Submitted to *Manufacturing & Service Operations Management*
manuscript

Store Network Design for Omnichannel Retailing

Mert Çetin
IESE Business School, University of Navarra, mcetin@iese.edu

Victor Martínez-de-Albéniz
IESE Business School, University of Navarra, valbeniz@iese.edu

Laura Wagner
Universidade Católica Portuguesa, Lisbon School of Business and Economicslwagner@ucp.pt

Problem definition: A store contributes to an omnichannel retailer in different ways, through direct brick-and-mortar sales but also by influencing online purchases. Yet, the contribution of the different store network dimensions is not well understood. What is the role of proximity, abundance of stores, assortment breadth or inventory availability on a consumer's propensity to buy products, in all channels? **Methodology/results:** We use consumer-level activity over time to infer the influence of the store network on pure online and offline channels, as well as on a hybrid channel where orders are placed in the store but fulfilled online. This is done by geolocating consumers and relating their sales activity to the varying store network characteristics in their catchment area. We separate factors related to ease of access, from those that are related to service quality in the form of assortment variety and product availability. We find that better access increases sales, but service quality provided by the physical store network increases offline sales but decreases online and hybrid sales. **Managerial implications:** We develop a counterfactual analysis to demonstrate that omnichannel retailers should opt for dense store networks to sustain high sales levels, in contrast with the recent trend of store closures. Our work paves the way for a more effective design of physical distribution strategies for omnichannel retailers: our framework can simulate and optimize the effects of store network modifications.

Key words: omnichannel, retail, store network, purchase propensity, channel choice, catchment area, service level

Submitted: April 25, 2022.

1. Introduction

Ever since the first e-commerce transaction in 1994, the fraction of consumers shopping online has been increasing (Lufkin 2020). In recent years, this growth has accelerated and in particular the COVID-19 pandemic has forced many consumers to transfer their transactions from physical stores to online channels (Paton 2020). Consequently, almost all retail companies either have complemented their traditional store networks with an online channel, or have strengthened their online sales infrastructure to survive in this challenging era (Hwang et al. 2020, Caro et al. 2020).

Combining online and offline channels in a single go-to-market strategy has been dubbed as omnichannel. The concept was developed about a decade ago and has been studied extensively by now (Gallino and Moreno 2019). Even before COVID, large retailers were already allowing consumers to move across channels for transactions, e.g., purchase a product online and pick it up at the store, or return a product bought online at physical stores (Ofek et al. 2011, Gallino and Moreno 2014, Gallino et al. 2017). Cross-channel activities make decisions about inventory, assortment, fulfillment, and consumer experience more challenging. Beyond the higher operational complexity, the increase in online penetration is triggering a rethinking of distribution strategies in the industry (Adhi et al. 2020). As a response to the migration to online channels, some of the largest fashion retailers are closing stores: Inditex and H&M announced the closure of 1,200 and 250 stores respectively as the pandemic drives sales online (Paton 2020, BBC 2020). Moreover, Inditex also plans to tailor its marketing strategy and daily operations to better fit online shoppers (The Economist 2021). In particular, there are opportunities to provide more choice to the consumer by offering online consumers fulfillment from store (Martínez-de Albéniz 2019). These innovations also open up a chance for operational improvements due to more flexibility in the matching between inventory units – available from more locations – and online orders (Andrews et al. 2019). Furthermore, the cross-channel flexibility allows consumers to 'showroom', where they can physically experience the products before buying online (Bell et al. 2017, 2020), as well as to 'webroom', where consumers can access online information on products before buying at the store (Gallino and Moreno 2014, Caro et al. 2021).

The developments above suggest that store networks need to take into consideration consumer cross-channel behaviors. In particular, a store performance should not be measured by revenues alone, but also by its contribution to the overall impact on all consumers, including those who did not shop in that store or those who shopped in the store but were previously shopping in another store or channel (Hearne et al. 2019). In other words, a store could bring in new consumers, capture existing consumers from other channels, or 'supercharge' new or existing consumers that end up spending in other channels (Bell et al. 2020). These are direct effects, with or without cannibalization, and indirect effects. These interactions are unfortunately not well understood. Specifically, some literature has studied the effect of store openings on multiple channels (Wang and Goldfarb 2017, Kumar et al. 2019), but we still do not know what elements of the store matter: what is the role of proximity, abundance of stores, assortment breadth or inventory availability? These elements all influence the experience of a visiting consumer and hence should impact its future purchase intentions across all channels, in the same way marketing communication shapes future purchases (Wiesel et al. 2011).

These insights are key to inform any store network reorganization. Indeed, if the value of a store comes from a billboard effect in which consumers get to see the brand, then it would be advisable to treat stores as advertising devices, and hence small permanent stores and pop-up stores should be the preferred option. In contrast, if the value comes from the ability to try products, then larger stores with broad assortments and high service levels should be the most effective way to engage with consumers. As a result, we can only determine optimal store networks for omnichannel retailers by breaking down the channel specific impact of stores on consumers into its functional components.

In this paper, we study the interaction of a given consumer (her) with a retailer, across offline and online channels. The consumer's surrounding conditions vary over time. That is, the number of stores in proximity may change as well as the distance to the closest store, the assortment available and the service level may fluctuate. As a result of these variations, the consumer may modify her purchase decisions. We estimate these effects with transaction data from a major shoe retailer, serving close to 170,000 consumers geolocated in space, and generating nearly 270,000 transactions over two years. The data allows us to estimate the sensitivity of the consumers to (1) ease of access, via the number of stores nearby and the distance to the closest one; (2) product offer, via the number of available products in nearby stores, for the consumer's gender; and (3) service level, via nearby in-store product availability, for the consumer's gender and size. Our validation thus entails a granular approach where we take advantage of the variation of conditions across geographies, time, and consumers, because, even when they are located in the same place in the same period, their experience may change for different genders and sizes.

We find that, consistent with previous research on the spillover effects of stores, online consumers exposed to physical stores are more likely to shop. However, online interactions are reduced when the store network around them provides better assortment breadth and higher service level, which means that the online channel is best activated with physical stores with low variety and low product availability. The same is true for the hybrid channel in which orders are placed in store but fulfilled online. In contrast, for the offline channel, more stores do not generally translate into higher sales, but the effect of distance is very strong: when a consumer is closer to a store she is more likely to purchase offline. There, a wider assortment and better service quality increase the offline purchase probability of the consumer. In other words, opening a new store enhances the offline purchases only if it results in better service quality and/or bringing stores closer to consumers.

Once we run our regression models, we present a counterfactual analysis where we measure the expected impact on purchases across different channels, when we change the store network. We find that store openings remain the most effective lever to increase sales, in comparison to increasing variety or service levels. Our model thus provides a useful tool for managers to help them predict the trade-off between opening a store and improving the service quality of the current stores specific to a geographical location.

To our knowledge, this is the first study to analyze the impact of store network characteristics on multi-channel individual consumer purchases. While our empirical context is limited to a specific product category (fashionable shoes), the methodology and the direction of the cross-channel influences seem generally applicable, even though their quantitative values may change. As a result, our work paves the way for a more effective design of physical distribution strategies for omnichannel retailers.

The rest of the paper is organized as follows. Section 2 provides a review of the relevant literature. The institutional setup and data are described in §3. The identification strategy and the main model are provided in §4, followed by the main results in §5 and the robustness tests in §6. After presenting the counterfactual analysis in §7, we conclude in §8. Additional methodological details and results are included in the Appendix.

2. Literature Review

Our work is related to three main areas of research. First, it is connected to the marketing and operations management literature on channel choice and store visit decisions. Second, it falls within a nascent stream of work around omnichannel, that combines empirical documentation of consumer behaviors, and decision models to best utilize omnichannel capabilities. Third, our prescriptions are based on the classical Operations Research frameworks for facility location.

2.1. Channel and store visit decisions

Exploiting past consumer purchase data to forecast future consumer behavior has been an active area of research in marketing, see e.g., Geyskens et al. (2002), Valentini et al. (2011), Avery et al. (2012) or Chintagunta et al. (2012). The central questions in this literature focus on understanding consumer decisions towards the point of sales, and how these are affected by firm actions. In particular, store choice is of primary interest. Early papers usually focus on the *which* store consumer visits and how this visit affects the purchase behavior of the consumer (Bell et al. 1998). With the emergence of online shopping, however, research tilted toward *whether* consumer visits a store, that is, to inform about the consumer channel migration from catalogue to online (Ansari et al. 2008). Moreover, earlier research showed that online channel investments are, on average, positive net-present-value investments considering the anticipated consumer behavior shift (Geyskens et al. 2002). Chintagunta et al. (2012) explore this channel migration by incorporating the transaction costs involved in grocery shopping. They further integrate those costs into a channel choice framework and offer a new strategy for the retailers to target consumers in both channels. All of those studies focus on how consumer behavior changes in a multichannel environment. One study that focuses on store network as we do is Avery et al. (2012), where authors empirically test cross-channel elasticities for a multichannel retailer using store openings and they find short-run cross-channel cannibalization along with long-run increase in both channels.

