# The Value of Online Interactions for Store Execution

Felipe Caro[1] • Victor Martínez-de-Albéniz[2] • Borja Apaolaza[3]

**Abstract**

***Problem definition:*** Omnichannel retailers interact with customers both online and offline. So far, they have used this richer information to optimize the sales process by designing the right channel and supply chain structures, and by personalizing offer, pricing and promotions. We advance an additional dimension of omnichannel value: retailers can use online clickstreams to better understand customer needs, and optimize store layouts to maximize conversion. ***Methodology/results:*** We develop a model where in-store purchases depend on the customer's product shopping list, and the effort required to locate and reach the products within the store. Product location in the store thus drives conversion. We then apply our model to a large home improvement retailer and find that preferences of store visitors are revealed by nearby online traffic to product pages, and hard-to-reach locations lead to lower conversion. We also do not find evidence of cross-selling. Finally, we optimize product-location assignments using our demand model and find that putting higher-interest and higher-price items in the most effective locations can increase revenues by about 10%. ***Managerial implications:*** Our results show how using online clickstream information for optimizing offline operations can create significant value. We also show that webrooming can be beneficial even in the absence of cross-selling.

Keywords: Shopping lists, shopping baskets, clickstream, webrooming, demand estimation, layout

Submitted: November 18, 2021.

## 1 Introduction

In the last decade, omnichannel has become a dominating retail strategy in which retailers do not see online and offline as independent channels, but manage them jointly (Gallino and Moreno 2019, Caro et al. 2020). Omnichannel delivers value on multiple dimensions, because it allows customers to learn about the product in one channel, and fulfill the demand in another (Bell et al. 2014). This flexibility implies that retailers are no longer constrained to run a single-channel sales process, and have more freedom to optimize the funnel from need to purchase (Wiesel et al. 2011). The additional flexibility requires closer coordination of the operations and marketing functions (Bijmolt et al. 2021), but has the potential to groom more effective interactions with the customer, increasing their satisfaction and delivering higher profits to the retailer.

The literature has identified different ways to extract value from omnichannel. On the one hand, traditional marketing actions can be refined with more precise customer histories, such as targeted advertising or promotions (Goic and Olivares 2019). On the other hand, many decisions in the operational realm have been improved. The design of channels can be optimized by better understanding how offline and online affect each other (Bell et al. 2017, 2020, Kumar et al. 2019,

[1]UCLA Anderson School of Management, Los Angeles, CA 90095, felipe.caro@anderson.ucla.edu

[2]IESE Business School, University of Navarra, Av. Pearson 21, 08034 Barcelona, Spain, valbeniz@iese.edu.

[3]The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, apaolaza@upenn.edu.

Bar-Gill and Reichman 2020). Information provision can drive the channel choices of the consumers (Gallino and Moreno 2014). Information from online sources can also help predict demand better so inventory levels can be optimized (Huang and Van Mieghem 2014, Cui et al. 2018). Finally, fulfillment flexibility allows firms to better run their supply networks (Hübner et al. 2019).

Most of the strategies described above are effective for firms that have a large online sales channel, but it is less clear how offline-heavy retailers can take advantage of an omnichannel strategy. Moreover, despite the increase of online shopping, retailers keep investing in stores as these remain the primary channel to interact with consumers (Schaverien 2018, Dowsett 2019). According to Bell et al. (2014), the value can be obtained by providing better information online, in a webrooming model. In this paper, we intend to uncover one additional value creation strategy available to omnichannel retailers, which is especially important under webrooming: one can exploit online interactions to detect (true) customer needs at a granular level – product or category level– and study the determinants of store conversion, measured in the number of sales generated by these potential customers.

This conversion process is complex, as visitors' initial shopping intention must translate into exposure to the wanted products (and others), then into consideration sets, and finally, possibly after reflection, consultation with store staff, and try-on, into purchase. It involves time and effort from the consumers, so a well-engineered layout can help them access their desired products quicker, and they might end up buying with a higher probability (Underhill 2009). Indeed, convenience increases the chances that customers buy: more formally, time pressure and higher search costs decrease sales (Hui et al. 2009b, Brynjolfsson et al. 2011). This the reason why impulse items such as chocolates are often located near the check-out line, and Amazon has patented the One-Click button to reduce cart abandonment (Wagner and Jeitschko 2017). Unfortunately, the understanding of the relationship between layouts and conversion is limited. While richer displays are known to increase conversion (Boada-Collado and Martínez-de Albéniz 2020), there is a lack of empirical evidence linking product position in the store with sales. Causal evidence of this kind is hard to obtain, because retailers generally do not know the store visitor's shopping list, and hence they only observe sales performance of a particular store area but not how effective it was in capturing potential purchase intentions. As Goic and Olivares (2019) put it, "In contrast [to online channels], data regarding browsing behavior in retail stores have been, for the most part, nonexistent. Studies that seek to measure the effect of changes in the layout and display of a store have typically used aggregate store-level data to conduct causal analysis." In this paper, we provide one novel way to assess the effects of layout on conversion, which can be fed into a layout optimization model and generate improved store layouts.

For this purpose, we first build a theoretical model where conversion is affected by the physical effort invested by the visitor to locate products in her shopping list. We then work with a large home improvement retailer for which we observe, during 7 months, all offline and online activities.

For 16 stores, we observe full transaction records, i.e., composition of individual tickets, and product details and precise location within each store. For the online channel, we observe full clickstreams, i.e., all the clicks with timestamps by distinct geolocated origins of internet traffic. For each store and product, we are thus able to count how many different potential customers might be interested in the product. This is a proxy for the number of store visitors genuinely interested in purchasing the product, and we show that it is indeed a strong predictor of product sales. We are then in a position to study how conversion is moderated by product location in the store. We find that the distance from the store entrance is the most critical determinant of conversion and items easier to reach – closer to the store entrance – exhibit significantly higher conversion, after controlling for product fixed effects. In contrast, we find that spillovers from adjacent products are not significant (recall that these are home improvement products for which there is little impulse shopping), which suggests that using store visits to create cross-selling revenue may not always be possible or desirable, as also suggested by Gao and Su (2017a). Our empirical findings pave the way for optimizing store layout. We formulate this question as an assignment optimization problem, and show that revenues can be increased by 10.2% when online information is used to decide product locations. This involves a one-time layout change. The revenue lift marginally increases up to 10.5% if we further allow the retailer to change the layout weekly.

Our work thus contributes to the growing literature on retail analytics, specifically by showing that having access to product preference lists – available in online interactions – as opposed to simply shopping baskets – typical in store transaction records – is very valuable. Our approach is thus a simpler alternative to in-store customer tracking (Hui et al. 2009a), and more importantly gives access to information about which products attracted visitors to the store (see Chen et al. 1999 for a similar idea applied to advertising). Besides establishing the connection between store layout and sales, we provide an integrative perspective where customer behavior is combined with layout design decisions, which goes beyond minimization of average travel distance (De Koster et al. 2007) or consideration of product adjancencies (Ozgormus and Smith 2020).

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 formulates the model of a shopping visit and formulates the layout optimization problem. We estimate the impact of layout on sales in Section 4. Section 5 includes a counterfactual analysis of alternative store layouts. Section 6 concludes the paper.

## 2 Literature Review

Our work is mainly related to three streams of literature. First, we build on the operations and marketing literature that has studied the shopping funnel. Second, we are connected to the works about offline-online channel interactions. Third, we contribute to the literature on prescriptive models for retail execution.

## 2.1 Shopping funnel

The concept of funnel is a natural approach to study the effects of different marketing strategies on the customer. The funnel applies to both physical channels, where store visits are transformed into units sold, and online channels, where leads become visits which in turn generate orders. Wiesel et al. (2011) provides a framework to integrate both channels. In terms of modelling, hierarchical models are a convenient way to capture that only a fraction of those showing an initial interest in shopping end up making a purchase (Arora et al. 1998, Martínez-de Albéniz et al. 2020). Detailed decision processes have been developed, such as the use of consideration sets (Wang and Sahin 2018) or product evaluation heuristics (Aouad et al. 2021).

An interesting variation of the study of shopping funnels is possible when interactions between customer and firm occur multiple times, requiring us to consider the sequence in which they occur. Sequential decisions have been considered in the marketing and economics literature across visits (Chintagunta et al. 2012) or within a single visit (Hui et al. 2009a,b, Ruiz et al. 2020). Hui et al. (2009a) provides a review of marketing research that considers paths of consumers in store settings. Ruiz et al. (2020) include memory effects as well as one-step forward considerations. Closest to our context is Hui et al. (2009b), who model the path between supermarket categories, using a conditional model where transitions from category to category are driven by destination characteristics and path history, which they validate with path data in one store. Interestingly, store layout is included in the decision process through the distance between store regions. We use a similar approach, with some differences. First, we do not have individual path data, but have access to product-level aggregates. Second, we focus on conversion from needs to purchases, so our consideration of cross-product interactions is operationalized in conversion spill-overs, between needs for one product and sales for another, which is moderated by the location of products in the store. Third, we see variation of product locations across different stores, which allows us to control for product characteristics and separately measure the impact of product location.

Furthermore, the literature, except for the papers that employ path data, typically considers transaction data in the form of baskets, i.e., the set of products appearing in a given ticket. To the best of our knowledge, only one paper theorizes about the importance of order within the basket. Chen et al. (1999) highlights that some categories are more important than others as they are the reason behind a store visit. They develop the concept *marketing profits* to reflect that profit should be attributed to the category that brought the customer to the store. While our focus is very different, we are able to separate online activities into initial clicks, those that appeared in the first position of an online session, and those that were clicked afterwards. We find that first-time clicks are more important in predicting store sales, in comparison with later clicks.

