How Physical Retail Channels Impact Customers' Online Purchase Behavior?

Anuj Kumar
Assistant Professor
Warrington College of Business Administration
University of Florida
akumar1@ufl.edu

Amit Mehra
Associate Professor
Indian School of Business
Hyderabad, India
amit_mehra@isb.edu

Subodha Kumar
Associate Professor
Mays Business School
Texas A&M University
subodha@tamu.edu

This Version
December 2014

Abstract

How does reducing customers' store access costs affect the quantity and nature of their online purchase? We answer this question by designing a quasi-natural experiment around the event of store openings by a large US fashion apparel and accessories multichannel retailer. While the distance from the nearest store (store access costs) was significantly reduced for customers living in the vicinity of the opened store (affected customers), it remained unaffected for other customers (unaffected customers). We utilize both propensity score based inexact matching and coarsened exact matching methods to estimate the causal treatment effect of reduction in affected customers' store access costs on their purchase and return behaviors. Our estimates reveal that reduction in customers' store access costs results in their (1) higher store purchases and returns, (2) higher, more diverse, and more expensive online purchases, and (3) higher net total purchases on all channels combined. We propose two mechanisms, information effect and return effect that could explain the effect of easier store access on customers' online purchase behavior and provide empirical evidence for them on our field data.

Key Words: *Electronic commerce; multichannel customer behavior; store versus online channel interactions; causal inferences; average treatment effect; propensity score estimators; coarsened exact matching estimators*

1. Introduction

Retailers who have traditionally relied on the store channel are increasingly depending on their online channels to deliver sales growth. For example, the sales on Walmart.com increased by 30% compared to overall sales growth of 1% in the first quarter of 2013.[1] The e-commerce sales at Gap increased by 13% while sales in their department stores declined in the last quarter of 2013 leading to a meager growth rate of 1.2% in the overall sales.[2] This trend has led to many retailers scaling back their store presence while investing more in their online channels. Thus, Gap has closed down 20% of their stores over the years (McIntyre and Hess 2014) and Nordstrom is considering increased investments in their online channel (Cook 2014). Hence, the impact of store presence on the sales at the online channel must be well understood as retailers rebalance their investments across the two channels.

A significant body of the existing literature examines how sales on store channel get affected with increased adoption of the online channel (Ansari et al. 2008, Biyalogorsky and Naik 2003, Danaher et al. 2003, Deleersnyder et al. 2002, Geykens et al. 2002). However, the causal impact of improving customer access to the store channel of a retailer on the online sales of the same retailer has not yet been definitively studied. In the context of competing store and online retailers, Foreman et al. 2009 show that the online sales reduce when a store opens in a particular location. A similar cannibalistic outcome is possible when the store and online channels belong to the same retailer. However, there is some evidence that the online sales of a multichannel retailer from a particular location increase when the retailer opens a new store in that area (Avery et al. 2012, Bell et al. 2013). Given the conflicting and unclear nature of the evidence, it is an open empirical question whether improving access to the store of a multichannel retailer would result in increase or decrease in existing customers' purchases at the online channel of that retailer. This is the first research question that we examine.

[1] http://news.walmart.com/news-archive/investors/walmart-reports-a-46-percent-increase-for-q1-eps-of-114-us-businesses-forecast-positive-comp-sales-for-q2-1820850
[2] http://www.internetretailer.com/2014/05/22/e-commerce-accounts-all-q1-growth-gap

A second aspect of impact of store access on the online sales is how the basket of goods sold at the online channel changes. As far as we know, no academic study has analyzed this question so far. There are important implications of this question on inventory management at the online channel. For example, with increased store access, if customers purchase more of the same products they buy, the inventory of these products must be increased. However, if customers increase the diversity of the products they purchase, the inventory mix should be changed accordingly.

Finally, it will be useful to understand how easier store access could increase the sales on the online channel. Based on prior literature, we posit two possible mechanisms: one, more convenient store visits due to easier store access could increase the extent of product search by customers at the online as well as the store channel (*information effect*). Hence, customers are exposed to wider variety of products. Second, the risk of a wrong purchase decreases since product returns are easier with easier store access (*return effect*). Therefore, customers are more amenable to buy, and may even consider purchasing more expensive products because of reduced risk. Overall, facilitating customers' store access could affect the quantity and nature (product diversity and average product price) of their online purchases because of these two mechanisms. Accordingly, our third research question is to analyze whether empirical evidence exists in support of these two mechanisms.

To answer the research questions identified above, we utilized the event of new store openings by a large retailer of fashion apparel, accessories and home products in the US. Due to such store openings, while the distance from the nearest store significantly reduced for customers (affected customers) living near the newly opened store, it remained unchanged for customers (unaffected customers) living in other parts of the US. We designed a quasi-natural experiment around the event of such store openings to estimate the treatment effect of reduction in customers' store access costs on their purchase and return behaviors on the store, online, and all channels combined. We obtained our treatment effect estimates by computing the difference in purchase and return behaviors of the affected customers with that of their matched unaffected customers. We used customer-level data and utilized propensity score based inexact matching, as well as coarsened exact matching methods to match the two groups of customers, and found qualitatively similar treatment effect estimates from the two methods.

Our treatment effect estimates indicate that the reduction in store access costs for customers results in their (1) higher purchases and returns (in number of transactions, quantity, and revenue) on the store channel, online channel, and on all channels combined (store, online and others like catalogue etc.), and (2) higher net total purchases (purchase – returns) on all channels combined. We also found that customers purchased more diverse and higher priced products on the online channel due to store openings.