A separate literature shows how assortment affects consumers' visit and purchase behaviors. Bernstein and Martínez-de-Albéniz (2017) find that consumer visits are triggered by assortment rotation and suggest an optimal product-rotation policy. Ferreira and Goh (2021) show that not disclosing future product introductions may push consumers to purchase more. Kök and Simsek (2021) show that higher variety is associated with higher sales. These studies highlight that variety is a lever to engage with consumers, which leads us to consider assortment breadth as an important characteristic of the physical stores that a consumer can access.

Similarly, inventory availability is a critical factor to convert potential demand into realized sales. Campo et al. (2000, 2004), Musalem et al. (2010) and Boada-Collado and Martínez-de Albéniz (2020) show that higher inventory availability – in particular to avoid stock-outs – increase sales. We thus include availability as another metric that describes the quality of the physical stores in the consumer vicinity.

Finally, Lemon and Verhoef (2016) underline the importance of understanding the consumer journey as a whole and exploring all its aspects, especially in the presence of complex decision making. This is particulary relevant to omnichannel retailing, so we aim to include all possible metrics to evaluate consumers' decisions.

2.2. Omnichannel

Offline-online interactions have been the focus of omnichannel research. Interactions go in both ways: online actions affect offline behaviors, and vice versa. Specifically, online actions by the firm affect consumer channel choice and offline behaviors: Gallino and Moreno (2014) show that 'buy-online, pick-up-in-store' implementation unexpectedly results in a decrease of online sales due to additional sales driven by the pick-up visit and the channel-shift effect. Gao and Su (2017a) explains the same phenomenon by the channel-shift effect, yet they also point out that this can help retailer acquiring new consumers. The effect also impacts competitors: Zhang et al. (2019) shows a similar spillover effect of pop-up stores on other sales channels. Spillovers from offline into online behaviors have also been established. Bell et al. (2015) and Bell et al. (2017) show that visiting a showroom where products are displayed and experienced increase total demand, including in online channels. Bell et al. (2020) find that, after visiting a showroom, consumers spend more, shop at a higher velocity, and are less likely to return items. Gao and Su (2017b) evaluate physical showrooms, virtual showrooms, and availability information as different mechanisms to provide more information to consumers.

Additionally, there is a stream of research relating store access and sales in an omnichannel system. There are two main concepts explaining why proximity may drive more sales: convenience and exposure. When proximity provides more convenience to online consumers via faster delivery, Fisher et al. (2019) show that both online and offline sales increase, and both channels complement each other. Kumar et al. (2019) test how opening of new stores effect both online and offline sales. Their results identify a convenience effect due to a reduction in the risks of online purchase, given the in-store return option; and an exposure effect from higher engagement after store interactions, in line with Bell et al. (2020). Wang and Goldfarb (2017) further study the effect of opening a store both in places where the retailer has a strong presence and where not. They suggest that online and offline channels are substitutes in distribution, but complements in marketing communications. Ofek et al. (2011) argue that the introduction of online channel induces higher offline channel costs from the extra work associated with in-store returns, yet it might be beneficial because store assistance makes consumers try the products and ultimately reduce the likelihood of returns. Gao et al. (2022) list showrooming, return, and fulfillment as the different functions of physical stores. They show that the more functions given to a store, the necessary number and size of stores is reduced.

The impact of variety on omnichannel behavior has also been studied. Rooderkerk and Kök (2019) explore how omnichannel retailers coordinate their assortment decisions across multiple channels and further discuss the strategic, tactical and operational challenges of this emerging omnichannel assortment planning problem. Brynjolfsson et al. (2009) test whether there is cross-channel competition within the same company. Their analysis indicates that competition is stronger for mainstream products, whereas niche products sold online are not subject to competition from the brick-to-mortar retailers. Gallino et al. (2017) study the impact of increased variety, obtained from the ability to receive products that are unavailable at the store. They show that sales dispersion increases. In contrast with these studies, we look at the impact of physical store variety on all channels. More generally, we provide a consumer-level analysis that decomposes the different dimensions of the store network.

2.3. Facility location and management

Facility location has been one of the fundamental research topics of operations research for decades (Hekmatfar and Farahani 2009). Drezner and Hamacher (2001) provide an outlook of facility location related problems such as set covering problem, hierarchical location, hub location, etc. Although this is a well established field, there are only few studies focusing on facility location aspects of omnichannel retailing. Liu et al. (2010) formulate a capacitated location model that in a multi-channel supply chain and solve synthetic two-echelon instances. Millstein and Campbell (2018) study reverse cannibalization between online and offline channels of a retailer, and develop an general multiple-facility location model to optimize the location of an omnichannel warehouse. However, they only consider the optimal location of warehouse – not the stores. In addition, Nishida (2015) analyzes the data from a multistore retailing firm to investigate location choice and presents an equilibrium in store networks regarding the business-stealing effect and the cost-saving effect. Furthermore, Aguirregabiria and Vicentini (2016) study a similar problem considering the effects of competition between multi-store retailers. Our approach is similar, except that we consider the entire store network redesign, incorporating cross-channel effects.

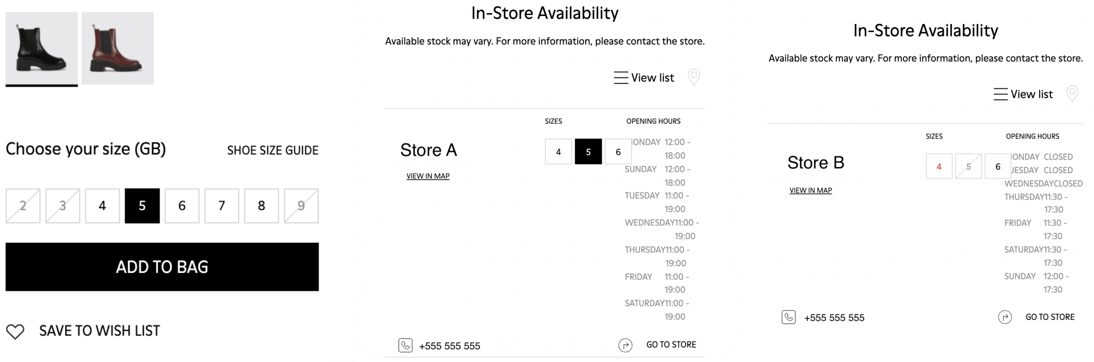
Finally, there is literature that takes the store network as given, and manages operational decisions, although without cross-channel considerations. For instance, Federgruen and Zipkin (1984) consider inventory distribution assuming independent demands across stores, and Burns et al. (1985), Iyer and Schrage (1992), Martin et al. (1993), Chan and Simchi-Levi (1998), Fumero and Vercellis (1999), Godfrey and Powell (2001), Iyer et al. (2007), Gürbüz et al. (2007), among others, provide further studies on the distribution questions. Caro and Gallien (2010) investigate inventory distribution in fashion apparel, and implements the theoretical prescriptions at Zara, showing that sales increase significantly; Gallien et al. (2015) consider the same problem with a focus on initial shipment decisions.

3. Institutional Setup and Data

We partnered with a major omnichannel shoe retailer to study consumers' sensitivity to the retailer's store network characteristics. The retailer's network consists of many different store types, including traditional brick-and-mortar stores, outlet stores, its own online store, and external marketplaces. Consumers interested in shopping with them can find products either *online* or in *stores*.

When a consumer browses the website, she is offered the entire available assortment, thanks to the retailer's fully connected physical network, so online orders can be fulfilled by any of the retailer's physical stores as well as by a central warehouse. She, additionally, has the option to check whether her favourite pair of shoes is available in a nearby store. Figure 1 illustrates these points. In this instance, a consumer interested in buying a pair of shoes can do so online if her shoe size is between 4 8, and is out of luck otherwise. Checking the nearby stores in London area, reveals that Store A stocks these shoes, while Store B is out of stock for size 5. While the online information on the assortment and service level increases visibility, it does not allow a consumer to immediately reserve a pair of shoe in store. The retailer additionally integrated another common omnichannel experience: the possibility to return products in-store when purchased online. Our research question aims to address how online consumers react to changes to service provided by the store network. Thus, online consumers' awareness of these parameters can be ensured through this online system to check physical stores.

Figure 1: Check in-store feature.



If on the other hand a consumer decides to visit a brick-and-mortar store, she has to first search through the assortment available in that particular store and then, depending on the store design, either ask the store assistants for her desired size, or pick them out herself from the shelves. If a product she desires is out of stock, she also has the option of placing an ***online order*** via the store manager's tablet, essentially giving her access to the entire network's inventory. This hybrid channel resembles the pure offline shopping experience, with the difference that fulfillment is carried out online, directly to the consumer's delivery address.

In sum, our partner retailer has established a true omnichannel experience where online browsers have visibility of the service provided by the store network, and offline consumers benefit from the online offerings via the tablet options.

Our aim is to study how consumer's propensity to buy products via these three channels (online, traditional brick and mortar, and tablet) changes with her exposure to a retailer's changes in their brick-and mortar store network composition and related service quality. We focus on the retailer's operations in Europe from January 2018 to February 2020 (previous to the COVID pandemic). The raw data comprises of three databases concerning individual transaction data, store data and inventory data, which we fused to obtain individual consumer exposure to retailer's physical presence.

Transaction Data. The company records purchase transactions for each channel separately, along with encrypted consumer identifiers, time stamps, home addresses (if available), sign up date, gender and purchased product information (product description, size, product and basket value) which allows us to track individual purchases across channels.

