I2, in sample

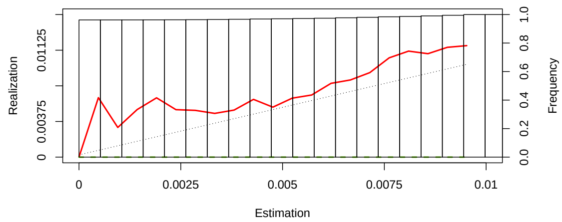


(d) Specification I2, out of sample

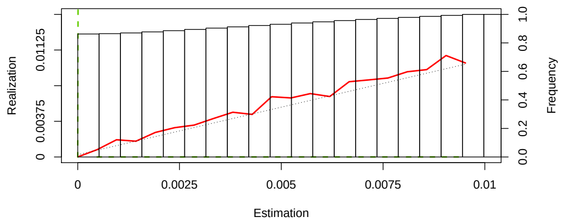
Figure 10: Comparison between product visits or add-to-cart $\hat{I}_{jt}$ from model (3) and actual I_{jt}, conditional on knowledge of $E_{c(j)t}$.



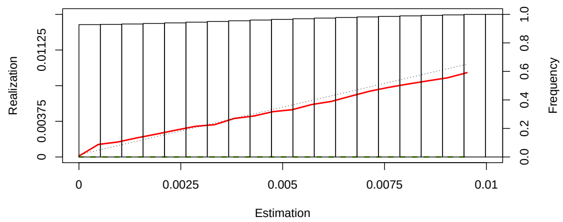
(a) Specification O1, in sample



(b) Specification O1, out of sample



(c) Specification O2, in sample



(d) Specification O2, out of sample

Figure 11: Comparison between product purchases $\hat{O}_{jt}$ from model (4) and actual O_{jt}, conditional on knowledge of I_{jt}. Note that dashed lines do not appear here because actual purchases take values of 0, 1 or higher integers, thus not shown on the graph.