We further separately estimated the treatment effect of store openings for two subcategories of affected customers based on whether their distances from the newly opened store is in the lowest quartile (near customers) or the highest quartile (far customers) of their distribution of distances from the nearest store. Due to store openings, while the distance from the nearest store had significantly reduced for the near customers, it still remained prohibitive for the far customers. We found the effect of store openings in case of the near customers but not for far customers. We also conducted a similar analysis for other stores opened by the retailer in areas where it had pre-existing stores and therefore customers' distances from the nearest store did not change significantly due to such store openings. In this case, we did not find any significant effect of store openings on customers' purchase and return behaviors. These results provide clear evidence that the effect of store openings on customers' purchase and return behaviors is caused by the change in their store access costs.

To shed light on the mechanisms through which easier store access may affect the quantity and nature of customers' online purchases, we estimated the treatment effect of store openings for different subcategories of affected customers based on whether they make higher/less or equal store purchase and return transactions with the retailers after the store openings. We found the increase in both the quantity (number of transactions, quantity, and revenue) and nature (purchase diversity and average product price) of online purchases for only those affected customers who make either more store purchase or return transactions in the post store opening period. These results provide direct empirical support to our proposed mechanisms that higher store visits due to lower access costs not only exposes customers to a wider variety of retailer's products but also mitigates their risks of online purchases, which, in turn, affects their online purchase behavior.

Our paper makes several contributions to the literature. First, our study combines the power of customer-level data with a quasi-natural experiment to provide causal estimate of facilitating customers' interactions on the store channel on their purchase and return behaviors on the online channel. Previous studies either used experimental setup with zip code-level aggregate data (Avery et al. 2012, Bell et al. 2013) or used customer-level data without an experimental setup (kushwaha and Shanker 2013, Venkatesan et al. 2007). Second, we propose two mechanisms through which higher store visits due to easier store access could increase the quantity, diversity, and average product price of online sales. We provide empirical evidence of these mechanisms on our field data. Finally, prior studies have only examined how the option to return on a channel would affect the sales on that channel (Anderson et al. 2009). Our study studies the novel aspect of how (and why) facilitating returns on one channel (due to easier access to store channel) affects the demand on another channel (online channel). To our knowledge, we are the first to provide evidence of such cross channel interactions between product returns and purchases.

The rest of the paper is organized as follows. We present the related literature and develop a conceptual framework in Section 2; describe our field setup, empirical design, and data description in Section 3; present our empirical results and robustness checks in Section 4; and conclude with managerial implications and opportunities for future extensions in Section 5.

2. Related Literature and Conceptual Framework

A direct implication of easier store access is that customers may visit the store more often and fulfill their needs by increasing their purchases at the store while decreasing their purchases at the online channel of the retailer. Some empirical support for this substitutive relationship between sales on stores and the online channel exists in extant literature. For example, in the context of competition between pure-play online retailers and brick-and-mortar stores, Foreman et al. (2009) show that customers substitute away from online purchasing when a store opens in their vicinity, while Brynjolfsson et al. (2009) show that Internet retailers face significant cannibalization from brick-and-mortar retailers when selling mainstream products.

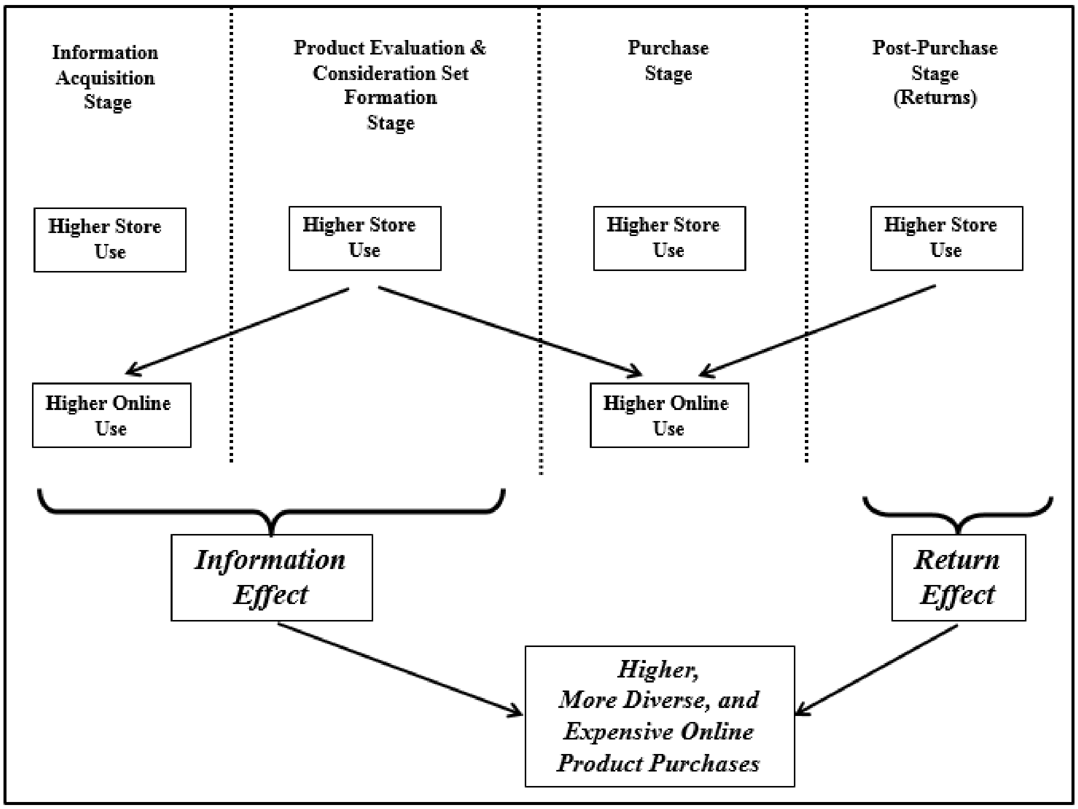


Figure 1. Conceptual framework for impact of easier store access on online sales

Therefore, while the rationale and evidence exists in favor of substitutive effect between the online and store channel sales, it is also possible that synergies between the two channels drive a complementary relationship when the two channels belong to the same retailer. For better understanding of possible synergies between these channels, we develop a comprehensive framework of the various mechanisms though which easier store access may affect store and online sales as shown in Figure 1. This framework is based on Verhoef et al. (2007) and marries the needs of a customer during different stages of the purchase process: information acquisition, consideration set formation, purchase, and post-purchase transactions with the capabilities of the online and store channels in the context of sensory products like apparel and their accessories (our focus in this study).