## 2.2 Omnichannel

The omnichannel phenomenon has been extensively studied in the last decade. Brynjolfsson et al. (2013) provides an early discussion of the potential of omnichannel for retailers. Gao and Su (2017a,b) develop analytical models for channel choice under omnichannel capabilities. As conceptualized in Bell et al. (2014), the benefits of omnichannel come from better product information, and from better fulfillment possibilities. In other words, there advantages in showrooming, and in webrooming.

Showroom physical interactions in the store allows to engage with customers more effectively. Bell et al. (2015, 2017, 2020) show how the convenience and the store experience can help online pure players sell more. Kumar et al. (2019) identifies the possibility of making in-store returns as another driver of sales increases.

Webrooming can also be valuable. Gallino and Moreno (2014) study the effect of Buy Online Pickup in Store (BOPS) on online and offline sales, and find that store traffic increases due to better information about in-store product availability. Interestingly, the quality of the online experience has an impact on offline sales (Bar-Gill and Reichman 2020, Gallino et al. 2018).

Our study reveals a different value driver of omnichannel. It can be used to identify demand needs at the store level and hence study the impact of store layout on the conversion process.

## 2.3 Retail Execution

Our work is also connected to works that develop models to improve retail execution. Interventions have focused on different dimensions, which we briefly review. Perdikaki et al. (2012), and Mani et al. (2015) measure the impact of staffing levels on sales, which Chuang et al. (2016) use to develop a labor planning methodology. Caro and Gallien (2010) study inventory distribution across stores and combine demand forecasting and inventory allocation optimization to improve sales at Zara; Gallien et al. (2015) apply a similar approach to new product distribution. Inventory inaccuracy is another cause of suboptimal retail performance. DeHoratius et al. (2008) measure the extent of inaccuracies and DeHoratius and Raman (2008) use inventory replenishment and audits to mitigate their effects. Montoya and Gonzalez (2019) develop a hidden Markov chain model to predict phantom stock-outs based on sales time-series. The effect of store congestion has also been explored: Lu et al. (2013) measure how queues reduce sales conversion.

We discuss here an understudied aspect of retail execution. Indeed there is no existing work that studies the role of store layout in generating sales. In particular, Hui et al. (2009b) do not investigate store design because they only have data on one store and hence cannot disentangle the effect of product location from the product itself. In contrast, we study layout decisions. The design of a store layout resembles that of designing a warehouse. There exists a broad literature on warehouse layout optimization, see De Koster et al. (2007) for an excellent review. Usually, the design problem is formulated as a large integer program that is solved with heuristic techniques. The methods have also been applied to store layout design, e.g., Mowrey et al. (2018). In these models, customer behavior is integrated through simplified customer behavior assumptions such as considering penalties for categories that are not adjacent (Ozgormus and Smith 2020). In contrast, we use the moderating effect of location on conversion to propose improve layouts.

# 3 Model

## 3.1 The Shopping Process

In the same vein as the shopping funnel discussed in Section 2, we make the following assumptions for the shoppers in our model:

1. Store choice: consumers prefer buying at a store that is closest to where they live. Hence, each store has a "natural catchment area" that consists of all the households within a certain radius.

2. Shopping lists: a significant fraction of consumers starts their purchasing process with a prioritized list of items in mind that they would like to buy or are considering buying. Products that are more important to the consumer are held higher in the list.

3. Webrooming: a significant fraction of consumers does product research online on the retailer's website, and then follow through by visiting the store to purchase (some of) the items they researched online. A consumer's (mental) shopping list dictates the order in which they search the items on the retailer's website. The first item on the shopping list can be understood as the "lead category" for that given consumer (Chen et al. 1999).

4. Store sales moderated by effort: once at the store, consumers try to purchase all the items on their shopping list but might give up on some if they run out of time or if they are not willing to exert the necessary effort to find and fetch the item.

This sequential funnel makes some assumptions regarding customer behavior. First, it requires, implicitly, that consumers highly value their time, so they make their store choice based on proximity, and limit their willingness to shop to fill functional needs, thereby disregarding potential impulse purchases that would require extra effort for a small additional utility. This assumption is reasonable for a grocery or home improvement store chain in which all stores are alike and that carry most of the products offered online.

Second, we ignore competing stores. Note that we are not assuming that consumers are captive to a particular store, but rather that households are representative of the demand faced by the neighboring stores, even if they do not necessarily shop there.

Third, the shopping list assumption can be justified in retail settings where choices are made before entering the store. Hence, there is prior choice set that is mostly unaffected by the layout. This assumption is consistent with choice models where each customer has a preference list. In the literature, these products are substitutes and the customer ends up buying a single, preferred product out of the available ones. In our context, we extend this view to consider a preference list of complementary products, so this can be interpreted as a shopping list.

Fourth, the webrooming assumption is based on a common pattern observed in omnichannel retailing. In fact, industry reports show that the percentage of shoppers doing online research prior to visiting the store can range from 69% to 88% (Accenture 2013, Harris 2013, Deloitte 2017).

Finally, the moderation effect that effort has on sales is justified by the value of time premise. This assumption is consistent with behavioral models in which consumers have a time budget for in-store purchases (Hui et al. 2009b), and is more amenable to functional products such as home improvement, for which the time spent enjoying the store experience is not a major driver of conversion.

### 3.2 An Empirical Specification

Our empirical specification is based on the assumptions presented in the previous section. Conceptually at a high level, it has the following form:

$$sales_{ist} = \alpha_i + \alpha_s + \alpha_t + f\big(online\_visits_{ist}, effort_{is}\big) + \epsilon_{ist} \tag{1}$$

The dependent variable $sales_{ist}$ should be considered in log form, so as to justify an additive structure of independent drivers. The terms $\alpha_i, \alpha_s, \alpha_t$ correspond to product ($i$), store ($s$) and time ($t$) fixed effects, which represents the baseline demand. The next term amplifies demand as a function of webrooming moderated by effort, through a generic function $f(\cdot)$ that increases with online visits and decreases with effort. Here, $online\_visits_{ist}$ represents a vector of relevant metrics that characterize online traffic, and $effort_{is}$ should capture the time (disutility) involved in finding product $i$ at store $s$. Note that the latter excludes the fixed time/cost it takes to arrive to the store, which would be captured by the store fixed effect. Finally, $\epsilon_{ist}$ is the usual error term.

A few more remarks are noteworthy. We consider two amplification components in Equation (1). Namely, (i) primary demand: people that came to the store with the intention of buying, and exerted the effort to find the product; and (ii) secondary demand: people that came to the store searching for something else, but got exposed to the product and ended up buying (spillover in path, spillover nearby). Both components should be captured by $online\_visits_{ist}$.

We can observe that there are no substitution effects included in Equation (1). This formulation is appropriate when different products are solutions to non-overlapping functional needs. As will be seen next, our analysis is performed at the subcategory level, and hence substitution effects across categories should be negligible.

Finally, the effect of store execution, via assortment and service level decisions, is captured by the store fixed effect. Specifically, note that the model in Equation (1) does not take into account inventory, which could censor demand when there are stockouts. This can pose a challenge. However, in our case it was not a major issue because the service level of our industrial partner was overall high.

# 4 Application to Home Improvement Retailing

## 4.1 Context

We collaborated with a South American chain of home improvement stores, a leader in this industry, which operated 61 stores across Chile and an online channel at the time of the collaboration. We obtained a comprehensive proprietary dataset providing information about stores, products and customer interactions, which we describe below.

The retailer sells a variety of home improvement products, tools and materials. For the sake of illustration, the items in the assortment belong to categories such as Paint and Accessories, Wood, Plumbing, Gardening, Decor, Furniture, Lighting or Car Accessories, to name a few. The same assortment is sold in stores and online. In the products available in the data, we list 380,134 SKUs that are categorized in different hierarchical levels in the following manner: 5 Level-D clusters, 21 Level-0 clusters, 168 Level-1 clusters, 787 Level-2 clusters, and finally other more fine-grained clusters.

At this retailer, the weight of the online channel is small, as it is responsible for only 2.63% and 6.20% of total tickets and revenues, respectively. At the same time, in this industry webrooming is known to be an important factor affecting the shopping process; for instance, Home Depot states that it influences about 60% of store purchases even though the online channel only contributes to 6% of sales (Digital Commerce 360 2017). Because products are functional and product research is typically done in advance, this seems to be the ideal setting to assume that customers build a shopping list before entering the store, and to empirically connect online browsing to purchases.

Three types of data are available to us, which reflect customer behavior in online and offline channels:

- *Transaction data.* It describes the subset of the assortment's products that are purchased together. Each product bought belongs to a ticket, which is assigned to a physical store and a date. We refer to the purchase data as shopping-cart or shopping-basket data hereafter. From the raw information, we compute how many tickets issued by a certain store in a certain date included products of each category.
- *Clickstream data.* It describes the online journey that potential customers navigate when

visiting the retailer's website. It consists of time-stamped observations of product-level visits, with an IP address identifier (totalling 3,691,442 different identifiers). We refer to the clickstream data as shopping-list data or webrooming data hereafter. To process clickstream data, we first define a session as the web journey that a potential customer (given by an IP identifier) navigates in one natural day. One session is formed by a list of ranked products, represented in a ranked vector. IP identifiers are geolocated, so we are able to associate each session with stores nearby. Specifically, the catchment area of a given store is a 5km radius for stores in the Santiago Metropolitan area and 20km elsewhere. There is one special IP identifier that is worth mentioning: it corresponds to a gateway assigned to all wireless connections from mobile networks. Despite this point being geolocated in Santiago, it comprises all the mobile connections that originate in Chile. For this reason, mobile traffic cannot reliably be assigned to a nearby stores, so it is left out of our analysis. Furthermore, we do not consider web visits that are thought to be generated by bots through web-scraping. To remove those visits, we filter the visits by those IP identifiers that either visit one product more than 40 times or visit more than 300 products in one certain day.