Store Data. We obtain the characteristics of all the physical stores from 12 countries, amounting to 98 stores in total. During the observed time period, the retailer opened 12 brick-and-mortar stores and closed 14. Each of these stores is associated with a (i) unique address and a (ii) store type (brick-and mortar store, outlet store).

Inventory data. The retailer records all products available in each store at any given week, including information about shoe sizes.

Constructed data and data cleaning. Based on the information available in the raw data set, we constructed and imputed missing information for several variables. First, home addresses of every individual consumer were geolocated (latitude and longitudes); we were able to identify 95.4% of all *available addresses*. Second, we imputed missing gender information for both, consumers and products using product description, consumer gender, product size range information, and purchase frequency. The exact approach is detailed in Appendix A. Finally, we labeled consumers as new or existing ones based on their sign up date.

Sample Selection. The raw data set contains 1,557,858 transactions. To obtain our final sample we included *all* consumer transactions which met the following criteria:

1. The consumer must be identifiable through a unique identification with exactly one location.
2. The consumer must have a (European) shoe size between 36 and 47, so as to focus primarily on adults.
3. Transactions are made either via the retailer's own website, in a traditional brick and mortar store or via a store manager's tablet.

Criterion 1 ensures that we are able to capture consumers' reactions to local store network changes. This principle keeps 33.23 % of all transactions (517,732). Note that, while online transactions are naturally tied to both unique consumer identifiers and consumer addresses, physical transactions may lack this information. We must be careful though, this non-random identification of consumers, could bias subsequent analysis. We address such concerns with an aggregated analysis in Appendix B that is fully consistent with the main individual-level analysis. The rationale behind Criterion 2 is that children are unlikely to make purchase decisions on their own. This filter retrieves 81.73% of all transactions surviving Criterion 1 (423,136). Criterion 3 removes transactions in Marketplaces and Outlet stores, which are mostly driven by promotions and discounts, and have different dynamics than full-price channels. For completeness, however, we assure ourselves of this intuition via a robustness check in §6. The final transaction data set thus includes 268,573 distinct purchases placed by 169,894 unique consumers purchasing 7,333 distinct products.

The descriptive statistics are provided in Table 1. We additionally find that the shoe retailer attracts women and men equally (51% versus 49% of total purchases). Moreover, the percentage of identified consumers who have purchased in store and tablet channels is rather low, topping 4.8% (Store & Online 3.1%, Store & Tablet 0.7%, Online & Tablet 0.6%, 0.4% use all channels).

4. Identification Strategy

The service that retailers provide changes over time and hence may serve different populations in a heterogeneous way. For instance, not only stores may open or close, but also the service provided

Table 1: Descriptive statistics.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Purchase frequency per consumer	169,894	1.623	1.235	1	1	2	19
Average basket size	169,894	1.273	0.659	1	1	1	16
Average basket value	169,894	89.913	29.359	3.086	68.070	108.400	420.170
Total spending	169,894	112.646	65.322	4.130	73.950	128.100	1,941.579

by the network of stores varies. These variations may affect *all* consumers in a region (such as store openings or variations in assortment), or only *some* (such as variations in a particular shoe size). How these variations affect consumers propensity to purchase is at the core of this research. In other words, we study whether and how consumers react to variations in retailer's brick-and-mortar network composition and service provided through it. Our identification strategy is based on a weekly individual consumer panel data approach with two way (time and consumer) fixed effects. To be more precise, consider a consumer i that lives in a certain location. Over multiple periods $t = 1, \ldots, T$, she is exposed to the physical presence of a retailer *in her catchment area*. To measure this area for consumer i, we draw a circle around her with a radius of 5 to 10 km, and we document with X_{it} the exposure characteristics of consumer i in week t to a retailer's changes in physical activities. These characteristics can include any explanatory variable provided it is consumer-time dependent. In this research context, we are interested in how retailer's changes in (i) the physical and (ii) operational aspects of the brick-and-mortar store network affect consumer's propensity to buy via particular channels (*online*, *store* and *tablet*) over time. To determine the various components of the vector $X_{i,t}$ we analyze the following dynamic variables:

Physical aspects of the store network:

— *Distance.* This variable measures the straight-line distance between the consumer's address and the closest brick-to-mortar store, in log scale.

— *Store accessible.* This binary variable measures whether at least one brick-and-mortar store is within a consumers' catchment area when buying online.

— *Number of stores.* This variable measures the log of numbers of stores reachable within the catchment area.

Operational aspects of the store network:

— *Assortment breadth.* The average number of distinct model-color products available in the local network for a consumer's gender per store, in log scale.

— *Service level.* Measures the proportion of assortment breadth available in a consumer's size *and* gender, in log scale.

Promotion controls:

—*Discount.* It is crucial to control for promotion patterns, since consumer behavior changes drastically when goods are discounted (Haghighatnia et al. 2018). In an ideal world, we would control for the fraction of products being offered on discount in any given week to ensure that our results are not driven by consumers' price sensitivity. Unfortunately, we do not have direct access to this variable, but not including it could lead to unobserved variable bias. Faced with this challenge, we approximate the fraction of discounted products via our sales records. In the presence of both list prices and actual prices, one could simply take the country specific list price for each offered product, compare it with the actual selling price and count the fraction of products offered below. Since we do not have access to price data for all *o ered products*, we approximate them via data obtained from *sold products*. We thus proceed as follows: to infer the list price for each product, we take the median price per country over the season (list prices typically vary by country, but not by stores or channels within a country). Then for every week, country and channel, we count the fraction of products sold at a discount. To sum up, the fraction of discounted products is a proxy for promotional activity, and varies by channel, week and country.

Our final data set has a weekly panel structure, which is shown for a sample trajectory of a single customer in Table 2.

Table 2: Panel Data with Covariates

			Store Network			Operational Characteristics		Promotion Control
ID	Week-Year	Purchase	Distance	Store accessible	Number of stores	Assortment breath	Service level	Discount
1	24 2019	0	4.17	TRUE	6	123.86	0.69	0.06
1	25 2019	0	4.17	TRUE	6	123.29	0.68	0.04
1	26 2019	0	4.17	TRUE	6	123.14	0.66	0.03
1	27 2019	0	4.17	TRUE	6	124.29	0.66	0.20
1	28 2019	0	3.04	TRUE	7	129.57	0.61	0.33
1	29 2019	0	3.04	TRUE	7	119.00	0.58	0.32
1	30 2019	0	3.04	TRUE	7	121.75	0.54	0.28
1	31 2019	0	3.04	TRUE	7	118.50	0.52	0.46
1	32 2019	1	3.04	TRUE	7	123.00	0.56	0.45
1	33 2019	0	3.04	TRUE	7	129.75	0.56	0.53
1	34 2019	0	3.04	TRUE	7	138.38	0.59	0.43
1	35 2019	0	3.04	TRUE	7	140.12	0.59	0.28

Additionally, we provide descriptive statistics of the main variable of interest in Table 3. To ensure that the catchment area of 5-10 km captures most retailer-consumer physical interactions, we illustrate in Figure 2 the purchase frequency per distance. As one can observe, 36% 48% of all online purchases and 72% 80% of all in-store purchases (*store* and *tablet*) are made by consumers living within 5km and 10km, respectively.

Finally, we provide the time-varying assortment breadth and service level variables in a region for a a given consumer in Figure 3.

Table 3: Descriptive statistics of final data.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Purchase	18,211,380	0.014	0.119	0	0	0	1
Network composition							
Distance	18,211,380	64.937	87.201	0.00001	3.596	100.920	991.314
Store accessible	18,211,380	0.299	0.458	0	0	1	1
Number of stores	18,211,380	0.823	1.740	0	0	1	8
Operational aspect							
Assortment breadth	18,211,380	25.949	43.091	0	0	55	226
Service level	18,211,380	0.214	0.342	0	0	0.6	1
Promotion control							
Discount	18,211,380	0.20	0.19	0.00	0.04	0.32	1

Figure 2: Purchase frequency as a function of *Distance* to the closest store.

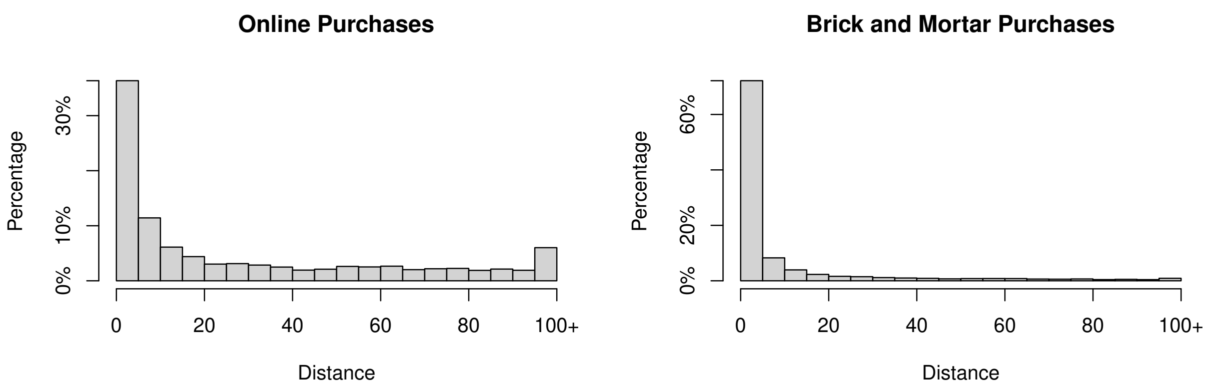
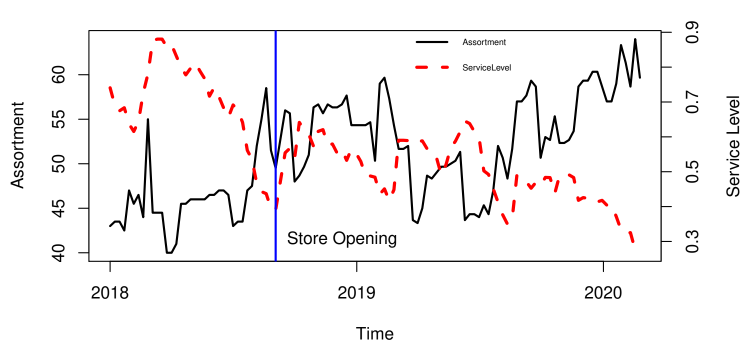


Figure 3: Changes in covariates for an individual consumer's catchment area.