Easier store access could affect customers' use of the two channels in information acquisition stage. When customers acquire product (especially sensory products) information at the online channel, they may not be able to form their consideration set solely based on this information because they cannot touch and try the products physically. A store visit allows for a more comprehensive product evaluation, enabling customers to form their consideration sets. Ratchford et al. (2003) suggest that the extent of information search on the online channel depends upon the expected benefit of the search. With the opening of a store nearby, customers can conduct product evaluations by visiting the store at low cost (Pauwels et al. 2011). Therefore, they may find it beneficial to conduct increased information search on the online channel. Such increased online search may inform customers not only about the product they were searching for, but also about other products, thus exposing customers to a wider variety of products.

Second, when customers visit a store, they are exposed to not only the products they intended to evaluate but also other displayed products in store. Moreover, the sales representatives also recommend products to customers during such store visits. The information acquired about products not under active consideration during store visits could widen customers' exposure to products for future purchase. Therefore, increased store visits due to easier access to stores would result in wider exposure to the retailer's product assortment.

We refer to the increased exposure to product information due to the two processes described above as the *information effect*. This effect results in formation of a wider consideration set of products, perhaps also comprising of more expensive products due to increased opportunities of product evaluation at the store. The implication is an increased quantity, diversity and average purchase price of products not only at the store channel, but also at the online channel. This is because customers, even after having formed their consideration sets after product evaluation at the store, may not immediately purchase the product. They may want to take more time to consider their purchase, or consult with their family and friends before they commit to the purchase. Once they are ready to purchase, they can easily do this using the convenience at the online channel without travelling to the store.

Due to limited product information and evaluation opportunities on the online channel, customers face the uncertainty of fit between attributes of sensory products and their needs/expectations. Improved access to a store reduces customers' uncertainty in online purchase as they can easily return the purchased products at the nearby store if it does not match their needs. Anderson et al. (2009) show that the availability of return option increases customers' demand on the catalog channel. Therefore, we expect increased availability of the option to return at nearby store to result in customers purchasing an increased quantity, diversity and higher priced products on average from the online channel. We refer to the impact of this mechanism as the *return effect*.

Overall, while there is evidence of substitutive effect, there is also a compelling reason for complementary effect between online and store channels. Therefore, the net effect of facilitating store access is an open empirical question.

3. Research Setting, Empirical Strategy and Data Description

3.1 Field Setup

We conduct our study on a large fashion apparel, accessories and home products retailer in the US.[3] The retailer mainly sells products through physical stores and a website.[4] The retailer has annual revenue of over US $1.0 billion. In the present study, we utilize the event of store opening by the retailer to examine the effect of reduction in customers' store access costs on their purchase and return transactions on the store and online channels as well as their total purchases and returns with the retailer.

Out of the total customer population of the retailer, we collected transaction data for a random sample of 1.56 million customer households, who made at least one purchase with the retailer in the period from 1999 to 2006. Specifically, we collected data on age of the head of household, income level

[3] The identity of the retailer is not disclosed due to a non-disclosure agreement.
[4] Store and online sales account for roughly 95% of the total sales of the retailer.

of household, and zip code of the household. We also collected zip codes for all stores opened by the retailer in the year 2003-04. For each customer household (hereafter referred as customer) in our sample, we computed the distance from the nearest store before and after the opening of each store in the 2003-04 period. For most customers in our sample, the distance from their nearest store remained unaffected due to the opening of new stores. However, for some customers, the distance and hence their access costs from their nearest store decreased due to opening of a store in vicinity of their residence. Hence the event of store opening results in a treatment of reduction in store access costs for some customers (treated customers) but not for others (control customers). We exploit this differential effect of store openings in our quasi-natural experiment design to estimate the causal effect of reduction in customers' store access costs on their purchase and return behaviors.

3.2 Experimental Design

We identified two stores opened by the retailer in fall 2003 in areas where they did not have any pre-existing stores in a radius of 200 kilometers, hereafter referred to as type-1 store openings.[5] For customers residing in the vicinity of a newly opened store, the distance from the nearest store reduced substantially, hereafter referred to as affected customers. But for customers residing in other parts of the US, the distance from their nearest store remained unaffected, hereafter referred to as unaffected customers. The treatment effect of change in store access costs on customers' purchase/return behaviors can be estimated by computing the difference in the purchase/return behaviors of affected customers from that for unaffected customers. However, the purchase and return behaviors of affected and unaffected customers may be affected differently due to differences in *customer-specific*, *time-specific*, and *retailer-specific* factors for the two groups of customers. In figure 1, we describe how we control for these effects by using a combination of difference-in-difference design with matching estimators in our empirical design.