- *Store layout data*. It describes the layout of the store, i.e., it details each Level-1 category's location in each store. The layouts of 16 brick-and-mortar stores are available in *pdf* files. We process these files automatically and we obtain the locations of the category labels within the layouts. These locations are described in $(x, y)$ coordinates, and measured in pixels, but for each file, the scale conversion is available, through the width of checkout corridors which measures 1.65 meters. Hence, we can compute the distance in meters that a potential customer has to walk in the retailer's store, so as to buy a product. From this map, we can thus compute the distance between products and between an item, the store entrance and the checkout lanes. We use Manhattan distances in meters, so as to reflect the true walking distance given the existence of horizontal and vertical aisles in the stores.

Given the information about layouts, we focus our study on 16 of the retailer's brick-and-mortar stores (26.2% of the total), and its online channel. From these stores, nine are located in the Santiago Metropolitan area, while the remaining stores belong to other regions. We use daily data from December 1st, 2018 to June 30th, 2019, with the exception of 19 days that were removed from the analysis due to missing values. The total study period is thus 30 weeks long.

Tables 1 and 2 compare the number of tickets and the composition of basket value and size for these 16 stores compared to the entire network. We observe that our subsample has stores that are slightly larger in scale (number of tickets) but does not significantly change the basket composition, hence suggesting that no bias is introduced by focusing on our chosen store subset.

Table 1: Week and store distribution of tickets in stores

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| # tickets per week & store (16 stores) | 468 | 14,804.69 | 5,844.029 | 711 | 10,487 | 18,471 | 34,428 |
| # tickets per week & store (all stores) | 1,788 | 11,677.26 | 5,395.334 | 711 | 7,662.2 | 14,659 | 34,428 |

Table 2: Basket value and size distribution in stores

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Focal sample (16 stores) | | | | | | | |
| Basket value | 6,928,597 | 39,224.43 | 138,006.5 | 0 | 6,110 | 39,280 | 71,200,000 |
| # categories per basket | 6,928,597 | 2.180 | 1.798 | 1 | 1 | 3 | 36 |
| All stores | | | | | | | |
| Basket value | 20,878,950 | 39,242.500 | 124,958.1 | 0 | 5,990 | 38,600 | 136,596,250 |
| # categories per basket | 20,878,950 | 2.161 | 1.780 | 1 | 1 | 3 | 39 |

## 4.2 Descriptive Statistics

In this section we operationalize the variables from our conceptual model (see Section 3). The data is aggregated weekly to avoid within-week fluctuations: each observation corresponds to a week $t$, a Level-1 category $i$ (which we call product for simplicity), and a store $s$. Hence, we define the following variables of interest:

- $N_{ist}$: Number of tickets that include product $i$ issued at store $s$ during week $t$.
- $N_{st}$: Total number of tickets issued at store $s$ during week $t$. Note that $N_{st} \leq \sum_i N_{ist}$ because a ticket may include multiple products.
- $V_{1,ist}$: Number of online sessions in which product $i$ is viewed as the first item, within the catchment area of store $s$ during week $t$.
- $V_{2-4,ist}$: Number of online sessions in which product $i$ is viewed as the second, third, or fourth item, within the catchment area of store $s$ during week $t$.
- $V_{>4,ist}$: Number of online sessions in which product $i$ is viewed as the fifth item or further, within the catchment area of store $s$ during week $t$.
- $V_{ist}$: Number of online sessions in which product $i$ is viewed in any order, within the catchment area of store $s$ during week $t$. It follows that $V_{ist} = V_{1,ist} + V_{2-4,ist} + V_{>4,ist}$.
- $V_{st}$: Total number of online sessions within the catchment area of store $s$ during week $t$. It follows that $V_{st} = \sum_i V_{ist}$.

- $D_{is}$: Distance to pick item $i$ in store $s$ measured in meters, i.e., the distance between the store entrance and product $i$ plus the distance between product $i$ and the checkout lanes.

In our study, we use the variables in log form for ease of interpretation of the coefficients and to remove skewness, i.e., we transform variable $X$ into $x := \log(1+X)$ (we add one to avoid problems with zero values of $X$). With this notation, the variable $n_{ist}$ is our proxy for $sales_{ist}$, and $d_{is}$ is our for proxy $effort_{is}$. Our proxy for $online\_visits_{ist}$ includes $v_{ist}, v_{1,ist}, v_{2-4,ist}, v_{>4,ist}$, and might also include these same variables for other products $j$ whose traffic is relevant to the sales of product $i$.

Table 3 contains the descriptive statistics of the logged variables, and Table 4 their correlations. One can observe that the amount of generic online traffic $v_{st}$ has a small correlation with sales indicators $n_{st}$ or $n_{ist}$. However, product-specific online traffic $v_{ist}, v_{1,ist}, v_{2-4,ist}$ and $v_{>4,ist}$ has a high positive correlation with product-specific sales $n_{ist}$. This indicates that indeed online activity can be used as a key input for store demand forecasting, and this insight is a promising starting point to develop a more sophisticated model as discussed in §3.

Table 3: Descriptive statistics of the main model variables, log-transformed.

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| $n_{ist}$ | 80,190 | 4.057 | 1.917 | 0.000 | 3.091 | 5.485 | 8.713 |
| $n_{st}$ | 80,190 | 9.283 | 1.550 | 0.000 | 9.230 | 9.824 | 10.447 |
| $v_{1,ist}$ | 80,190 | 2.901 | 1.449 | 0.000 | 1.946 | 3.932 | 7.771 |
| $v_{2-4,ist}$ | 80,190 | 2.715 | 1.377 | 0.000 | 1.792 | 3.689 | 7.594 |
| $v_{>4,ist}$ | 80,190 | 1.819 | 1.263 | 0.000 | 0.693 | 2.708 | 6.632 |
| $v_{ist}$ | 80,190 | 3.668 | 1.479 | 0.000 | 2.773 | 4.691 | 8.419 |
| $v_{st}$ | 80,190 | 8.660 | 0.839 | 6.737 | 8.013 | 9.308 | 10.239 |
| $d_{is}$ | 80,190 | 4.694 | 0.439 | 3.386 | 4.368 | 5.026 | 5.936 |

Table 4: Correlation matrix between the variables of interest (log).

| | $n_{ist}$ | $n_{st}$ | $v_{1,ist}$ | $v_{2-4,ist}$ | $v_{>4,ist}$ | $v_{ist}$ | $v_{st}$ |
|---|---|---|---|---|---|---|---|
| $n_{ist}$ | 1 | | | | | | |
| $n_{st}$ | 0.382 | 1 | | | | | |
| $v_{1,ist}$ | 0.446 | -0.037 | 1 | | | | |
| $v_{2-4,ist}$ | 0.355 | -0.056 | 0.888 | 1 | | | |
| $v_{>4,ist}$ | 0.263 | -0.065 | 0.755 | 0.883 | 1 | | |
| $v_{ist}$ | 0.415 | -0.049 | 0.952 | 0.967 | 0.869 | 1 | |
| $v_{st}$ | -0.007 | -0.062 | 0.515 | 0.600 | 0.605 | 0.577 | 1 |

To further illustrate the available data, Figure 1 shows the joint evolution of $V_{ist}$ and $N_{ist}$ for two stores and two products. We can see that both series tend to move together, although their relative values (i.e., their ratio) changes across stores and products, which is natural given that some categories may require relatively more browsing to achieve a certain level of sales, and the customers around some stores may have higher natural conversion between browsing and purchasing, compared to others. These structural, static differences will be captured by store and product fixed effects in our model.

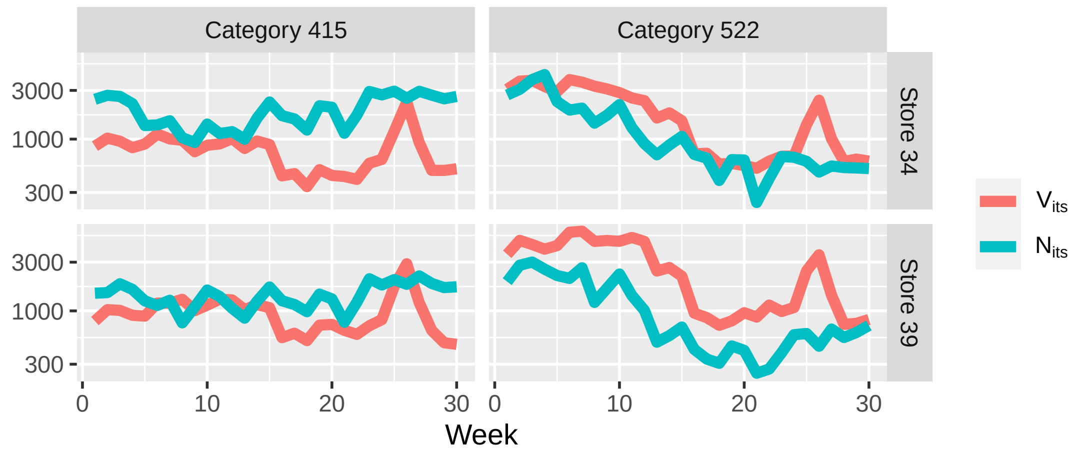


Figure 1: Evolution of clickstream and sales figures, for two products and two stores.

As shown in Table 4, $n_{ist}$ and $v_{ist}$ are highly correlated. This suggests that clickstream activity seems a useful lead indicator of product $i$'s performance. Taking into account this relationship, we can further study the impact of the store layout on conversion. Figure 2 plots conversion for a given store, measured as $n_{ist} - v_{ist} = \log\left((1 + N_{ist})/(1 + V_{ist})\right) \approx \log\left(N_{ist}/V_{ist}\right)$, averaged over 30 weeks. We observe that, while conversion fluctuates, we see a clear trend showing that the conversion of distant products is lower than those near the entrance or center. This model-free evidence suggests that a product's location in the store strongly affects the conversion from product interest to actual sales.