Our main model focuses on three different channels, $c\ 2\ fonline; store; tabletg$. In week t, consumer i will thus obtain, in channel c, a utility

$$U_{ict} := {}_{ic} + {}^{0}_{tc} + {}_{c}X_{it} + {}''_{itc}$$

where ${}_{ic}$ and ${}^{0}_{tc}$ are fixed effects, X_{it} are the covariates of interest which include store network features, and ${}''_{ict}$ is a Gumbel-distributed shock.

As a result, the probability of purchase can be written as a logistic function:

$$\begin{aligned} \mathrm{P}(\mathrm{Purchase}_{itc}) &= LOGIT({}_{ic} + {}^{0}_{tc} + {}_{c}X_{it}) \\ &= LOGIT({}_{ic} + {}^{0}_{tc} \\ &\quad + \underbrace{{}_{1}\mathrm{Distance}_{it} + {}_{2}\mathrm{Store\ accessible}_{it} + {}_{3}\mathrm{Number\ of\ stores}_{it}}_{\text{Network composition}} \\ &\quad + \underbrace{{}_{4}\mathrm{Assortment\ breadth}_{it} + {}_{5}\mathrm{Service\ level}_{it}}_{\text{Operational aspect of the network}} + \underbrace{{}_{6}\mathrm{Discount}_{it}}_{\text{Promotion control}}) \end{aligned} \tag{1}$$

To obtain our estimates, we use the binary variable that indicates whether a consumer i in week t purchased a product in channel c, and run the logistic regression specified by Equation (1).

This formulation thus provides three separate equations for online, store and tablet purchases, each of which is estimated separately. Recall that logit models remove any observations whose value remains constant, and as such, we do not need to be concerned with gathering data from consumers who do not change behaviour across the entire observation period. Finally, to account for possible cross-channel effects, we also analyze a Multinomial logit model in §6, which results in qualitatively similar findings.

5. Main Results

In this section we explore how consumers' react to retailer's variation in network composition and service provided by the network. We begin by exploring consumers' sensitivity to physical changes in her catchment area when considering the online channel.

5.1. Impact on consumers' propensity to purchase online

The results of our estimation is given in Table 4. Column 1 of that table presents the estimates with a catchment area of 5km. We see that consumers once exposed to at least one physical store are on average more likely to shop online compared to not having access to the physical network. Additionally, the more stores a consumer can reach in her catchment area the higher her propensity to shop online. These two results are consistent with previous research (Bell et al. 2015, 2017, Fisher et al. 2019, Zhang et al. 2019), attributing the increase in cross-channel activity to the stores' billboard effect. One may wonder how far such a store billboard effect applies. In column 2, we illustrate the estimates for a wider catchment area of up to 10km and find that neither store accessibility nor the number of stores in an area can explain a consumer's purchase propensity (non-significant estimates). It thus seems that the billboard effect of brick-and-mortar stores is limited to consumers living in a very short proximity to stores (5km). But how does the assortment breadth and service level provided within the stores alter a consumer's propensity to purchase?

Interestingly, we find that the more products are available to a consumer offline and the better the service level of nearby stores, the less likely she is to purchase products online (Columns 1 and 2). Consistently, Column 3 shows that the further a consumer is away from a store the less sensitive she is to variations in the operational aspects of stores. While this seems contradictory to the billboard effect (which should be more intense when the stores have more to offer), the results are entirely consistent with predictions from consumer choice theory: according to this theory, consumers assign a utility to each channel and buy from the channel that provides the highest utility (including the outside option). The online channel's main drawback, compared to brick and mortar store, is that the product is not available immediately and may not have been tried on (so there may be uncertainty about its fit, see Gallino and Moreno 2018). In contrast, the utility from buying a product in a brick-and-mortar store directly increases with the assortment breadth and service level offered. Thus a better assortment and service level of the store network in a consumer's catchment area comparatively improves the value of the offline channel and reduces a consumer's willingness to buy online. Moreover, we find that *online* discounts have a marginally smaller positive effect on willingness to purchase when consumer is located further away from a store (Column 3), although discount enhances the willingness to buy in general (Columns 1, 2, and 3).

The implication of these findings is important: to activate online channels, it is important to have stores in close proximity of the consumer (less than 5km), but these stores do not need to be full-service ones. Small, showroom-type locations with limited variety and availability are more effective to increase online sales – even though these formats are less effective to generate offline sales as we will see next.

5.2. Impact on consumers' Brick and Mortar Purchases (Store or Tablet)

We now explore the effect of store network variations on consumers' propensity to purchase either in store or placing the order via a tablet. To do so, we concentrate on purchases within the catchment area and remove purchases further away, since these consumers are unlikely to have permanent access to the store network. Columns 1-2 in Table 5 present the estimates for transactions made in store, while Columns 3-4 show the coefficients for transactions made via tablet. We begin with analyzing the effect on regular store transactions and find that the larger the distance to the nearest store, the lower the likelihood of purchasing. Having more stores within a catchment area does not

Table 4: Impact on propensity to purchase online.

Catchment area	P(Online) 5km	P(Online) 10km	P(Online) 10km
Distance	0:02	0:04	0:02
	(0:02)	(0:02)	(0:02)
Store accessible	0:73	0:15	0:52
	(0:24)	(0:12)	(0:17)
Number of stores	1:38	0:14	0:14
	(0:18)	(0:09)	(0:10)
Assortment breadth	0:26	0:02	0:18
	(0:04)	(0:02)	(0:03)
Service level	1:44	1:11	0:63
	(0:10)	(0:08)	(0:11)
Discount	3:50	3:49	3:99
	(0:03)	(0:03)	(0:04)
Store accessible Distance			0:27
			(0:07)
Number of stores Distance			0:12
			(0:05)
Assortment breadth Distance			0:05
			(0:01)
Service level Distance			0:03
			(0:05)
Discount Distance			0:15
			(0:01)
Consumer FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Pseudo R^2	0:08	0:08	0:08
Observations	15406380	15406380	15406380
AIC	2310807:00	2310785:45	2310311:87
BIC	4350442:88	4350343:57	4349942:74
Log Likelihood	1015225:50	1015219:73	1014977:93

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

necessarily alter a consumer's propensity to purchase (insignificant). Thus, as long as a consumer has access to a store in her catchment area (whether 5 or 10 km), opening alternative stores will not change her propensity to purchase in-store, unless it alters the distance to the nearest store. We also find that the larger the assortment breadth and the better the service level to a consumer, the more likely she is inclined to place a transaction in store, irrespective of the catchment area (5 or 10 km). This is in line with the previous discussion for online sales. Note that the impact of service level is very large: when the service level improves by 10% (e.g., from 50 to 55%), sales are increased by about 20% ($e^{2:03 \quad \ln(1:1)} = 1:21$). Given that not all SKUs are stocked in store – the average service level in the sample is 0.214, see Table 3–, this seems a promising direction of improvement for the retailer.

Transactions via the tablet are impacted differently (Columns 3-4). The shorter the distance of a consumer to a store, the more likely she is placing an order via the tablet. Similarly, as the number of stores in her catchment area increases, the higher her willingness to use this channel. Thus changes in the physical aspects of the store network around a consumer will alter a consumer's willingness to experiment with this alternative channel. Recall that a transaction made via the tablet triggers an order online. As such, the consumer using this tablet channel may experience uncertainty with the purchased product. One possible explanation why she is more likely to try the channel when she has more stores around her is her convenience to return it in any of the stores near her. Finally, we observe that higher service levels reduce her probability to purchase via this channel. Again the impact of service level is extremely high, which is consistent with consumers using the tablet channel when products are unavailable in the store. Lastly, we find that discounts effect the purchase behavior positively as expected, yet by a smaller magnitude for tablet channel comparing to store channel. This can be easily explained by the nature of the tablet channel being triggered only for stock-outs, *after* consumer makes the purchase decision, hence she should be less price-sensitive.

Table 5: Impact on propensity to purchase in brick and mortar store.