[5] The exact store opening date is not reported to keep the identity of the retailer confidential

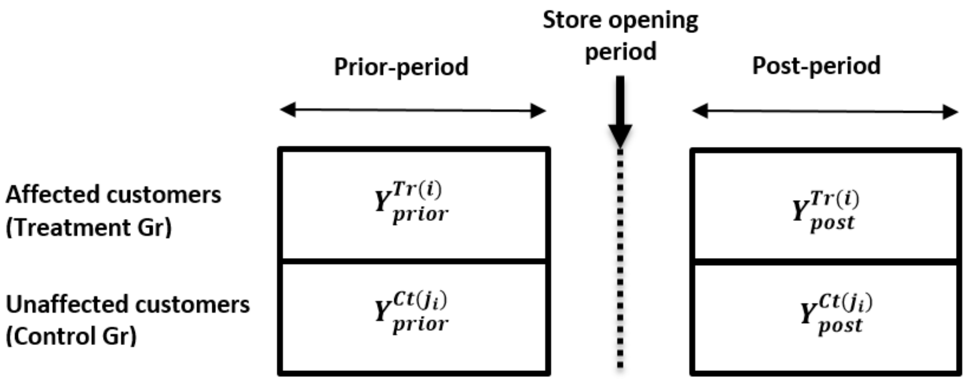


Figure 1: Quasi-natural Experimental Design

The average treatment effect of store opening on treated customers (*ATT*) is computed with the following specification (1).

$$ATT = \frac{1}{N_i}\sum_i \left[\left(Y^{Tr(i)}_{post} - Y^{Tr(i)}_{prior}\right) - \frac{1}{N_{j_i}}\sum_{j_i}\left(Y^{Ct(j_i)}_{post} - Y^{Ct(j_i)}_{prior}\right)\right], \qquad \text{---------} \ (1)$$

where, i = 1, 2, ….. N_i denotes the affected (treated) customers and j_i = 1, 2, …. N_{j_i} denotes the matched unaffected (control) customers for the affected customer i. The details on how we match affected and unaffected customers are provided in the next section. $Y^{Tr(i)}_{prior}$ and $Y^{Tr(i)}_{post}$, respectively, denote the purchase/return variables for the affected customer i in prior- and post-periods. Similarly, $Y^{Ct(j_i)}_{prior}$ and $Y^{Ct(j_i)}_{post}$, respectively, denote the purchase/return variables for the matched unaffected customer j_i in prior- and post-periods.

Because the area to open a new store is not decided randomly by the retailer, there may be systematic differences between affected and unaffected customers. For example, the purchase and return behaviors of the affected and unaffected customers could be different before the store opening. We control for these differences between the two groups of customers by matching each affected customer (i) with a sample of unaffected customers (j_i) on their characteristics observed in our data such as demographic characteristics, distance from the nearest store, and cumulative purchase and return behaviors prior to the store opening. As long as we assume that the differences in purchase and return behaviors of the affected and unaffected customers can be accounted for by the differences in their aforementioned observed characteristics, the counterfactual purchase and return behaviors of the affected customer (i), in absence of store opening, can be obtained from the purchase and return behaviors of their matched sample of unaffected customers (j_i). Therefore, the difference between the purchase/return behaviors of the affected customer and the average purchase/return behaviors of her matched sample of unaffected customers during the post store opening period would provide the causal treatment effect of reduction in store access costs on the affected customer's purchase and return behaviors.

Besides on observed characteristics, the affected customers may differ from unaffected customers on a variety of factors unobserved to us but taken into consideration by the retailer while deciding to open a store at a specific location. For example, the retailer may decide to open a store in an area based on several area-specific factors such as socio-demographic profiles of residents, level of competition for its products, trends in population preferences, pace of development, and trends in earned income. Due to these unobserved area-specific factors in the area of store opening, the purchase and return behaviors of the affected customers may differ from that of the unaffected customers. As most of these factors are not observed in our data, such differences in the behaviors of affected and unaffected customers would not be captured in the observed pre-treatment variables used in the matching algorithm. Therefore, the *ATT* computed by taking the difference of purchase/return behaviors of the affected customers with their matched unaffected customers in the post-period may capture the effect of the differences in their behavior due to these unobserved area-specific factors. As long as the effect of these unobserved factors on customer behavior do not change with time, it can be differenced out by computing the *ATT* by taking the difference of change in purchase/return behaviors of the affected customers from prior- to post-period $\left(Y_{post}^{Tr(i)} - Y_{prior}^{Tr(i)}\right)$ from the average of corresponding values for their matched unaffected customers [$\frac{1}{N_{j_i}}\sum_{j_i}\left(Y_{post}^{Ct(j_i)} - Y_{prior}^{Ct(j_i)}\right)$], a difference-in-difference design as proposed in specification (1). Note that this specification is analogous to combining fixed-effects with matching estimators.

During the store opening period, there may be increased local publicity and word-of-mouth about the retailer. Such high awareness about the retailer around store opening period could additionally influence the purchase/return behaviors of affected customers as compared to their matched sample of unaffected customers. Therefore, the differences in purchase and return behaviors of the affected and unaffected customers could also include the effect of increased awareness about retailer along with the effect of reduction in the affected customers' distance from the nearest store. From our discussions with the retailer's representatives, we learnt that the increased publicity about the new store lasts for about two to three months around the date of store opening. Therefore, we remove a total of six months period (three months before and three months after) around the date of store opening from our analysis (shown as store opening period in Figure 1). This way, we get rid of the effect of increased publicity during the store opening period on customers' purchase and return behaviors but still pick up the effect of reduction in their store access costs.

To control for the *time-specific unobserved factors* that may have a similar effect on the affected and unaffected customers such as seasonality and inflation, we compare the purchase/return behaviors of the affected and unaffected customers in the same calendar time period. In the present context, the retailer uniformly applies promotions across different channels (store and online channels) and across the geographical areas in the US. Therefore, by comparing the purchase/return behaviors of the two groups of customers located at different geographic locations in the same calendar time periods, we account for *retailer-specific unobserved factors* that may influence customers' purchase and return behaviors.