### 4.3 Results

We can now rewrite Equation (1) into a main specification:

$$n_{ist} = \alpha_i + \alpha_s + \alpha_t + \beta v_{ist} + \gamma d_{is} + \epsilon_{ist} \tag{2}$$

We also consider variations of this specification, in which we incorporate quality-segregated online visits, via $v_{1,ist}, v_{2-4,ist}, v_{>4,ist}$ instead of $v_{ist}$; categorical values of $d_{is}$; interactions between $v_{ist}$ and $d_{is}$; or spillover effects. We estimate Equation (2) using standard Ordinary Least Squares.

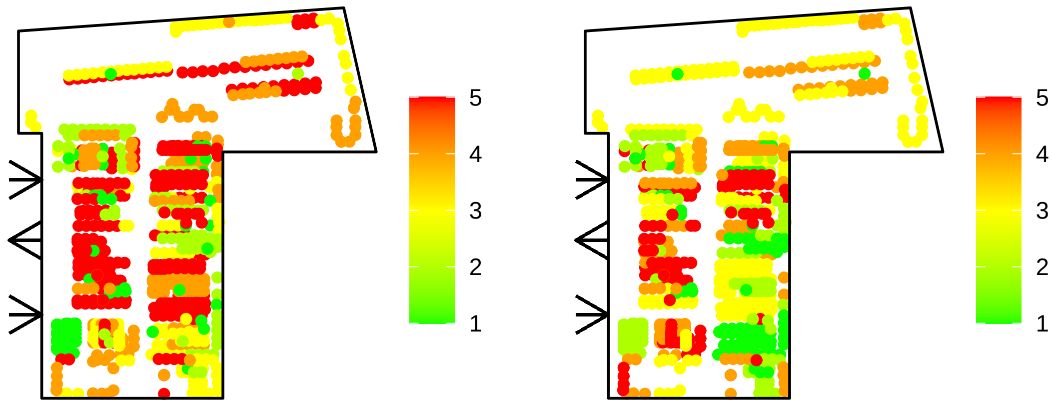


Figure 2: Model-free representation, with sales (left) and conversion (right) in quintiles. The arrows correspond to store entrances and exit.

We first study the use of nearby online interactions as determinants of store sales. For that purpose, we set $\gamma = 0$ in Equation (2). The results are reported in Table 5. Model (1) presents a benchmark model that only incorporates fixed effects for product, week and store. As we can see, fixed effects only lead to a $R^2$ of 0.87, which suggests that cross-product and cross-store heterogeneity, as well as seasonality (cross-week variation), are high in our context. Model (2) incorporates the total online traffic for each store and week, in the same way Gallino and Moreno (2014) used online interactions as a driver of store sales. In comparison with them, we find general online traffic is non-significant and does not help predict product-level sales, suggesting that accounting both for store and time variation through fixed effects is sufficient and online traffic simply contains redundant information. In contrast, when we consider product-level online interactions in Models (3) and (4), prediction improves steeply, to $R^2 = 0.92$. This implies that product-level clicks provide a strong signal about sales. Moreover, the coefficient in Model (3) is equal to 1.0382 and highly significant, which suggests that the relationship between clicks and sales is approximately proportional, i.e., we can write $N_{ist} \approx kV_{ist}$. In other words, if online clicks double, sales also double. Model (4) breaks down clicks into different 'quality grades', by considering separately clicks in which the focal product was the first one in the session (the sequence of products viewed by the consumer; the first one should be the most important for the consumer), and the clicks in which the focal product was in positions 2 to 4, or $> 4$. We can observe that indeed clicks in the first position have the highest coefficient 0.7477, while later clicks had lower coefficients 0.2864 and 0.0716 (all of them are statistically significant). This supports our interpretation that online interactions are a proxy for true consumer interest, and it is revealed especially when it appears early in the online search sequence of the consumer.

With Models (1) through (4) we have established that online interactions are significant drivers of sales. This is a direct influence where sales are simply the consequence of a true demand need existing prior to a store visit. Beyond this direct influence, the literature has identified other indirect influences, namely spill-over effects between products. In other words, if there is a flow of shoppers interested in buying a certain product, these visitors will be exposed to other products on their way to their primary shopping objective. We thus consider two additional drivers of sales arising from spill-over effects.

First, for a certain product $i$, we consider the primary demand associated with products $j$ that require the shopper to walk by $i$ in their path to $j$. For this purpose, we define the binary variable $INPATH_{ijs}$ which is equal to one if the shortest path from entrance to $j$ to exit, and the shortest path from entrance to $i$ to $j$ to exit coincide, i.e., they have the same distance; and zero otherwise. We then define $PATH_{ist}$ as the number of online sessions within 5km of the store that include any product $j \neq i$ such that $INPATH_{ijs} = 1$:

$$PATH_{ist} = \sum_{j \neq i} INPATH_{ijs} \times V_{jst}, \tag{3}$$

and let $path_{ist} = \log(1 + PATH_{ist})$. This variable should thus capture spill-overs into items that are in central locations within the store, that see a high amount of traffic for primary items that are further inside the store.

Second, we consider the primary demand of products $j$ in the vicinity of $i$. We thus define the binary variable $NEARBY_{ijs}$ which is equal to one if the distance between $i$ and $j$ is less than 20 meters. We then let $NEAR_{ist}$ as the number of online sessions within 5km of the store that include any product $j \neq i$ such that $NEARBY_{ijs} = 1$:

$$NEAR_{ist} = \sum_{j \neq i} NEARBY_{ijs} \times V_{jst}, \tag{4}$$

and let $near_{ist} = \log(1 + NEAR_{ist})$. This variable captures spill-overs related to proximity to store hot spots.

We incorporate these two variables in Model (5). We can see that primary demand $v_{ist}$ remains significant and with a coefficient similar to that in Model (3). In contrast, $near_{ist}$ is not significant and $path_{ist}$ is negative and significant but very small in magnitude. This suggests that spill-over effects are negligible, which is understandable given the functional, non-impulse nature of the products sold in our home improvement context. Another possible interpretation of this result is that webrooming informs a more focused consumer that will spend less time roaming at the store, and therefore, opportunities for cross-selling are diminished. Other authors have discussed similar effects of webrooming, see for instance Gao and Su (2017a).

The previous models establish that online clicks are a valuable determinant of store sales. We can now study the impact of product location on sales, corresponding to Equation (2) with $\gamma \neq 0$. We thus operationalize location through the ease of access to the product in the store,

Table 5: Models using online interactions.

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | $n_{ist}$ | | | | |
| | (1) | (2) | (3) | (4) | (5) |
| $v_{st}$ | | $-0.0749$ | | | |
| | | (0.0781) | | | |
| $v_{ist}$ | | | $1.0382^{***}$ | | $1.0382^{***}$ |
| | | | (0.0044) | | (0.0044) |
| $v_{1,ist}$ | | | | $0.7477^{***}$ | |
| | | | | (0.0075) | |
| $v_{2-4,ist}$ | | | | $0.2864^{***}$ | |
| | | | | (0.0102) | |
| $v_{>4,ist}$ | | | | $0.0716^{***}$ | |
| | | | | (0.0086) | |
| $path_{ist}$ | | | | | $-0.0006^{***}$ |
| | | | | | (0.00004) |
| $near_{ist}$ | | | | | $-0.0001$ |
| | | | | | (0.0001) |
| Fixed effects | week product store | week product store | week product store | week product store | week product store |
| Observations | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 |
| $R^2$ | 0.8707 | 0.8707 | 0.9243 | 0.9249 | 0.9244 |
| Adjusted $R^2$ | 0.8706 | 0.8706 | 0.9242 | 0.9248 | 0.9244 |
| Residual Std. Error | 1.6142 | 1.6142 | 1.2352 | 1.2300 | 1.2338 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

via the distance from entrance to product to exit. Table 6 shows the result of the estimation. The table shows two different ways of incorporating distance, directly as a continuous variable and as a piece-wise constant function with intervals of 50 meters. We first observe, in Models (6) and (7), that distance alone (without online clicks) is significant but only marginally improves the result of Model (1). In contrast, Models (8) and (10) improve on Models (3) and (4). The coefficients for $v_{ist}$ and $v_{k,ist}, k \in \{1, 2-4, > 4\}$ do not change, which means that the role of distance, driver of conversion, seems orthogonal to that of online interactions, a proxy for true consumer needs. These models show a strong effect of distance, with a coefficient of -0.0012 and -0.0013 respectively. This is a relatively high value: distance within the store roughly varies between 50 and 250 meters, which implies that the difference in sales between closest and furthest products is about $-0.0012 \times (250 - 50) = -0.24$, a 21% decrease (since $e^{-0.24} = 0.79$). Similarly, Models (9) and (11) show that the effect of distance is monotonic, with higher distances reducing sales more and more, with a drop of -0.24 ($= -0.46 - (-0.22)$) between the closest products to the furthest ones. Figure 3 graphically compares the effect of distance on sales in Models (10) and (11).

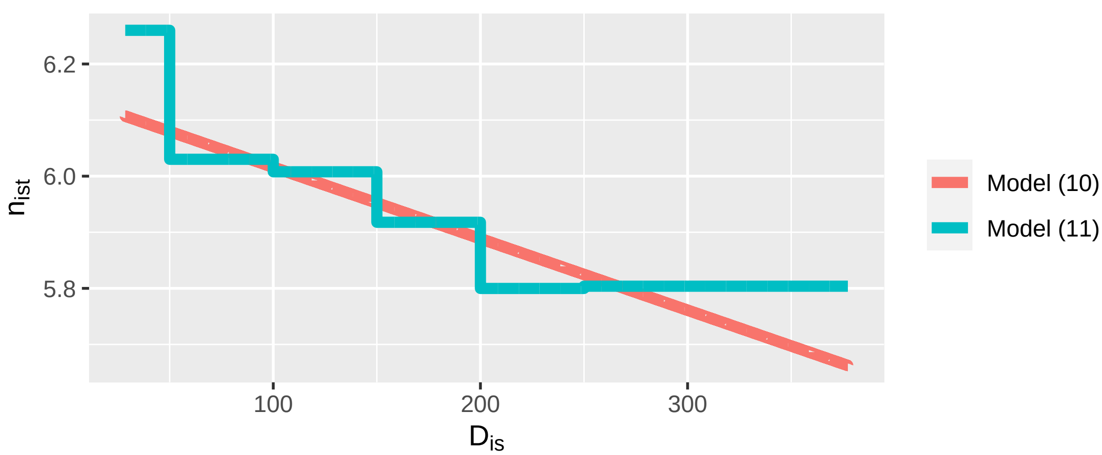


Figure 3: Variation of sales as a function of in-store distance in Models (10) and (11).