Catchment area	P(Store) 5km	P(Store) 10km	P(Tablet) 5km	P(Tablet) 10km
Distance	0:22	0:28	0:84	0:80
	(0:09)	(0:08)	(0:11)	(0:11)
Number of stores	0:14	0:08	2:38	0:90
	(0:35)	(0:23)	(0:51)	(0:32)
Assortment breadth	0:39	0:38	0:12	0:12
	(0:08)	(0:06)	(0:08)	(0:06)
Service level	2:03	2:10	2:19	1:72
	(0:23)	(0:20)	(0:27)	(0:23)
Discount	2:25	2:10	1:01	1:04
	(0:14)	(0:12)	(0:13)	(0:12)
Consumer FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Pseudo R^2	0:06	0:06	0:05	0:05
Observations	1268136	1633583	851180	1041260
AIC	179589:30	228909:87	111319:43	135413:80
BIC	322466:26	414734:79	202829:61	248993:72
Log Likelihood	77940:65	99354:93	47807:71	58126:90

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

6. Robustness Checks

We conduct various robustness checks to address some concerns about our results. For these checks, we focus on a catchment area of 5km.

6.1. Endogeneity Concerns

One possible challenge is that our independent variables might be endogenous. Specifically, it is plausible that the retailer did not choose the location of a new store randomly, but rather projected potential sales for various locations and chose the best one. Similarly, the service level a retailer offered may have been set in anticipation of a surge in demand. Such situations could bias our estimation results and would make our reported estimations non causal. To overcome the aforementioned concern, we conduct an instrumental variables (IV) analysis with the aim of ensuring robustness and causality. We follow the common procedure of *a two step control function approach* to tackle such situations (Semykina and Wooldridge 2010, Tan and Netessine 2014, Chuang et al.

2016, Boada-Collado and Martínez-de Albéniz 2020). First we regress our potentially endogenous vector X_{it} as a function of the instrument Z_{it}, controlling for time and consumer fixed effects ($_i$, $_t$):

$$X_{it} = \ _i + \ _t + \ Z_{it} + \ _{it}. \tag{2}$$

Then, we predict values for the endogenous variables and repeat the estimation of Equation (1) with those predicted values from Equation (2), which can be rewritten as follows:

$$P(Purchase_{itc}) = LOGIT(\ _{ic} + \ _{tc} + \ _1\widehat{\text{Distance}}_{it} + \ _2\widehat{\text{Store accessible}}_{it} + \ _3\widehat{\text{Number of stores}}_{it} \tag{3}$$

$$+ \ _4\widehat{\text{Assortment breadth}}_{it} + \ _5\widehat{\text{Service level}}_{it} + \ _6\widehat{\text{Discount}}_{it} + "_{itc}) \tag{4}$$

A requirement for choosing instrumental variables is that each IV should be correlated with the corresponding endogenous explanatory variable, but uncorrelated with the error term $_{i;t}$.

To implement this in our context, we need six instruments, one for each independent variable in the vector X_{it} (distance, store accessible, number of stores, assortment breadth, service level, and the discount). We choose the following instruments:

1. Distance$_{it}$: As an instrument, we use the lagged Distance$_{i;t\ \ 1}$. To justify the instrumental variable choice, we reason that the minimum distance at time $t \ \ 1$ should not affect the purchase probability of a consumer at time t, because at that time the store network observed in the past should not affect her present choices. However, the minimum distance at time $t \ \ 1$ is strongly related to the minimum distance at time t, since it is a direct measure of the consumer's location – which is constant throughout our time horizon. This approach is quite common in the literature – see Boada-Collado and Martínez-de Albéniz (2020).
2. Store accessible$_{it}$ and Number of stores$_{it}$: As an instrument, we use the total inventory in a region (subnational divisions), divided the assortment breadth. We reason that since inventory increases with the number of stores, but assortment breadth usually does not, it should be highly correlated with the endogenous variables. At the same time, total inventory and assortment breadth in a region are invisible to the consumer (who only sees her catchment area), so it should not affect her choices.
3. Assortment breadth$_{it}$: Similar to the distance measure, we use the lagged Assortment breadth$_{i;t\ \ 1}$. To justify its use, it seems clear that the assortment at time $t \ \ 1$ does not directly affect the purchase choice at time t, because the consumer chooses based on the options available to her now.
4. Service level$_{it}$: Similarly, we use the lagged of the endogenous variable to create our instrument Service level$_{i;t\ \ 1}$, with the same logic as before.

5. Discount_{it}: Again, we use the lagged of the endogenous variable to create our instrument $\text{Discount}_{i;t\ \ 1}$, with the same logic as before.

We depict the results of the second stage regressions in Table 6. As we can see, the signs and significance of our estimates remain robust, although the magnitude of the coefficients may slightly vary.

Table 6: Impact on propensity to purchase in respective channels - IV second stage results.

	P(Online Purchase)	P(Store Purchase)	P(Tablet Purchase)
Distance	0:02	3:77	10:18
	(0:01)	(1:28)	(4:09)
Store accessible	0:10		
	(0:02)		
Number of stores	1:63	0:23	1:78
	(0:16)	(0:35)	(0:76)
Assortment breadth	0:19	0:32	0:31
	(0:03)	(0:09)	(0:18)
Service level	0:94	1:78	1:84
	(0:08)	(0:24)	(0:29)
Discount	4:99	4:74	3:38
	(0:09)	(1:02)	(0:76)
Consumer FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Pseudo R^2	0:08	0:06	0:05
Observations	15099443	1242163	832651
AIC	2284736:85	177229:26	109782:35
BIC	4299213:93	318248:58	199956:48
Log Likelihood	1003727:42	76894:63	47139:18

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

6.2. Analysis on Existing consumers

One valid concern about our results is that some consumers do not know the brand in the first place, and as a result are not aware of retailer's network changes prior to that. To address this concern, we label consumers as newcomers and existing consumers based on their sign-up information. If a consumer has already expressed interest in the retailer's brand before the start of our data-set (January 1, 2018) we count her as an existing consumer, otherwise she is a newcomer. Note that 73.65% of our consumers in the data set are classified as existing. The results are shown in Table 7, and are consistent with the main findings in terms of sign and significance.

6.3. Marketplaces, Outlets and Far Away consumers Analyses

We estimate our Equation (1) for consumers purchasing in marketplaces, outlet stores, and in stores that are far away from their homes. Note that we cannot include our channel specific discount measure for the Marketplaces – since there are multiple marketplaces involved with independent discount policies, and outlet stores – since by definition outlets always sell discounted products.

We conjecture that because marketplace consumers are exposed to multiple brands at the same time, they do not monitor variations in the store network of a particular brand and hence should

Table 7: Impact on propensity to purchase considering only existing consumers.

	P(Online Purchase)	P(Store Purchase)	P(Tablet Purchase)
Distance	0:02	0:21	0:83
	(0:02)	(0:09)	(0:12)
Store accessible	0:66		
	(0:25)		
Number of stores	1:42	0:28	2:38
	(0:19)	(0:36)	(0:53)
Assortment breadth	0:26	0:40	0:10
	(0:04)	(0:08)	(0:09)
Service level	1:41	2:03	2:25
	(0:10)	(0:23)	(0:28)
Discount	3:51	2:23	1:04
	(0:03)	(0:14)	(0:13)
Consumer FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Pseudo R^2	0:08	0:06	0:05
Observations	11195030	1155708	606100
AIC	2194829:45	169228:06	105343:88
BIC	4126110:54	303372:38	191708:47
Log Likelihood	964215:72	73430:03	45226:94

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

not be affected by it. Similarly, outlet consumers, who tend to be price sensitive, may be driven to these channels due to the physical aspects of the brick-and-mortar store network but should not alter their purchase probabilities when the service quality varies. Finally, a consumer purchasing products in a store that is far away from her residency may or may not be more inclined to do so when being aware of the brand (i.e., the physical aspects of the store), but her purchase behaviour should not be affected by the variety and availability in her own catchment area.

Table 8 presents the results, and show that common logic holds: we see that the billboard effect related to distance remains, as well as the positive effect of discounts, but density of the store network, variety and availability are not significant. This suggests that having a store nearby creates awareness about the brand, but adding nearby service intensity or quality does not create positive spillovers into these alternative channels.

6.4. Multinomial Logit Model

Our main model essentially estimates one logit model for each channel, since we introduce consumer-channel fixed effects. One concern with this modeling approach is that we do not directly consider the substitutability across channels. To account for this, we can use a multinomial logit framework. Note that both *Store* and *Tablet* channels are directly associated with a store visit, hence any substitution mechanism between these two channels would not be independent, i.e. including these two channels in the same multinomial logit model would undermine the essential assumption of independence of irrelevant alternatives. We thus change the channel structure to being either *O ine* or *Online* for this analysis, and only consider consumers who demonstrated activity in both channels (3.4% of all consumers – after accounting for the aggregation of *Store* and *Tablet* channels) to focus on

Table 8: Impact on propensity to purchase through an online marketplace, outlet and far away stores.