To further establish that the difference in purchase and return behaviors of the matched affected and unaffected customers is due to reduction in affected customers' store access costs only, we identified two stores opened by the retailer in fall 2003 in the areas where it had a pre-existing store within the radius of 50 kilometers, hereafter referred to as type-2 store openings. Therefore, in case of type-2 store openings, the changes in affected customers' distances from their nearest store were marginal. If we can show significant differences between the purchase/return behaviors of the affected customer and their matched sample of unaffected customers in type-1 store openings but insignificant differences in case of type-2 store openings, it would provide a strong evidence that the effect on customers' purchase and return behaviors in case of type-1 store openings is caused by the reduction in their distance from the nearest store only.

In line with above empirical design, we first describe the data in our field setup and then discuss the *ATT* estimates in the following sections.

3.3 Data Description

Out of the total sample of 1.56 million customers, we first identified the affected customers for the two type-1 openings and pooled them together for our analysis. Similarly, we identified the sample of affected customers for the two type-2 store openings. Thereafter, from the sample of affected customers for type-1 store openings, we only kept those customers whose distance from the nearest store is reduced below 150 kilometers due to the store openings, as a reduced distance of more than 150 kilometers would still keep the store transportation costs prohibitive for the affected customers.[6] Then, from the total unaffected customers, we identified a sample of 200,000 unaffected customers with a similar distribution of distance from the nearest store as for the sample of affected customers in the period prior to store opening.[7] Next, for this analysis, we only considered pre-existing affected and unaffected customers who made at least one purchase or return with the retailer prior to the date of store opening. In other words, we did not consider the purchases/returns of newly acquired customers after the store openings, as the post-store opening purchases/returns for them would be naturally higher. We got our final samples of 8883 affected and 84911 unaffected customers in case of type-1 store openings and 7900 affected and 82524 unaffected customers in case of type-2 store openings.

[6] We also tried several other threshold distance values in kilometers such as 100, 125, 175, and 200 instead of 150 and find qualitatively similar results.

[7] There were over one million unaffected customers with similar distribution of distances from the nearest store as for our sample of affected customers. We only selected 200,000 unaffected customers to keep the computational load in our analysis reasonable.

Table 1 reports the distribution of distance from the nearest store for the samples of two groups of customers in each type of store openings. Table 1 reveals a similar distribution of distance from the nearest store for the affected and unaffected customers in both types of store openings in the prior period. Moreover, Table 1 further reveals a substantial reduction in the distance values for the affected customers in case of type-1 store openings but marginal reduction in case of type-2 store openings.

Table 1: Summary statistics of distance from the nearest store

Distance from the nearest store in kilometers	Type-1 store openings			Type-2 store openings		
	Affected customers (8883)		Unaffected customers (84911)	Affected customers (7900)		Unaffected customers (82524)
	Prior-period	Post-period	Prior/post period	Prior-period	Post-period	Prior/post period
Mean	207.2	47.3	176.4	32.1	26.1	26.3
Std. Dev.	67.7	44.4	80.0	39.4	37.0	33.8
1 percentile	92.7	0.0	76.5	1.7	0.4	2.0
25 percentile	130.8	10.5	107.8	9.6	7.4	7.3
50 percentile	228.3	22.9	156.9	22.0	12.6	12.9
75 percentile	245.7	94.9	232.2	34.4	23.0	26.8
99 percentile	323.2	145.4	363.8	164.2	148.4	158.1

Next, we collected data to capture the quantitative and qualitative aspects of the purchase and return behaviors of the customers. For quantitative aspect, we collected data on the number of transactions, quantities, and revenue of the purchases and returns made by a customer on a channel in the prior- and post-periods. Data was collected for the purchase and return variables on the store, online, and other channels offered by the retailer. We report the summary statistics for change in these variables from prior- to post-period on store, online and all channels combined in Table 2, which indicates a higher change in number of transactions, quantities, and revenue for purchase transactions on store and online channels for affected customers as compared to the unaffected customers in case of type-1 store openings and smaller differences in the change of corresponding values between the two groups of customers in case of type-2 store openings.[8] This provides preliminary evidence for a higher store and online purchases

[8] We also performed Welch's t-test for difference in the mean values of change in online and store purchase behaviors for the affected and unaffected customers and mostly found statistically different mean values in case of type-1 store openings but statistically similar mean values for type-2 store openings.

by affected customers due to reduction in their store access costs. Moreover, we find a higher increase in number of transactions, quantities, and revenue for return transactions on the store channel only for affected customers as compared to the unaffected customers in case of type-1 store openings. From Table 2, we also find a higher change in total purchase and return variables for the affected customers as compared to unaffected customers in the type-1 store openings but similar values for the two customer groups in type-2 store openings.