### 4.4 Robustness

While our main models in Tables 5 and 6 are kept simple to focus on the direct impact of online visits and effort, we run several robustness checks to discard possible confounders and to identify possible interactions between model variables. We discuss below the findings and include details in the Appendix.

First, there could be store-specific factors that influence product success and seasonality, which may be related to online visit patterns and thus bias our estimation. To discard this possible confounder, we replicate the estimation of our main Models (9) and (11) with alternative fixed effect configurations, which incorporate possible interactions between week, product and store.

Table 6: Models using online interactions and product location.

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | $n_{ist}$ | | | | | |
| | (6) | (7) | (8) | (9) | (10) | (11) |
| $v_{ist}$ | | | $1.0387^{***}$ | $1.0389^{***}$ | | |
| | | | (0.0044) | (0.0044) | | |
| $v_{1,ist}$ | | | | | $0.7491^{***}$ | $0.7500^{***}$ |
| | | | | | (0.0075) | (0.0075) |
| $v_{2-4,ist}$ | | | | | $0.2873^{***}$ | $0.2872^{***}$ |
| | | | | | (0.0102) | (0.0101) |
| $v_{>4,ist}$ | | | | | $0.0689^{***}$ | $0.0681^{***}$ |
| | | | | | (0.0086) | (0.0086) |
| $D_{is}$ | $-0.0009^{***}$ | | $-0.0012^{***}$ | | $-0.0013^{***}$ | |
| | (0.0002) | | (0.0001) | | (0.0001) | |
| $\mathbb{1}_{D_{is}\in[50,100]}$ | | $-0.2467^{***}$ | | $-0.2208^{***}$ | | $-0.2299^{***}$ |
| | | (0.0296) | | (0.0227) | | (0.0226) |
| $\mathbb{1}_{D_{is}\in[100,150]}$ | | $-0.2448^{***}$ | | $-0.2501^{***}$ | | $-0.2522^{***}$ |
| | | (0.0309) | | (0.0237) | | (0.0236) |
| $\mathbb{1}_{D_{is}\in[150,200]}$ | | $-0.3223^{***}$ | | $-0.3325^{***}$ | | $-0.3422^{***}$ |
| | | (0.0333) | | (0.0255) | | (0.0254) |
| $\mathbb{1}_{D_{is}\in[200,250]}$ | | $-0.3278^{***}$ | | $-0.4325^{***}$ | | $-0.4602^{***}$ |
| | | (0.0430) | | (0.0328) | | (0.0327) |
| $\mathbb{1}_{D_{is}\in[250,\infty]}$ | | $-0.3976^{***}$ | | $-0.4341^{***}$ | | $-0.4562^{***}$ |
| | | (0.0559) | | (0.0427) | | (0.0425) |
| Fixed effects | week | week | week | week | week | week |
| | product | product | product | product | product | product |
| | store | store | store | store | store | store |
| Observations | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 |
| $R^2$ | 0.8707 | 0.8708 | 0.9244 | 0.9245 | 0.9250 | 0.9251 |
| Adjusted $R^2$ | 0.8706 | 0.8707 | 0.9243 | 0.9244 | 0.9250 | 0.9251 |
| Residual Std. Error | 1.6138 | 1.6132 | 1.2344 | 1.2336 | 1.2291 | 1.2282 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

The results are shown in Models (12) to (17) in Table 7. We confirm that our main findings are preserved.

Second, even though we do have cross-store layout variation, one may think that, if stores are far apart, then customers may be intrinsically different so in reality separate estimations should be conducted for each of the stores. To remedy this, we focus on the nine stores located in the Santiago Metropolitan area, which should serve a common pool of customers with homogeneous tastes. We replicate the estimation Models (6) through (11) with the data from these nine stores, see Table 8. Again, we observe that the main findings are preserved.

Third, we may be concerned that online visits may be endogenous and influenced by external shocks that are also moving sales. To alleviate this concern, for each store we measure the online interactions that happen far from it – a distance of more than 10km for stores in Santiago Metropolitan area and more than 40km for stores in other region – as a Hausman-type instrument for online visits. We argue that the online visits far from the store of interest will be exogenous and correlated with the visits near the store, but uncorrelated with the dependent variable for sales. Thus, this approach protects against possible endogeneity of the $v_{ist}$ variables. Table 9 replicates Models (8) through (11) using a two-stage least squares (2SLS) approach. We find that the results of the 2SLS do not change our findings.

Finally, in our main specification the effect of distance is assumed to be independent of the amount of online visits. We consider a possible interaction between distance and online traffic to enrich Model (9), shown in Model (19) in Table 7. We see that the coefficients are similar in sign and size. The main insight from this specification is that the impact of distance is smaller for items with higher online activity, suggesting that products with a high amount of webrooming may be less sensitive to in-store location.

# 5 Store Layout Optimization

## 5.1 A Product-Position Assignment Problem

Our model assumes and empirically demonstrates that a product's location within the store has a significant impact on its conversion. In this section, we are interested in prescribing improved layouts that increase total sales, taking consumer true needs captured via online interactions as fixed.

Product location optimization is a relatively well-studied area of research, mainly in warehouse settings, see De Koster et al. (2007) for a review. In these contexts, one usually minimizes picking costs, which results in placing high-rotation products in easily accessible locations, while slow-movers are sent to more costly locations. In a store, the costs to bring a product to the shelf are relatively small and insensitive to location within the store. As a consequence, we focus on the main driver of profits coming from the impact of product location on sales conversion.

We can formulate the layout design problem as the following assignment problem. Let $x_{istp} = 0,1$ denote whether product $i \in \mathcal{I}$ was located in position $p \in \mathcal{P}$, in store $s$ and time period $t$. One product can go into one position, and one position can only take one product.

Let $d_p$ be the distance a consumer must travel from the entrance when a product is located in position $p$ (at a given store). Then, the location-dependent demand of product $i$ can be written as $r_{istp} = r_i d_{istp}$, where $d_{istp} = exp\Big(\alpha_i + \alpha_s + \alpha_t + \beta v_{ist} + \gamma d_p\Big)$, and $\beta$ and $\gamma$ come from the estimation of Model (8). We can now formulate the layout design problem as

$$J_{st} := \max_x \sum_{i\in\mathcal{I}} \sum_{p\in\mathcal{P}} r_{istp} x_{istp} \tag{5}$$

$$s.t. \sum_{i\in\mathcal{I}} x_{istp} \leq 1 \quad \forall p \in \mathcal{P} \tag{6}$$

$$\sum_{p\in\mathcal{P}} x_{istp} \leq 1 \quad \forall i \in \mathcal{I} \tag{7}$$

$$x_{istp} \in \{0,1\}. \tag{8}$$

Note that this formulation only includes constraints pertaining to the impossibility of placing two products in the same location, or one product being sent to two locations. It is easy to incorporate additional linear constraints reflecting business conditions for the product in the store. For example, if a product can only be located in a particular part of the store, then we can set $x_{istp} = 0$ for infeasible locations. If products $i$ and $j$ must be adjacent, then we can set $x_{istp} \leq \sum_{p'} A_{pp'} x_{jstp'}$ with $A_{pp'} = 1$ if $p$ and $p'$ are adjacent and zero otherwise; in other words, if $x_{istp} = 1$, then one adjacent $p'$ (such that $A_{pp'} = 1$) is such that $x_{jstp'} = 1$.

In the absence of additional constraints, Equations (6)-(7) make a Totally-Unimodular Matrix (TUM), and hence constraint $x_{istp} \in \{0,1\}$ can be replaced with $0 \leq x_{istp} \leq 1$ without changing the optimal solution of (5). In other words, $J_{st}$ can be obtained by solving a linear program. Otherwise, we solve an integer program.

Notice that we can write $r_i d_{istp} = \bar{r}_{ist} g_p$, which will allow us to find the optimal assignment in closed form. Indeed, we can write the objective as $\sum_{i\in\mathcal{I}} \sum_{p\in\mathcal{P}} \bar{r}_{ist} g_p x_{istp}$. This is maximized by assigning the largest $g_p$ to the largest $\bar{r}_{ist}$: assign the best in-store position (highest $g_p$) to the best-selling product (highest $\bar{r}_{ist}$).

## 5.2 Improving on Existing Layouts

We can now apply the method of §5.1 to reengineer the actual layouts observed in our data. We first provide an in-depth analysis for one store and then provide results for the complete set of stores.