	P(Marketplace Purchase)	P(Outlet Purchase)	P(Far Away Purchase)
Distance	0.02	0.92	0.38
	(0.34)	(0.08)	(0.04)
Store accessible	12.01	6.17	1.54
	(11.10)	(3.50)	(0.79)
Number of stores	9.13	1.65	1.55
	(15.43)	(0.85)	(0.84)
Assortment breadth	1.61	0.43	0.38
	(0.86)	(0.29)	(0.18)
Service level	3.31	0.24	0.85
	(2.28)	(0.57)	(0.48)
Discount			1.21
			(0.08)
Consumer FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Pseudo R^2	0.07	0.06	0.04
Observations	96300	469302	761420
AIC	11867.78	69664.71	352318.79
BIC	21447.24	121719.43	647673.85
Log Likelihood	4922.89	30125.36	152987.39

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

consumers' cross-channel behavior. Formally, we define the utility that the store network provides to them using the same set of independent variables of our main model: consumer i receives a utility U_{itc} in channel c at each week t defined as

$$U_{itc} = {}_{ic} + {}_{tc} + {}_{c} X_{it} + \varepsilon_{itc}. \tag{5}$$

In this expression, ${}_{ic}$ is the utility driven by personal preference of consumer i for channel c, and ${}_{tc}$ is the utility driven by the conditions specific to week t for channel c, and X_{it} includes the same set of variables explained in Equation (1), except *Store accessible*, which must be equal to one because these consumers always have access to a brick-to-mortar store. Note that, as before, we assume that ε_{itc} is a Gumbel distributed random variable (Anderson et al. 1992).

Given this utility structure, when a consumer chooses a channel to purchase (including no purchase as an alternative channel, i.e. $C = \{\text{Offline}, \text{Online}, \text{No purchase}\}$) that provides the highest utility, the probability of consumer purchases in channel c follows the multinomial logit (MNL), and can be written as

$$P_{itc} = \frac{e^{{}_{c} X_{it} + {}_{ic} + {}_{tc}}}{1 + \sum_{c' \in C} e^{{}_{c'} X_{it} + {}_{ic'} + {}_{tc'}}}. \tag{6}$$

Estimation results for Equation (6) are presented in Table 9.

Results are qualitatively in line with the main findings in §5, but with some differences given the new context and the channel structure. We begin with discussing the effect of geographical proximity to a store. Similar to the main findings, the further consumers are located from a store, the less likely they are to purchase in a brick-to-mortar store. However, for the online channel this is

Table 9: Multinomial logit model results.

	Dependent variable:
	Purchase
ChannelOffline	1.343
	(0.097)
ChannelOnline	2.425
	(0.089)
Discount	4.266
	(0.058)
ChannelOffline:Distance	0.076
	(0.006)
ChannelOnline:Distance	0.148
	(0.009)
ChannelOffline:Number of stores	0.009
	(0.020)
ChannelOnline:Number of stores	0.027
	(0.022)
ChannelOffline:Assortment	0.037
	(0.031)
ChannelOnline:Assortment	0.023
	(0.021)
ChannelOffline:Service level	0.046
	(0.062)
ChannelOnline:Service level	0.376
	(0.066)
ChannelOffline:Discount	2.694
	(0.080)
R^2	0.473
Max. Possible R^2	0.519
Observations	2,545,290
Log Likelihood	116,041.400
Wald Test	380,366.200 (df = 12)
LR Test	1,632,109.000 (df = 12)
Score (Logrank) Test	1,564,948.000 (df = 12)
Note:	$p<0.1$; $p<0.05$; $p<0.01$

reversed, i.e., consumers are more likely to purchase online when they are located further away from a store. Although this might seem contradictory to our initial findings, note that analysis in this section is based only on the consumers who live in the catchment area of at least one store, whereas the previous estimation explained in Section 5.1 considered *all* consumers who showed activity in the *Online* channel. Hence, we repeat the previous analysis on online purchase behavior to consider only the consumers who live in the catchment area of a store, and present the results in Table 14 in Appendix D, where we observe a positive and insignificant coefficient for the distance measure. Thus, we find that the closer consumers are to a store, the more they tend to favor brick-to-mortar over online.

For the remaining variables, we find that the number of stores in the catchment area does not affect the purchase probability (this effect is now exclusively contained in the distance variable). Similarly, larger assortments do not change cross-channel behavior. The coefficient of the service level remains large and negative for the online channel, but becomes insignificant for the offline channel, suggesting that higher service levels from nearby stores reduce the value of the online channel.

Figure 4: Best location picked for Barcelona.

Lastly, discounts increase the likelihood of shopping in-store, but the effect is more pronounced for the online channel.

7. Prescription of Optimal Store Network Design

In the previous sections we have learned that the way a retailer sets up and runs its brick-and-mortar stores has implications for consumers' interactions with the retailer, not only offline but also online. How can the retailer then exploit these findings to optimize its store network in different cities? In this section, we present a counterfactual analysis. We aim to understand the impact of exploiting the store network in terms of structure and functionality on purchase propensity for different channels.

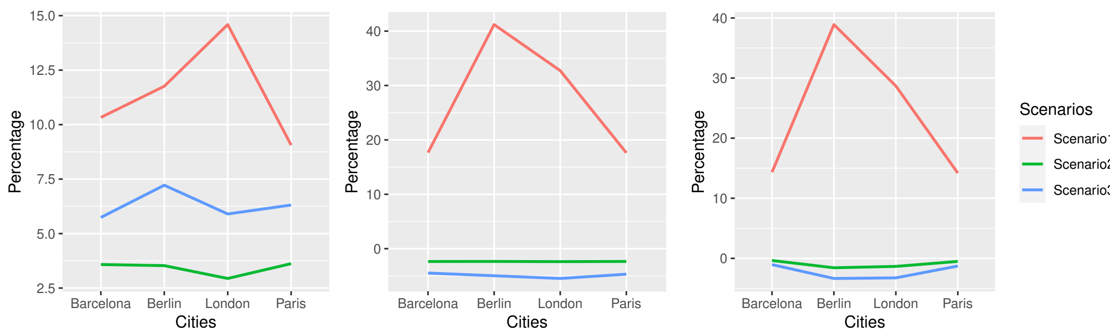
Our first analysis is based on Barcelona, Spain. We seek to find the best location to open a store within the city. We therefore create a grid of 200 200 equidistant geolocations covering the whole city to identify a large enough set of possible locations. Afterwards, we run a predictive analysis to decide which of these locations is the best in terms of its expected addition to average purchase propensity. Since our models include a consumer-level fixed effect, we consider the set of consumers that are within the catchment area of this new store, and manipulate the independent variables *Store accessible*, *Number of stores*, and *Distance*, yet we keep the original *Discount* variable to ensure that we consider the same promotion structure. We then predict the behavior of each consumer with regard to each channel considering the manipulated variables. Lastly, in order to transform our individual-level prediction into an aggregate revenue figure, we calculate the difference in the weighted average of purchase propensities regarding each channel on the basis of the fractions we observed for them in the raw data. Figure 4 illustrates the location picked after the analysis on a map – where our model predicts an increase of more than 14% in the average purchase propensity of consumers in the area.

After fixing the new store location for the region of interest, we execute different scenarios using the same predictive method. We aim to explore the marginal effects of having more stores, providing

Table 10: Results of prescriptive scenarios for Barcelona.

Scenarios	*Store Opening*	*Assortment*	*Service Level*	Brick-and-mortar	Online	All Channels
1	Yes	∅	∅	+10.1%	+16.1%	+14.1%
2	No	+10%	∅	+3.6%	-2.4%	-0.3%
3	No	∅	+10%	+5.8%	-4.5%	-1.0%

Figure 5: Same scenarios in multiple cities.



(a) Brick-and-mortar **(b)** Online **(c)** All channels

better assortment or better service level on each channel. We present our results in Table 10. Note that we consider a percentage increase for both assortment and service level. Consider for example an assortment of 100 units and a service level of 50%, in this case a 10% improvement would increase the assortment to 110 units and the service level to 55%.

We observe that opening a new store has a positive impact on both brick-and-mortar and the online channel, as well as on overall purchases. This is clearly aligned with the discussion on the cross-channel billboard effect of the existence of stores near consumers. Furthermore, providing a better assortment and service level without changing the store network enhances brick-and-mortar purchases, yet it causes a clear cannibalization of online purchases. Note that widening the assortment breadth results in a smaller change than when a higher service level is provided, which suggests that consumers place a much higher value upon what is available to them than upon the entire assortment. Interestingly, we see that the average overall purchase propensity in the region regardless of channel decreases after scenarios 2 and 3 are implemented. This is a crucial result: it shows the relative importance of service quality provided by the store network, which may be detrimental to online channels. Our analysis suggests that the retailer *must* seek to expand its store network in this region instead of improving the service quality, which supports the value of showrooming in omnichannel settings (Bell et al. 2015).

Nonetheless, it is important to remember that this result is based on the existing store network, service quality and consumer base in the area. Thus, we expand our counterfactual analysis to more geographical locations to generalize our results. Results are shown in Figure 5.

We find that scenario 1 has relative importance over scenarios 2 and 3 in all geographical areas. This suggests that retailers can be better off when they prioritize store network expansion over service quality improvement strategies. Furthermore, the magnitude of the predicted change across different areas varies much more for scenario 1 than for other scenarios. We conjecture that, because the intensity of the stores are different in each area, the marginal effect of opening a store might change depending on the location. For instance, we observe 8 stores in Barcelona area, whereas in Berlin there are only 3 stores despite its larger population. Then, if we implement a store opening to both areas, we predict an increase of 14% in the overall sales in Barcelona, yet an increase of 40% in Berlin – a difference of almost three-folds. However, the potential benefits of improving the assortment or the service level across different areas are rather similar in terms of magnitude.