Table 2: Summary statistics for change in purchase/return behavior

Change in Variables from Prior- to Post-period	Type-1 store openings				Type-2 store openings			
	Affected Customers (8883)		Unaffected Customers (84911)		Affected Customers (7900)		Unaffected Customers (82524)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Purchase transactions								
Total No. of transactions	0.51	5.04	0.25	4.85	0.34	5.13	0.32	5.12
No. of store transactions	0.46	4.46	0.27	4.27	0.34	4.61	0.29	4.55
No. of online transactions	0.06	1.15	0.02	1.05	0.04	0.91	0.04	0.94
Total quantity	1.63	20.01	0.76	20.80	1.05	21.04	0.84	30.86
Store quantity	1.50	18.18	0.86	18.96	1.09	19.62	0.90	19.76
Online quantity	0.19	4.49	0.07	3.52	0.11	2.92	0.03	22.46
Total revenue (US$)	112.9	1219.2	67.79	1246.1	91.06	1275.2	81.74	1570.6
Store revenue (US$)	98.22	1123.8	60.65	1118.5	80.94	1159.1	72.87	1186.8
Online revenue (US$)	14.42	223.5	10.51	228.5	11.64	184.5	8.42	908.0
Return transactions								
Total No. of transactions	0.16	2.18	0.08	2.07	0.08	2.01	0.11	2.26
No. of store transactions	0.12	1.50	0.06	1.39	0.05	1.39	0.08	1.54
No. of online transactions	0.00	0.35	0.00	0.41	0.01	0.36	0.00	0.39
Total quantity	0.36	5.41	0.16	5.35	0.11	4.77	0.23	6.09
Store quantity	0.29	4.04	0.13	3.80	0.08	3.59	0.18	4.50
Online quantity	0.01	0.71	0.00	0.97	0.00	0.72	0.00	0.81
Total revenue (US$)	32.50	506.2	18.03	428.4	15.10	446.5	22.54	542.6
Store revenue (US$)	25.71	408.5	11.77	305.8	9.33	379.1	15.20	414.2
Online revenue (US$)	1.78	54.63	1.07	66.15	1.50	57.02	1.08	60.62
Nature of store and online purchase								
Store purchase diversity	-0.029	0.492	-0.037	0.493	-0.039	0.494	-0.031	0.497
Online purchase diversity	0.004	0.365	-0.011	0.373	-0.010	0.361	-0.002	0.349
Avg. store product price	-1.85	78.77	-3.16	74.70	-2.89	106.2	-2.41	78.52
Avg. online product price	2.12	37.79	0.900	39.87	1.21	38.98	1.53	37.94

To capture the qualitative aspects of purchases, we compute the diversity in purchases and the average product price of purchases made by the customers on the store and online channels in the prior-and post-periods. The retailer categorizes its products into over hundred categories based on the classification of products (such as apparel, accessories, handbags sunglasses, cosmetics, fragrance, beauty, and home products), the target gender for products (such as for men, women, junior, kids, and babies), and some specific brands of products. We computed the diversity of a customer's purchases on a channel in a period by the ratio of number of distinct categories of products purchased to the total number of purchased products by the customer on that channel in that period, hereafter referred to as purchase diversity. We computed the average price of purchased products by a customer on a channel in a period by the ratio of total purchase revenue to the total number of purchased products by the customers on that channel in that period, hereafter referred as average product price. We report the summary statistics for the change of these variables from prior- to post-period for the two groups of customers in each type of store openings in Table 2, which indicates a higher change (either higher increase or lesser decrease) in purchase diversity and average product price for the affected customers as compared to the unaffected customers in case of type-1 store openings but smaller difference in these values for the two groups of customers in case of type-2 store openings.

The differences in the mean values in Table 2 for the two groups of customers could be simply because of inherent differences in their purchase and return proclivities. In order to check this, we computed the cumulative purchase and return behaviors of each customer in our sample till three months before the date of store opening.[9] We computed the summary statistics of the cumulative total, store, and online purchases and returns for the two groups of customers and report it in Table 3a, which reveals similar mean values of these variables for affected and unaffected customers, both for type-1 and type-2 store openings. However, we find high standard deviations for these cumulative purchase and return variables from Table 3a, which points to a high dispersion in purchase and return proclivities within each group of customers. Therefore, to precisely infer the counterfactual purchase and return behaviors of an

[9] Note that we removed data for three months before and three months after the date of store opening to minimize the publicity effect around the store opening period.

affected customer, we need to find the unaffected customers with similar cumulative purchase and return behaviors.

Table 3a: Cumulative purchase/return behavior of customers

Variables	Type-1 store openings				Type-2 store openings			
	Affected customers		Unaffected customers		Affected customers		Unaffected customers	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Purchase transactions								
Total No. of transactions	12.39	21.13	12.73	21.11	12.97	21.19	14.83	23.03
No. of store transactions	10.35	19.24	10.53	19.26	11.02	19.47	12.73	20.94
No. of online transactions	0.48	1.58	0.51	1.66	0.43	1.46	0.46	1.67
Total quantity	39.99	84.91	40.85	87.12	41.92	83.06	48.40	94.60
Store quantity	34.78	78.67	35.25	80.76	36.90	78.02	42.88	85.22
Online quantity	1.12	4.74	1.24	5.57	1.04	4.75	1.19	22.85
Total revenue (US$)	1971.8	4463.1	2000.2	4792.44	2037.7	4711.5	2363.5	5108.6
Store revenue (US$)	1667.1	4086.4	1676.6	4370.5	1749.8	4407.2	2045.6	4589.7
Online revenue (US$)	60.94	257.7	66.36	290.1	55.97	240.3	64.09	951.3
Return transactions								
Total No. of transactions	2.91	8.13	3.01	8.22	2.94	7.45	3.47	8.94
No. of store transactions	1.60	5.21	1.64	5.23	1.66	4.79	1.98	5.69
No. of online transactions	0.08	0.44	0.09	0.57	0.08	0.51	0.08	0.53
Total quantity	5.34	19.00	5.51	19.68	5.33	16.51	6.43	22.35
Store quantity	3.08	12.69	3.16	13.07	3.14	11.45	3.85	14.85
Online quantity	0.13	0.91	0.16	1.38	0.13	1.07	0.14	1.12
Total revenue (US$)	3614	1325.8	372.8	1467.4	373.8	1548.4	437.2	1637.7
Store revenue (US$)	207.4	886.0	214.4	963.9	225.4	1219.7	262.6	1090.0
Online revenue (US$)	8.87	58.49	10.65	96.32	8.99	70.65	9.28	70.94