We define positions $p$ in the same way as products, i.e., $p$ denotes the actual position of product $i = p$. We compute $r_i$ to be equal to the average spending per ticket in product $i$ over the season of 30 weeks. We force the assignment to remain stable, and hence define decision variables $x_{isp} = 0, 1$ in the following decision problem:

$$J_s(z) := \max_x \sum_{i \in \mathcal{I}} \sum_{p \in \mathcal{P}} \left( \sum_{t \in \mathcal{T}} r_{istp} \right) x_{isp} \tag{9}$$

$$s.t. \sum_{i \in \mathcal{I}} x_{isp} \leq 1 \quad \forall p \in \mathcal{P} \tag{10}$$

$$\sum_{p \in \mathcal{P}} x_{isp} \leq 1 \quad \forall i \in \mathcal{I} \tag{11}$$

$$\sum_{\substack{p \in \mathcal{P} \\ p = i}} x_{isp} \geq |\mathcal{P}| - z \tag{12}$$

$$x_{isp} \in \{0, 1\} \tag{13}$$

In contrast to Equation (5), Equation (9) considers a one-time product assignment change that applies to the entire season (hence $x_{isp}$ does not depend on $t$). It also includes an additional parametric constraint $\sum_{p \in \mathcal{P}, p=i} x_{isp} \geq |\mathcal{P}| - z$, where $z$ is an integer variable. This constraint limits the number of actual product assignment changes to be $z$ at the most. For example, if $z = 0$, the only feasible solution is to set $x_{isp} = 1$ when $i = p$ and zero otherwise. If $z = |\mathcal{P}|$, then the constraint is innocuous. When $z$ takes intermediate values, it provides us with interventions with varying degrees of complexity. Note, however, that constraint (12) breaks the TUM structure of the constraint matrix, and thus requires us to solve a set of integer programs. In addition, we consider two versions of the decision set $\mathcal{P}$: one that includes all products, and another one that excludes construction products that are typically bulkier and located at the side of the store, and hence, are difficult to place in any other store position.

Consider store 51, depicted in Figure 2. In this store, we have 168 different products (Level-1 categories) assigned to 168 positions shown in the map. Figure 4 shows the normalized value of $J_s(z)$ as a function of the number of changes allowed $z$. Of course, our formulation ignores all other business constraints, e.g., adjacencies, space limitations, etc. but still our results are useful to understand the potential of layout optimization as suggested by our empirical findings.

When there is no limit in the number of changes ($z = 168$), a rearrangement of the layout increases revenues by 13.8% when all products can be moved. The value decreases to 7.5% when construction products are excluded from the optimization. These are both significant lifts for a home improvement retailer where margins are thin and increasing the top line typically has a very strong effect on net margins. When examining in more detail the type of changes recommended in this scenario, we see that top-selling products located towards the back of the store are moved to the front, and some minor ones in front positions are relegated to the back of the store.

Additionally, Figure 4 informs us about how much complexity is needed to achieve a certain level of lift. Specifically, when all products can be moved, with just swapping two products with each other ($z = 2$), a revenue lift of 2.0% can be achieved; when changing 10 products, the lift is 7.3%, more than 50% of the unconstrained maximum lift possible 13.8%. The insight is similar when we exclude construction products. This suggests that with minimal effort, the retailer can achieve relevant improvements in its operations.

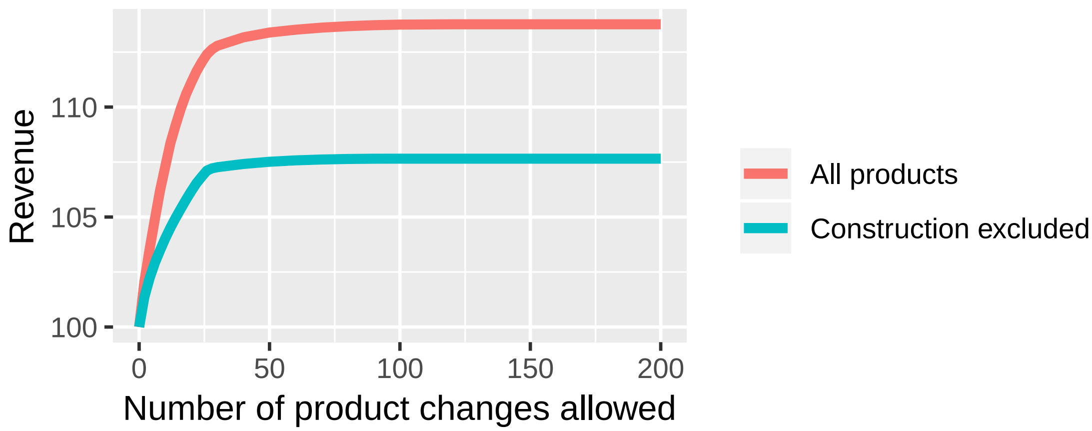


Figure 4: Expected revenue in store 51, after applying product-location assignment changes, as a function of the maximum amount of changes $z$ allowed. The revenue is shown relative to the revenue achieved with the current layout, $J_s(0)$.

Furthermore, through this type of analysis we can also evaluate the value of changing the layout on a recurrent basis. Namely, is there value in changing the layout every week, as opposed to just once? This may be a useful intervention motivated by the changing patterns of $v_{ist}$, which may change the potential of each product $\bar{r}_{ist}$ over time. To answer this question, we can compare $J_s^{recurring}(z) := \sum_{t \in \mathcal{T}} J_{st}(z)$ to $J_s(z)$. Figure 5 compares the values of $J_s^{recurring}(z)$ and $J_s(z)$ as a function of the number of changes made. We find that flexibility may just increase revenues by less that 0.5%, suggesting that a one-time intervention is sufficient to improve store performance. In other words, the order of products along their 'natural' revenue $\bar{r}_{ist}$ does not change significantly across periods $t$, and thus the optimal assignment given $z$ is stable over time.

Finally, we can extend this analysis to the entire network of 16 stores for which we can reengineer the layout. Figure 6 shows the distribution of the revenue improvements achieved with an unconstrained layout change and one limited excluding construction products. As we can see, the revenue lift can be significant, with one store achieving improvements of 17.5%, with the average in the sample being 10.2%. Average improvements are already 6.0% when the intervention excludes construction products.

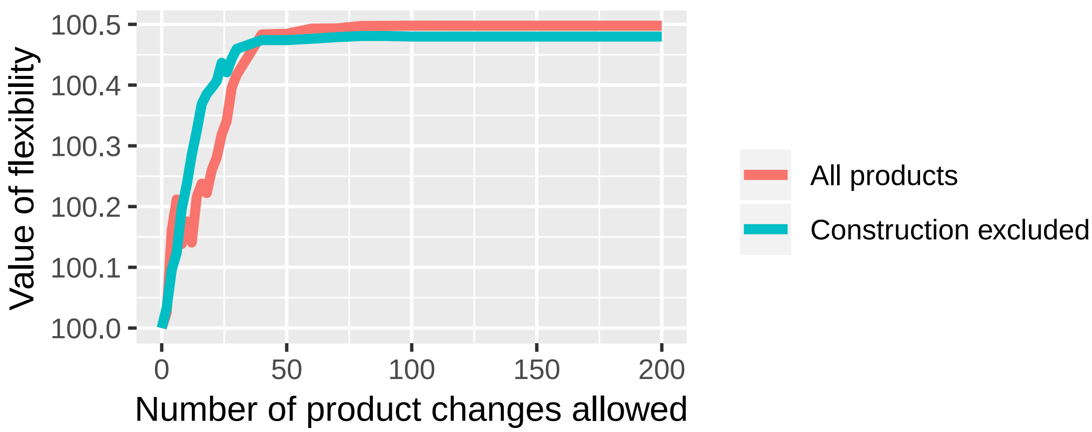


Figure 5: $J_s^{recurring}(z)/J_s(z)$ in store 51 as a function of the maximum amount of changes $z$ allowed.

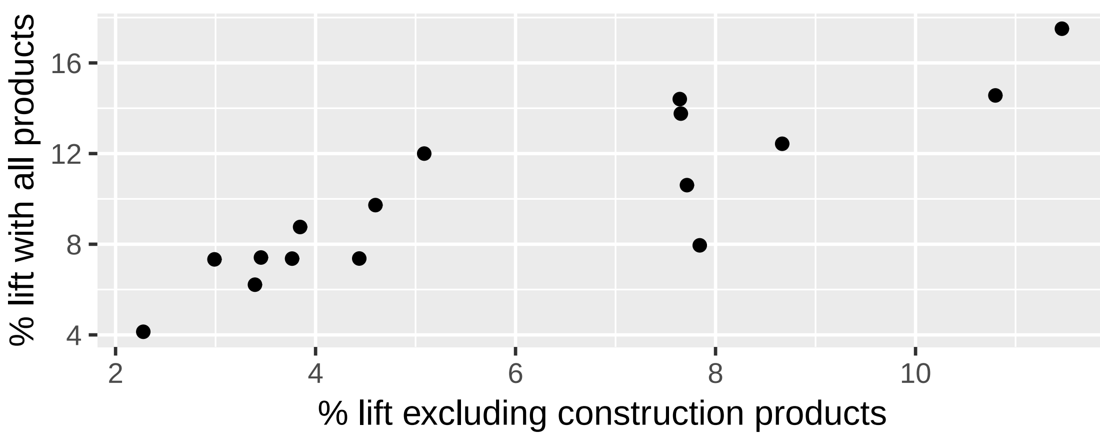


Figure 6: Distribution of $J_s(\infty)/J_s(0)$ across 16 stores (each point corresponds to a store).

# 6 Conclusion

In this paper, we have provided a new perspective on how omnichannel, via webrooming customer interactions, can help retailers manage better their physical stores. Specifically, we have posited that, when sales are preceded by a need that crystallizes into a shopping list and pre-purchase product search, then store sales are driven by both the amount of nearby online visits and the effort that it takes to fetch the product in the store. We validate our conceptual model with data from a home improvement chain, over multiple products, weeks and locations. The data provides variation of product interest and in-store location, and allow us to identify the effect of online visits and effort on sales. We find that sales grow proportionally with online visits, and that easy-to-reach store positions lead to significantly higher conversion (with about a 20% variation between best and worst locations). In addition, we generate evidence that there is negligible cross-selling in this context.

Our results furthermore have important implications for the management of physical stores. First, they suggest that layout reengineering can provide a tempting lift in revenues, of about 10% on average within the stores in our study. Second, they imply that the efforts to generate store visits, in the hope that they will generate unplanned purchases, may not be fruitful. In other words, it may be better that stores do not accept new roles as delivery points (Faithfull 2018, Jones 2019), if the products on sale are related to a functional need that requires previous research. Third, our results identify the effort to find products in the store as a hindrance to conversion. In other words, actions to make in-store product search simpler may lead to increased sales. One such action could be to provide product 'addresses' to consumers when they prepare their shopping lists, as Target does, see Figure 7.