There are multiple important takeaways from our counterfactual analysis. Firstly, managers should be aware that the decisions on the store network and the service that it provides might have different impacts regarding different channels and different geographical locations. An important generic result is that having more stores is beneficial for all channels due to the convenience that they provide. In other words, we observe dominance of store opening policy over assortment / service level improvement policies. However, it should be noted that opening a store and improving the existing stores are not necessarily complementary to each other, since there are multiple dimensions to be considered in terms of spill-over effects and cannibalization, as well as the relevant costs. Hence, managers should study the consumer base in the respective geographical location and consider their channel-specific strategies before making decisions on the structure and functionality of the store networks.

For instance, in the Barcelona region our models predict 10.1% increase for the brick and mortar channel and 16.1% increase for the online channel in the case of Scenario 1. Additionally, we observe that 53% of purchases in this region occurred through the brick-to-mortar stores whereas the remaining 47% through online. Then, a manager can easily turn this into an approximation of expected additional revenue as follows $\mathrm{E}[\text{Additional Revenue } \%_{Scenario1}] = 0.101 \quad 0.53 + 0.161 \quad 0.47 = 0.129.$ This methodology can easily be applied to any possible store network modification scenario for any respective location, which can be a practical and strong tool to guide managerial decisions.

8. Conclusion

In this study, we seek to explore the effect of store network on purchase propensity for omnichannel retailers as more and more retailers combine online and offline channels. We conduct a granular analysis using consumer-level data from a large retailer with omnichannel capabilities. We geolocate consumers and track the store network characteristics around a 5km radius from them, including the number of stores, available assortment and service levels within this ***catchment area***. Using these variables, we model the consumer purchase propensity for each channel – ***online***, ***store***, and ***tablet*** as a two way fixed effects logit model where we control for individual and time, as well as controlling for the promotions.

We find support for a strong billboard effect, in which the closer a consumer is to a store, the more likely she is to purchase through any channel. On the other hand, the impact of service quality depends on the channel: consumers are less (more) likely to purchase online (offline) when they get broader assortments and higher product availability from the store network around them. Using these results, we provide a counterfactual analysis to provide recommendations on store network design for an omnichannel retailer. In this context, it is best to increase store presence, while improving service quality triggers a shift from online to store sales. Since the marginal effect of changing an element of the store network depends highly on the respective location and the channel, we advise omnichannel retailers to customize their store networks to the their existing consumer base. Our methodology is easily applicable and provides a useful tool for omnichannel retailers to better manage their store networks.

Our study brings light to how omnichannel consumers' purchase behavior is affected by the physical store network around them in a granular level. Instead of focusing on individual purchase behavior, it is possible to examine other phenomena regarding the omnichannel store networks, such as cross-channel returns and exchanges. Moreover, many aspects of assortment planning in omnichannel settings can be enlightened such as product-store matching in line with Rooderkerk and Kök (2019). In addition to these future empirical research questions, there is room to develop analytical models to formally optimize the omnichannel store networks by considering the cross-channel functionalities, and related consumer purchase behavior.

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

Adhi, P., A. Davis, J. Jayakumar, and S. Touse. 2020. Reimagining stores for retail's next normal. *McKinsey* April 22:online.

Aguirregabiria, V., and G. Vicentini. 2016. Dynamic Spatial Competition Between Multi-Store Retailers. *The Journal of Industrial Economics* 64 (4): 710–754.

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

Andrews, J. M., V. F. Farias, A. I. Khojandi, and C. M. Yan. 2019. Primal–Dual Algorithms for Order Fulfillment at Urban Outfitters, Inc. *INFORMS Journal on Applied Analytics* 49 (5): 355–370.

Ansari, A., C. F. Mela, and S. A. Neslin. 2008. Customer channel migration. *Journal of marketing research* 45 (1): 60–76.

Avery, J., T. J. Steenburgh, J. Deighton, and M. Caravella. 2012. Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time. *Journal of Marketing* 76 (3): 96–111.

BBC 2020, Oct. Coronavirus: H&M to close 250 shops as Covid drives sales online. *BBC* October 1:online.

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. 2017. O ine 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.

Bell, D. R., T.-H. Ho, and C. S. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *Journal of marketing Research* 35 (3): 352–369.

Bernstein, F., and V. Martínez-de-Albéniz. 2017. Dynamic Product Rotation in the Presence of Strategic Customers. *Management Science* 63 (7): 2092–2107.

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 M. S. Rahman. 2009. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science* 55 (11): 1755–1765.

Burns, L. D., R. W. Hall, D. E. Blumenfeld, and C. F. Daganzo. 1985. Distribution strategies that minimize transportation and inventory costs. *Operations research* 33 (3): 469–490.

Campo, K., E. Gijsbrechts, and P. Nisol. 2000. Towards understanding consumer response to stock-outs. *Journal of Retailing* 76 (2): 219–242.

Campo, K., E. Gijsbrechts, and P. Nisol. 2004. Dynamics in consumer response to product unavailability: do stock-out reactions signal response to permanent assortment reductions? *Journal of Business Research* 57 (8): 834–843.

Caro, F., V. M. de Albéniz, and B. Apaolaza. 2021. The Value of Online Interactions for Store Execution. Working paper, IESE Business School.

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.

Chan, L. M. A., and D. Simchi-Levi. 1998. Probabilistic analyses and algorithms for three-level distribution systems. *Management Science* 44 (11-part-1): 1562–1576.

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.

Drezner, Z., and H. W. Hamacher. 2001. *Facility location: applications and theory*. Springer Science & Business Media.

Federgruen, A., and P. Zipkin. 1984. Approximations of dynamic, multilocation production and inventory problems. *Management Science* 30 (1): 69–84.

Ferreira, K. J., and J. Goh. 2021. Assortment rotation and the value of concealment. *Management Science* 67 (3): 1489–1507.

Fisher, M. L., S. Gallino, and J. J. Xu. 2019. The value of rapid delivery in omnichannel retailing. *Journal of Marketing Research* 56 (5): 732–748.

Fumero, F., and C. Vercellis. 1999. Synchronized development of production, inventory, and distribution schedules. *Transportation science* 33 (3): 330–340.

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. 2018. The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing & Service Operations Management* 20 (4): 767–787.

Gallino, S., and A. Moreno. 2019. *Operations in an omnichannel world*. Springer.

Gallino, S., A. Moreno, and I. Stamatopoulos. 2017. Channel integration, sales dispersion, and inventory management. *Management Science* 63 (9): 2813–2831.

Gao, F., V. V. Agrawal, and S. Cui. 2022. The effect of multichannel and omnichannel retailing on physical stores. *Management Science* 68 (2): 809–826.

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.

Geyskens, I., K. Gielens, and M. G. Dekimpe. 2002. The market valuation of internet channel additions. *Journal of marketing* 66 (2): 102–119.

Godfrey, G. A., and W. B. Powell. 2001. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science* 47 (8): 1101–1112.

Gürbüz, M. Ç., K. Moinzadeh, and Y.-P. Zhou. 2007. Coordinated replenishment strategies in inventory/distribution systems. *Management Science* 53 (2): 293–307.

Haghighatnia, S., N. Abdolvand, and S. Rajaee Harandi. 2018. Evaluating discounts as a dimension of customer behavior analysis. *Journal of Marketing Communications* 24 (4): 321–336.

Hearne, R., A. Podreciks, N. Uhlenbrock, and K. Ungerman. 2019. Supercharging retail sales through geospatial analytics. *McKinsey* March 15:online.

Hekmatfar, M., and R. Z. Farahani. 2009. *Facility location*. Springer.

Hwang, E. H., L. Nageswaran, and S.-H. Cho. 2020. Impact of COVID-19 on Omnichannel Retail: Drivers of Online Sales during Pandemic. *Available at SSRN 3657827*.

Iyer, A. V., and L. E. Schrage. 1992. Analysis of the deterministic (s, S) inventory problem. *Management Science* 38 (9): 1299–1313.

Iyer, G., C. Narasimhan, and R. Niraj. 2007. Information and inventory in distribution channels. *Management Science* 53 (10): 1551–1561.

Kök, A. G., and A. S. Simsek. 2021. Variety and Inventory Trade-off in Retailing: An Empirical Study. Working paper.

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.

Lemon, K. N., and P. C. Verhoef. 2016. Understanding customer experience throughout the customer journey. *Journal of marketing* 80 (6): 69–96.

Liu, K., Y. Zhou, and Z. Zhang. 2010. Capacitated location model with online demand pooling in a multi-channel supply chain. *European Journal of Operational Research* 207 (1): 218–231.

Lufkin, B. 2020, Jul. The curious origins of online shopping. *BBC* July 22:online.

Martin, C. H., D. C. Dent, and J. C. Eckhart. 1993. Integrated production, distribution, and inventory planning at Libbey-Owens-Ford. *Interfaces* 23 (3): 68–78.

Martínez-de Albéniz, V. 2019, January. Omnichannel strategy at Camper. IESE Business School case study.

Millstein, M. A., and J. F. Campbell. 2018. Total hockey optimizes omnichannel facility locations. *Interfaces* 48 (4): 340–356.

Musalem, A., M. Olivares, E. Bradlow, C. Terwiesch, and D. Corsten. 2010. Structural Estimation of the Effect of Out-of-Stocks. *Management Science* 56 (7): 1180–1197.