We also collected data on the age and income of the two groups of customers. The retailer categorizes its customers into six income categories [1 → < 50K, 2 → 50-75K, 3 → 75-100K, 4 → 100-150K, 5 → >150K, and 6 → unknown] and seven age categories [1 → < 25 years, 2→ 25-34 years, 3 → 35-44 years, 4 → 45-54 years, 5 → 55-64 years, 6 → >65 years, and 7 → unknown], where customers who do not reveal their income or age are assigned to the unknown category. We compared the proportions of customer households in each age and income category in the total across the samples of affected and unaffected customers in the two types of store openings and report it in Table 3b. From

Table 3b, we find a similar distribution of proportions of customers in different age and income categories for the two groups of customers in case of each type of store openings indicating a similar distribution of demographic characteristics for the two groups of customers. However, we find a wide variations in age and income across customers in each group in our data. As the purchase behavior of a 20 year old customer may be vastly different from the purchase behavior of a 70 years old customer, it is imperative to compare the purchase/return behaviors of an affected customer with the unaffected customers in the same age and income category.

Table 3b: Demographic characteristics of customers

Variables	Type-1 store openings				Type-2 store openings			
	Affected customers		Unaffected customers		Affected customers		Unaffected customers	
	Number	%	Number	%	Number	%	Number	%
Age Categories								
< 25 Years	352	3.96%	3113	3.67%	309	3.91%	3001	3.64%
25-34 Years	1534	17.28%	14140	16.65%	1399	17.71%	13532	16.40%
35-44 Years	1807	20.35%	17810	20.97%	1615	20.44%	17402	21.09%
45-54 Years	1552	17.48%	16030	18.88%	1405	17.78%	16160	19.58%
55-64 Years	1318	14.85%	13981	16.47%	1226	15.52%	13773	16.69%
>65 years	1703	19.18%	16164	19.04%	1597	20.22%	15339	18.59%
Unknown	612	6.89%	3673	4.33%	349	4.42%	3317	4.02%
Total	8878	100%	84911	100%	7900	100%	82524	100%
Household Income Categories								
< 50 K	2389	26.91%	21892	25.78%	2081	26.34%	21138	25.61%
50-75 K	1713	19.29%	16760	19.74%	1560	19.75%	16194	19.62%
75-100 K	1209	13.62%	11820	13.92%	1072	13.57%	11789	14.29%
100-150 K	1276	14.37%	12509	14.73%	1164	14.73%	12574	15.24%
>150 K	1227	13.82%	11668	13.74%	1061	13.43%	12044	14.59%
Unknown	1064	11.98%	10262	12.09%	962	12.18%	8785	10.65%
Total	8878	100%	84911	100%	7900	100%	82524	100%

4. Empirical Analysis and Results

In the previous section, we outlined the need for matching of affected and unaffected customers in our empirical design to find the causal effect of store openings on customers' purchase and return behaviors. In this section, we describe the estimation of *ATT* for customers' purchase and return behaviors from an inexact parametric matching (propensity score matching) method and an exact nonparametric matching (coarsened exact matching) method on our field data. Thereafter, we provide *ATT* estimates on online purchase behavior of different subsamples of affected customers based on their store interactions to provide empirical evidence of the mechanisms through which customers' store interactions affect their online purchase behavior.

4.1 Propensity Score Matching Estimates

The retailer's decision to open a store in an area is based on the characteristics of that area such as the average purchase behavior of customers in that area. Therefore, the probability of an affected customer receiving the treatment of store openings is based on the aggregate customer characteristics in that area.[10] However, in the following, we compute the probability of affected customers receiving the treatment of store openings on their individual characteristics and then estimate the *ATT* by matching affected and unaffected customers on these probability values. The rationale for this choice is that the unaffected customers with similar probability of receiving a treatment based on their individual characteristics more precisely captures the counterfactual behavior of affected customers than the unaffected customers with similar probability values based on their aggregate characteristics in a geographical area.

The propensity score is the Logit probability that a customer is an affected (treated) customer conditional on her demographic characteristics and pre-treatment cumulative purchase/return behaviors as below

> Propensity Score = Logit Probability (affected customer | *cumulative total, online, and store no. of purchase/return interactions, purchase/return quantities, and purchase/return revenue; age category of the head of household; income category of household; distance from the nearest store in the prior-period*),

An unaffected customer with similar propensity scores as an affected customer is likely to behave as the affected customer in absence of the treatment of store openings. Therefore, counterfactual

[10] Accordingly, we computed the propensity scores based on zip code level aggregate characteristics and then computed the *ATT* based on matching of affected and unaffected customers on such propensity scores. We found qualitatively similar results with such aggregate-level propensity scores matching than what are shown with individual-level propensity scores matching in the paper. The results are available on demand from the authors.

purchase/return behavior of the affected customers can be approximated by the purchase/return behavior of their propensity score matched unaffected customers and the difference in the purchase/return behaviors of these matched pair of customers would give the *ATT* of store opening on the purchase/return behavior of the affected customers. The *ATT* estimates shown in this paper are based on the nearest neighbor matching criterion, where each affected customer is matched with a fixed number of its nearest neighboring unaffected customers based on their propensity scores.[11] We also applied other propensity score matching criteria such as radius caliper matching. Mahalanobis distance matching, and kernel matching and found qualitatively similar results.

To match each affected customer with sufficient number of unaffected customers with similar propensity scores, we require reasonable overlap in the propensity score distributions for the affected and unaffected customers. This is called the *Overlap requirement*. Figure A1 in Appendix-A, we show that there is reasonable overlap in the propensity score distributions for the affected and unaffected customers for the two types of store openings indicating that the overlap requirement is satisfied in our case.

Table 4 reports the *ATT* based on specification (1) for change in number of transactions, quantities, and revenue for both purchase and return transactions on the store, online, and all channels from the prior- to post-period where matching of affected and unaffected customers is done based on their propensity scores.