This study highlights the importance of better understanding the role of store design on customer experiences. This is a promising direction for future research. Indeed, the adoption of Internet Of Things technologies in stores provides new data sources for a more granular understanding of the trajectories of customers over time (the funnel view) and space (transitions between home, work and shopping destinations). This requires the full digitalization of the store conditions, and precise product locations, a piece of information that to date is rarely available, with the exception of supermarket planograms, common in grocery retailing, or RFID sources, installed by Walmart or Zara among others. It can potentially reveal the causal impact of different interventions such as product viewing, product information provision, staff advice or fitting (Musalem et al. 2021), as well as environmental stimuli such as music or temperature (Martínez-de Albéniz and Belkaid 2021). Combining on-premise data with online interactions is particularly interesting, so that conceptual frameworks such as Bell et al. (2014) can be operationalized and translated into prescriptive advice for retailers.

## Acknowledgements

Victor Martínez-de-Albéniz acknowledges the financial support provided by the Agencia Estatal de Investigación (AEI) from the Spanish Ministry of Science and Innovation, with project reference PID2020-116135GB-I00 MCIN/ AEI / 10.13039/501100011033.

## References

Accenture 2013. Accenture Study Shows U.S. Consumers Want a Seamless Shopping Experience Across Store, Online and Mobile. Technical report, Accenture plc, Available: https://accntu.re/3oFRJo2. Last accessed Oct 7, 2021.

Aouad, A., V. Farias, and R. Levi. 2021. Assortment optimization under consider-then-choose choice models. *Management Science* 67 (6): 3368–3386.

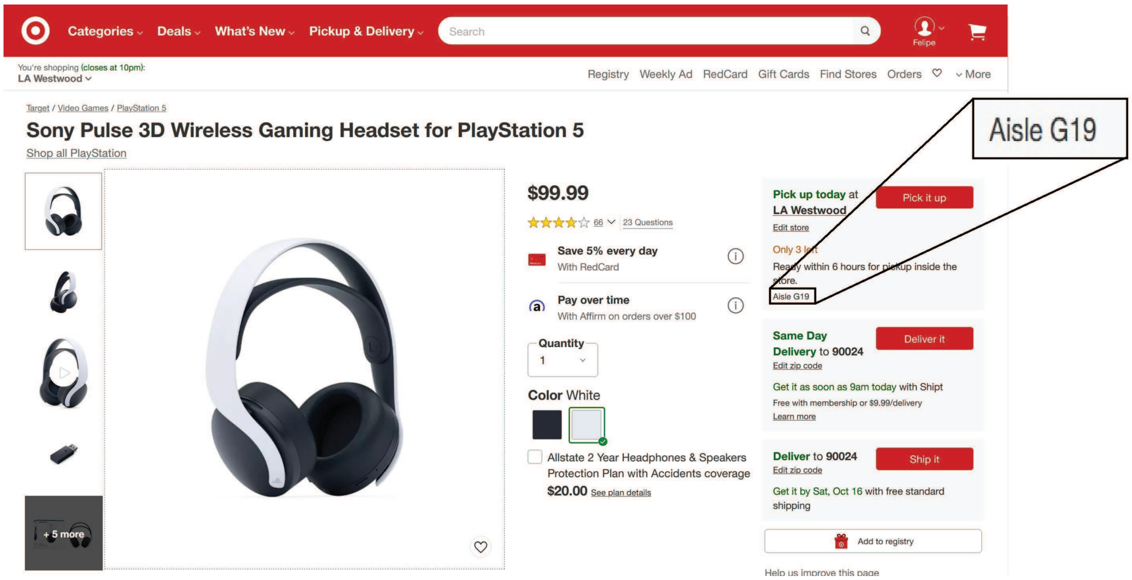


Figure 7: Standard product page on target.com, where the address of the product in the store of your choice is indicated.


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.

Bar-Gill, S., and S. Reichman. 2020. Stuck Online: When Online Engagement Gets in the Way of Offline Sales. *MIS Quarterly* Forthcoming:NA.

Bell, D., S. Gallino, and A. Moreno. 2015. Showrooms and Information Provision in Omni-channel Retail. *Production and Operations Management* 24 (3): 360–362.

Bell, D. R., S. Gallino, and A. Moreno. 2014. How to win in an omnichannel world. *MIT Sloan Management Review* 56 (1): 45.

Bell, D. R., S. Gallino, and A. Moreno. 2017. Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science* 64 (4): 1629–1651.

Bell, D. R., S. Gallino, and A. Moreno. 2020. Customer supercharging in experience-centric channels. *Management Science* 66 (9): 4096–4107.

Bijmolt, T. H., M. Broekhuis, S. De Leeuw, C. Hirche, R. P. Rooderkerk, R. Sousa, and S. X. Zhu. 2021. Challenges at the marketing–operations interface in omni-channel retail environments. *Journal of Business Research* 122:864–874.

Boada-Collado, P., and V. Martínez-de Albéniz. 2020. Estimating and optimizing the impact of inventory on consumer choices in a fashion retail setting. *Manufacturing & Service Operations Management* 22 (3): 582–597.

Brynjolfsson, E., Y. Hu, and D. Simester. 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* 57 (8): 1373–1386.

Brynjolfsson, E., Y. J. Hu, and M. S. Rahman. 2013. Competing in the Age of Omnichannel Retailing. *MIT Sloan Management Review* 54 (4): 23.

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

Caro, F., A. G. Kök, and V. Martínez-de Albéniz. 2020. The future of retail operations. *Manufacturing & Service Operations Management* 22 (1): 47–58.

Chen, Y., J. D. Hess, R. T. Wilcox, and Z. J. Zhang. 1999. Accounting profits versus marketing profits: A relevant metric for category management. *Marketing science* 18 (3): 208–229.

Chintagunta, P. K., J. Chu, and J. Cebollada. 2012. Quantifying transaction costs in online/off-line grocery channel choice. *Marketing Science* 31 (1): 96–114.

Chuang, H. H.-C., R. Oliva, and O. Perdikaki. 2016. Traffic-based labor planning in retail stores. *Production and Operations Management* 25 (1): 96–113.

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.

De Koster, R., T. Le-Duc, and K. J. Roodbergen. 2007. Design and control of warehouse order picking: A literature review. *European journal of operational research* 182 (2): 481–501.

DeHoratius, N., A. J. Mersereau, and L. Schrage. 2008. Retail inventory management when records are inaccurate. *Manufacturing & Service Operations Management* 10 (2): 257–277.

DeHoratius, N., and A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Management science* 54 (4): 627–641.

Deloitte 2017. 2017 pre-Thanksgiving pulse survey. Technical report, Deloitte Touche Tohmatsu Limited, Available: https://bit.ly/2IOKEg6. Last accessed Oct 7, 2021.

Digital Commerce 360 2017. Home Depot plans to spend $5.4 billion to sharpen its omnichannel strategy. *Digital Commerce 360* December 8:online.

Dowsett, S. 2019. Less is more? Inditex cuts stores but boosts space in home market Spain. *Reuters* June 20:online.

Faithfull, M. 2018. Innovation: do lockers create increased footfall? *Beyond Retail Industry* June 14:online.

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., N. Karacaoglu, and A. Moreno. 2018. Why retailers should care about net neutrality: The impact of website performance on online retail. Available at SSRN 3260203.

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. 2019. *Operations in an omnichannel world*. Springer.

Gao, F., and X. Su. 2017a. Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science* 63 (8): 2478–2492.

Gao, F., and X. Su. 2017b. Online and offline information for omnichannel retailing. *Manufacturing & Service Operations Management* 19 (1): 84–98.

Goic, M., and M. Olivares. 2019. Omnichannel Analytics. In *Operations in an Omnichannel World*, 115–150. Springer.

Harris 2013. Showrooming and Webrooming: A Tale of Two Trends. Technical report, Harris Interactive, Available: https://bit.ly/3AfNbqB. Last accessed Oct 7, 2021.

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.

Hübner, A., A. Holzapfel, H. Kuhn, and E. Obermair. 2019. Distribution in Omnichannel Grocery Retailing: An Analysis of Concepts Realized. In *Operations in an Omnichannel World*, 283–310. Springer.

Hui, S. K., P. S. Fader, and E. T. Bradlow. 2009a. Path data in marketing: An integrative framework and prospectus for model building. *Marketing Science* 28 (2): 320–335.

Hui, S. K., P. S. Fader, and E. T. Bradlow. 2009b. Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior. *Journal of consumer research* 36 (3): 478–493.

Jones, C. 2019. Can't wait for that delivery? Amazon, Rite Aid team up to make it easier to get packages. *USA Today* June 27:online.

Kumar, A., A. Mehra, and S. Kumar. 2019. Why do stores drive online sales? Evidence of underlying mechanisms from a multichannel retailer. *Information Systems Research* 30 (1): 319–338.

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

Mani, V., S. Kesavan, and J. M. Swaminathan. 2015. Estimating the impact of understaffing on sales and profitability in retail stores. *Production and Operations Management* 24 (2): 201–218.

Martínez-de Albéniz, V., and A. Belkaid. 2021. Here comes the sun: Fashion goods retailing under weather fluctuations. *European Journal of Operational Research* 294 (3): 820–830.

Martínez-de Albéniz, V., A. Planas, and S. Nasini. 2020. Using clickstream data to improve flash sales effectiveness. *Production and Operations Management* 29 (11): 2508–2531.

Montoya, R., and C. Gonzalez. 2019. A hidden Markov model to detect on-shelf out-of-stocks using point-of-sale data. *Manufacturing & Service Operations Management* 21 (4): 932–948.