Nishida, M. 2015. Estimating a model of strategic network choice: The convenience-store industry in Okinawa. *Marketing Science* 34 (1): 20–38.

Ofek, E., Z. Katona, and M. Sarvary. 2011. 'Bricks and clicks': The impact of product returns on the strategies of multichannel retailers. *Marketing Science* 30 (1): 42–60.

Paton, E. 2020. Shopping for Fashion, Six Months On. *The New York Times* September 18:online.

Rooderkerk, R. P., and A. G. Kök. 2019. Omnichannel assortment planning. In *Operations in an omnichannel world*, 51–86. Springer.

Semykina, A., and J. M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157 (2): 375–380.

Tan, T. F., and S. Netessine. 2014. When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* 60 (6): 1574–1593.

The Economist 2021, Jan. Schumpeter: How Inditex is refashioning its business model. *The Economist* January 16:online.

Valentini, S., E. Montaguti, and S. A. Neslin. 2011. Decision process evolution in customer channel choice. *Journal of Marketing* 75 (6): 72–86.

Wang, K., and A. Goldfarb. 2017. Can offline stores drive online sales? *Journal of Marketing Research* 54 (5): 706–719.

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.

Zhang, D. J., H. Dai, L. Dong, Q. Wu, L. Guo, and X. Liu. 2019. The value of pop-up stores on retailing platforms: Evidence from a field experiment with Alibaba. *Management Science* 65 (11): 5142–5151.

Appendices

A. Gender Assignment

In the raw data, there is no gender information attached to product identifiers. However, there is a gender identifier for the consumers in transaction data, yet it is only available for 26.11% of the transactions. Therefore, we substitute this missing information with the following methodology.

Gender assignment to products

To assign gender information to each product, we use transaction and inventory data sets. First, we search for gender keywords –*MAN*, *WOMAN*, Women, etc.– in the product description within the transaction data and assign a gender (*MAN* or *WOMAN*) to those products. For those that do not include this information, we count how many times a consumer whose gender is known bought a specific model, and if we observe more than 95% prevalence of a gender in transactions, we assign this product to the respective gender. For the remaining products, we go through size range of the products and assign gender as follows: sizes in range 35-51 must be a *MAN* product, whereas 23-42 must be a *WOMAN* product.

Gender assignment to consumers

Once we assigned genders to each product, we screen all the consumer transactions where gender information is missing, and calculate the purchase frequency of products from both genders. Afterwards, we simply assign the more frequently bought gender as the consumer's gender. We observe no occasions where both frequencies are equal.

B. Instrument Variable Regression – First Stage Results

First stage estimation results for the instrumental variables analysis are represented in the Table 11 for each endogeneous independent variable.

C. DID - Matching

We conduct a matching difference-in-differences analysis to study the effect of store openings to ensure that data cleaning process (see §3) does not bias our results. Hence, we choose the unit of analysis as wider regions, in order to recover the information on the transactions that were eliminated earlier for the main model because they were not linked to an identifiable consumer.

Initially, we go back to our raw data set and start by filtering for those who are associated with a full address. Furthermore, we assume that those who do not have the address information but associated with a store are located within the same wide region of the store. This allows us to recover as many transactions as possible. Since we focus on the sales observed in Europe for the previous analysis, we choose the NUTS 3 level regions of the European Union's Nomenclature of Territorial Units for Statistics to ensure a standard amongst the regions. We assign each transaction to the

Table 11: First stage estimation results for the instrumental variables analysis.

		First Stage Results per Endogeneous Variable					
	Distance	Number of stores	Assortment	Service level	Discount (Online)	Discount (Store)	Discount (Tablet)
Lag distance	0:01 (0:00)						
Log(Total inventory)		0:02 (0:01)					
Lag assortment			0:01 (0:00)				
Lag service level				0:68 (0:01)			
Lag discount					0:07 (0:00)	0:26 (0:00)	0:13 (0:00)
Consumer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0:99	0:99	0:99	0:99	0:81	0:87	0:76
Observations	19177834	21092699	21092699	21092699	1279878	15266867	843442
AIC	2243270:15	144192506:40	57647748:16	111323225:55	3188147:87	43030339:29	1831220:65
BIC	4851315:70	144188047:06	57643288:83	111318766:22	3045197:85	40992070:84	1739840:40
Log Likelihood	945049:08	72096553:20	28824174:08	55661912:78	1605924:94	21655341:65	923457:32

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

respective NUTS 3 level regions. Afterwards we aggregate the data for every region and create a panel data set where we observe the weekly total sales in each region along with total online sales and total physical store sales. We further locate the store openings and introduce those in the panel data as our treatment variable.

Since the final data set is not balanced in terms of the amount of regions in the treatment (8 regions) and control (1045 regions) groups, we apply a matching methodology. We incorporate further information on the regions in order to ensure a good quality matching. We introduce the following measures obtained through the official *Eurostat* website to represent the socioeconomic structure of regions.

- Population – Total number of people living in the region, as a proxy to potential and actual market size
- GDP – Total gross domestic product of the region, as a proxy to income level of individuals
- Population density – Average number of people living in the region per km^2, as a proxy to the difference in urban and rural population

Moreover, we introduce the following measures to be able to capture the presence of the brand in the regions at the time of the respective store opening.

- $StoreAccess_i$ – dummy variable showing whether at least one store had already been located in region i before the start of the time horizon of this study or not.
- $PreviousSales_i$ – the total overall sales observed in region i before the time of the respective store opening

- $PreviousOnlineSales_i$ – the total online sales observed in region i before the time of the respective store opening
- $PreviousPhysicalSales_i$ – the total physical store sales observed in region i before the time of the respective store opening

Note that we repeat the calculation of the variables $PreviousSales_i$, $PreviousOnlineSales_i$, and $PreviousPhysicalSales_i$ for each store opening, and match only one specific store-opened region, with a control region. This is because we observe the store openings scattered around our time horizon, and we do not take the homogeneity in regions throughout time as granted, so that this approach should deliver better matches. Descriptive statistics of these variables is shown in Table 12. Note that the variables are only shown for one sample store opening. After these covariates have been computed, we run a propensity score matching using the previously explained variables.

Table 12: Descriptive Statistics for the matching sample before matching.

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Treatment	1,053	0.008	0.087	0	0	0	1
Population	1,053	396,433.600	483,397.400	11,154	140,921	472,699	6,641,649
GDP	1,053	12,270.240	20,967.150	191.720	3,563.210	12,183.220	237,408.100
Population Density	1,053	435.024	1,077.323	1.900	72.900	324.800	21,069.800
Previous Sales	1,053	217.384	1,126.305	0	19	109	20,437
Previous Online Sales	1,053	120.472	360.517	0	18	100	5,836
Previous Physical Sales	1,053	96.800	809.713	0	0	7	15,227

Taking advantage of our panel data context, we propose the following DID model:

$$log(SALES_{itc}+1) = {}_{ic} + {}^{\theta}_{it} + {}_{1}TREAT_i + {}_{2}AFTER_t + {}_{3}(TREAT_i \quad AFTER_t) + \varepsilon_{itc} \quad (7)$$

In Equation (7), ${}_{ic}$ is region - channel fixed effect, ${}^{\theta}_{it}$ is the time - channel fixed effect, $SALES_{itc}$ is the total sales occurred in region i, at week t, through channel $c \in (overall; online; physical)$, $TREAT_i$ is a dummy variable representing whether a store opened in region i or not, and $AFTER_t$ is a dummy variable representing whether a specific week t is before or after the time of the respective store opening. Our variable of interest in this context is $TREAT_i \quad AFTER_t$, which would represent the effect of store openings after the time of the opening. We estimate Equation 7, and report the results in Table 13.

Our results suggest that sales increase in all channels after the introduction of a new store to a wide region, in alignment with our results in §5. Furthermore, the positive effect is less strong both in magnitude and significance for the online channel in comparison with the offline channel, as also seen in §5, and in line with the literature (Bell et al. 2015, Kumar et al. 2019). After this analysis, we conclude that the sample selection that we had to go through in the early stages of the study does not distort the results, since we observe the same trend when we include those lost observations using a more aggregate unit of analysis.

Table 13: Matching DID results.

	All Sales	Store Sales	Online Sales
TREAT	0:16	0:16	0:32
	(0:08)	(0:08)	(0:15)
AFTER	0:37	0:48	0:11
	(0:11)	(0:14)	(0:05)
TREAT:AFTER	1:16	2:28	0:16
	(0:10)	(0:12)	(0:06)
Region FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	1582	1582	1582
R^2	0:72	0:69	0:76
Adj. R^2	0:79	0:58	0:79

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).

D. Supplementary Results

Table 14: Online purchase propensity considering only the consumers who live within 5km catchment area of a store.

	P(Online Purchase)
Distance	0:09
	(0:07)
Number of stores	1:05
	(0:17)
Assortment	0:32
	(0:05)
Service level	0:81
	(0:13)
Discount	2:01
	(0:06)
Consumer FE	Yes
Time FE	Yes
Pseudo R^2	0:09
Observations	3740440
AIC	548806:02
BIC	996936:18
Log Likelihood	240285:01

Note. Significance levels: 5%( ), 1%( ), and 0.1%( ).