Mowrey, C. H., P. J. Parikh, and K. R. Gue. 2018. A model to optimize rack layout in a retail store. *European Journal of Operational Research* 271 (3): 1100–1112.

Musalem, A., M. Olivares, and A. Schilkrut. 2021. Retail in high definition: Monitoring customer assistance through video analytics. *Manufacturing & Service Operations Management* 23 (5): 1025–1042.

Ozgormus, E., and A. E. Smith. 2020. A data-driven approach to grocery store block layout. *Computers & Industrial Engineering* 139:105562.

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

Ruiz, F. J., S. Athey, D. M. Blei et al. 2020. Shopper: A probabilistic model of consumer choice with substitutes and complements. *Annals of Applied Statistics* 14 (1): 1–27.

Schaverien, A. 2018. Five Reasons Why Amazon Is Moving Into Bricks-And-Mortar Retail. *Forbes* December 2018:online.

Underhill, P. 2009. *Why we buy: The science of shopping–updated and revised for the internet, the global consumer, and beyond*. Simon and Schuster.

Wagner, R. P., and T. Jeitschko. 2017. Why Amazon's 1-Click Ordering Was A Game Changer. Knowledge@ Wharton.

Wang, R., and O. Sahin. 2018. The impact of consumer search cost on assortment planning and pricing. *Management Science* 64 (8): 3649–3666.

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.

# Appendices

## Supporting Tables for Robustness

Table 7: Alternative fixed effects for Models (9) and (11) and interaction model

| | Dependent variable: $n_{ist}$ | | | | | | |
|---|---|---|---|---|---|---|---|
| | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| $v_{ist}$ | 1.0453*** | | 1.0479*** | | 1.0546*** | | 1.0021*** |
| | (0.0043) | | (0.0044) | | (0.0044) | | (0.0085) |
| $v_{1,ist}$ | | 0.7562*** | | 0.7423*** | | 0.7484*** | |
| | | (0.0074) | | (0.0075) | | (0.0074) | |
| $v_{2-4,ist}$ | | 0.2913*** | | 0.2973*** | | 0.3019*** | |
| | | (0.0100) | | (0.0102) | | (0.0101) | |
| $v_{>4,ist}$ | | 0.0680*** | | 0.0854*** | | 0.0860*** | |
| | | (0.0086) | | (0.0087) | | (0.0087) | |
| $D_{is}$ | | | | | | | −0.0024*** |
| | | | | | | | (0.0003) |
| $\mathbb{1}_{D_{is}\in[50,100]}$ | −0.3470*** | −0.3729*** | −0.2206*** | −0.2282*** | −0.3469*** | −0.3712*** | |
| | (0.0325) | (0.0323) | (0.0225) | (0.0224) | (0.1001) | (0.0996) | |
| $\mathbb{1}_{D_{is}\in[100,150]}$ | −0.5075*** | −0.5119*** | −0.2501*** | −0.2510*** | −0.5084*** | −0.5121*** | |
| | (0.0371) | (0.0369) | (0.0235) | (0.0234) | (0.0987) | (0.0981) | |
| $\mathbb{1}_{D_{is}\in[150,200]}$ | −0.6147*** | −0.6151*** | −0.3325*** | −0.3407*** | −0.6157*** | −0.6146*** | |
| | (0.0401) | (0.0399) | (0.0253) | (0.0252) | (0.0991) | (0.0985) | |
| $\mathbb{1}_{D_{is}\in[200,250]}$ | −0.7187*** | −0.7247*** | −0.4334*** | −0.4590*** | −0.7212*** | −0.7253*** | |
| | (0.0504) | (0.0501) | (0.0326) | (0.0325) | (0.1033) | (0.1027) | |
| $\mathbb{1}_{D_{is}\in[250,\infty]}$ | −0.7368*** | −0.7558*** | −0.4345*** | −0.4547*** | −0.7383*** | −0.7554*** | |
| | (0.0622) | (0.0619) | (0.0424) | (0.0422) | (0.1096) | (0.1090) | |
| $D_{is} : v_{ist}$ | | | | | | | 0.0003*** |
| | | | | | | | (0.0001) |
| Fixed effects | week product-store | week product-store | week-product store | week-product store | week-product store-product | week-product store-product | week product store |
| Observations | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 | 80,190 |
| $R^2$ | 0.9264 | 0.9273 | 0.9261 | 0.9267 | 0.9280 | 0.9288 | 0.9244 |
| Adjusted $R^2$ | 0.9261 | 0.9269 | 0.9255 | 0.9261 | 0.9272 | 0.9280 | 0.9243 |
| Residual Std. Error | 1.2198 | 1.2128 | 1.2250 | 1.2197 | 1.2110 | 1.2040 | 1.2342 |

*p<0.1; **p<0.05; ***p<0.01

Table 8: Models (6) through (11), subset Santiago Metropolitan area

| | *Dependent variable:* | | | | | |
|---|---|---|---|---|---|---|
| | $n_{ist}$ | | | | | |
| | (19) | (20) | (21) | (22) | (23) | (24) |
| $v_{ist}$ | | | $1.0358^{***}$ | $1.0366^{***}$ | | |
| | | | (0.0062) | (0.0062) | | |
| $v_{1,ist}$ | | | | | $0.7379^{***}$ | $0.7393^{***}$ |
| | | | | | (0.0104) | (0.0104) |
| $v_{2-4,ist}$ | | | | | $0.3134^{***}$ | $0.3119^{***}$ |
| | | | | | (0.0139) | (0.0139) |
| $v_{>4,ist}$ | | | | | $0.0463^{***}$ | $0.0480^{***}$ |
| | | | | | (0.0120) | (0.0120) |
| $D_{is}$ | $-0.0010^{***}$ | | $-0.0013^{***}$ | | $-0.0015^{***}$ | |
| | (0.0002) | | (0.0002) | | (0.0002) | |
| $\mathbb{1}_{D_{is}\in[50,100]}$ | | $-0.3004^{***}$ | | $-0.3057^{***}$ | | $-0.3290^{***}$ |
| | | (0.0452) | | (0.0355) | | (0.0356) |
| $\mathbb{1}_{D_{is}\in[100,150]}$ | | $-0.2845^{***}$ | | $-0.3150^{***}$ | | $-0.3328^{***}$ |
| | | (0.0470) | | (0.0369) | | (0.0370) |
| $\mathbb{1}_{D_{is}\in[150,200]}$ | | $-0.3454^{***}$ | | $-0.3781^{***}$ | | $-0.4062^{***}$ |
| | | (0.0483) | | (0.0380) | | (0.0380) |
| $\mathbb{1}_{D_{is}\in[200,250]}$ | | $-0.3667^{***}$ | | $-0.5474^{***}$ | | $-0.5894^{***}$ |
| | | (0.0584) | | (0.0459) | | (0.0460) |
| $\mathbb{1}_{D_{is}\in[250,\infty]}$ | | $-0.6345^{***}$ | | $-0.5257^{***}$ | | $-0.5422^{***}$ |
| | | (0.0839) | | (0.0660) | | (0.0661) |
| Fixed effects | week | week | week | week | week | week |
| | product | product | product | product | product | product |
| | store | store | store | store | store | store |
| Observations | 44,910 | 44,910 | 44,910 | 44,910 | 44,910 | 44,910 |
| $R^2$ | 0.8677 | 0.8678 | 0.9181 | 0.9182 | 0.9179 | 0.9180 |
| Adjusted $R^2$ | 0.8675 | 0.8677 | 0.9180 | 0.9181 | 0.9178 | 0.9179 |
| Residual Std. Error | 1.6527 | 1.6519 | 1.3004 | 1.2994 | 1.3022 | 1.3009 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

Table 9: TSLS for Models (8), (9), (10) and (11)

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | $n_{ist}$ | | | |
| | (25) | (26) | (27) | (28) |
| $v_{ist}$ | $1.1184^{***}$ | $1.1186^{***}$ | | |
| | (0.0047) | (0.0047) | | |
| $v_{1,ist}$ | | | $0.9518^{***}$ | $0.9554^{***}$ |
| | | | (0.0159) | (0.0158) |
| $v_{2-4,ist}$ | | | $0.2379^{***}$ | $0.2289^{***}$ |
| | | | (0.0415) | (0.0414) |
| $v_{>4,ist}$ | | | 0.0213 | 0.0285 |
| | | | (0.0370) | (0.0369) |
| $D_{is}$ | $-0.0012^{***}$ | | $-0.0014^{***}$ | |
| | (0.0001) | | (0.0001) | |
| $\mathbb{1}_{D_{is}\in[50,100]}$ | | $-0.2188^{***}$ | | $-0.2353^{***}$ |
| | | (0.0227) | | (0.0228) |
| $\mathbb{1}_{D_{is}\in[100,150]}$ | | $-0.2505^{***}$ | | $-0.2590^{***}$ |
| | | (0.0237) | | (0.0238) |
| $\mathbb{1}_{D_{is}\in[150,200]}$ | | $-0.3332^{***}$ | | $-0.3525^{***}$ |
| | | (0.0255) | | (0.0256) |
| $\mathbb{1}_{D_{is}\in[200,250]}$ | | $-0.4405^{***}$ | | $-0.4893^{***}$ |
| | | (0.0329) | | (0.0331) |
| $\mathbb{1}_{D_{is}\in[250,\infty]}$ | | $-0.4369^{***}$ | | $-0.4736^{***}$ |
| | | (0.0428) | | (0.0429) |
| Fixed effects | week | week | week | week |
| | product | product | product | product |
| | store | store | store | store |
| Observations | 80,190 | 80,190 | 80,190 | 80,190 |
| $R^2$ | 0.9241 | 0.9242 | 0.9238 | 0.9239 |
| Adjusted $R^2$ | 0.9240 | 0.9241 | 0.9237 | 0.9238 |
| Residual Std. Error | 1.2370 | 1.2361 | 1.2391 | 1.2382 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01