From Table 4, we find a positive and significant estimate for *ATT* for change in the number of transactions, quantity, and revenue for purchase transactions from prior- to post-period for store, online, and on all channels for the affected customers in case of type-1 store openings but insignificant corresponding *ATT* estimates in case of type-2 store openings. This suggests that the customers' purchase on store, online, and on all channels increases in case of type-1 store openings but not in case of type-2 store openings. We further find a positive and significant *ATT* estimates for return transactions on the store channel and on all channels in case

[11] The number of nearest neighbors (unaffected customers) used for matching with an affected customer were determined based on the ratio of sample sizes of unaffected to affected customers. We also tried several other variations in the number of nearest neighbors and found qualitatively similar *ATT* estimates.

of type-1 store openings but not for type-2 store openings and insignificant *ATT* estimates for online return transactions in all cases. This suggests that with reduction in their store access costs, customers make more return transactions at the store channel.

Table 4: Average treatment effect on treated estimates

Change in Variables from Prior- to Post-period	Type-1 Store Openings				Type-2 Store Openings	
	ATT	t-value	Prior-period mean values	% increase	ATT	t-value
Purchase transactions						
Total No. of transactions	0.247***	3.85	3.58	6.90%	-0.013	0.2
No. of store transactions	0.171***	3.01	2.98	5.74%	0.007	0.12
No. of online transactions	0.047***	3.27	0.22	21.31%	0.008	0.68
Total quantity	0.739***	2.87	11.89	6.22%	0.099	0.37
Store quantity	0.538**	2.28	10.37	5.19%	0.129	0.52
Online quantity	0.111**	2.08	0.51	21.72%	0.002	0.05
Total revenue (US$)	44.39***	2.79	578.44	7.67%	6.14	0.38
Store revenue (US$)	37.34**	2.58	494.79	7.55%	8.03	0.54
Online revenue (US$)	4.00*	1.94	27.68	14.45%	0.017	0.01
Return transactions						
Total No. of transactions	0.078***	2.84	0.84	9.31%	0.037	1.32
No. of store transactions	0.048**	2.58	0.48	10.10%	-0.019	1.02
No. of online transactions	0.001	0.22	0.04		0	0.05
Total quantity	0.189***	2.75	1.56	12.09%	-0.133	1.05
Store quantity	0.127**	2.52	0.94	13.53%	-0.114	1.25
Online quantity	0.005	0.53	0.06		-0.001	0.09
Total revenue (US$)	15.24**	2.46	105.24	14.48%	-6.42	0.96
Store revenue (US$)	13.55***	2.66	62.33	21.74%	-4.58	0.84
Online revenue (US$)	1.20	0.33	4.13		0.528	0.67
Net transactions						
Total purchase quantity	0.573**	2.55	10.32	5.55%	0.208	0.88
Total revenue	33.31**	2.57	473.20	7.04%	11.79	0.88
Nature of purchase transactions						
Store purchase diversity	0.009*	1.95	0.358	2.51%	-0.004	0.57
Online purchase diversity	0.018***	3.69	0.100	17.99%	-0.002	0.40
Avg. store product price	1.12	1.10	36.31		-0.104	0.08
Avg. online product price (US$)	1.56***	3.13	8.29	18.82%	-0.022	0.04

***, **, * = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.
The percentage change values are only reported for statistically significant changes.

We further find positive and significant *ATT* estimates for net total purchase quantities and revenue for the affected customers in case of type-1 store openings but not for type-2 store openings. This suggests that the reduction in the store access costs results in higher net total purchases for the affected customers, i.e., an overall net benefit for the retailer. We use the prior-period mean values of purchase and return variables for the affected customers to translate the *ATT* estimates into the percentage increase values and report them in the fifth column of Table 4. These values indicate an economically significant percentage increase in the customers' total purchases and purchases on the store and online channels due to type-1 store openings. Specifically, the net total purchase quantities and revenues, respectively, increase by 5.55 and 7.04 percent due to type-1 store openings.

From Table 4, we also found a positive and significant *ATT* estimates for the purchase diversity on the store and online channels and positive and significant *ATT* estimate for the average product price on the online channel only. This suggests that easier store access with the store openings results in more diverse purchases on the store and online channels. This is in line with the argument that with easier store access, customers do higher search on the online and store channels that, in turn, exposes them to a wider product variety of the retailer. Moreover, the results of increased average product price on the online channel suggest that availability of a nearby store mitigates customers' risk of online purchases and thus encourages them to buy higher priced products online.

4.2 Coarsened Exact Matching (CEM) Estimators

Our *ATT* estimates are unbiased as long as the purchase and return behaviors of the unaffected customers provide a good counterfactual for the purchase and return behaviors of the affected customers sans store opening. Thus far, we have estimated the counterfactual purchase of return behaviors of affected customers from that of their propensity score matched unaffected customers. However, propensity score matching method suffers from the following criticism : (1) a pair of affected and unaffected customers, matched on a single measure of propensity scores, may have widely different values of individual variables, e.g., two customers having same propensity score value may have widely different ages and cumulative store purchase values, (2) matching is based on a parametric logit maximum likelihood model, and (3) researchers do not have ex-ante control on the extent of mismatch (measured as imbalance) between the samples of affected and unaffected customers and thus the error in *ATT* estimates.

To address these criticisms of the propensity score matching estimators, we used a nonparametric exact matching method, Coarsened Exact Matching (CEM) estimators, to match the sample of affected and unaffected customers on several different levels of imbalance and accordingly estimate the *ATT*. Unlike propensity score matching method that uses maximum likelihood estimators (a parametric model) to control for the differences in pretreatment variables across the treated and control groups, CEM is a nonparametric matching method that allows researchers to ex-ante bound the imbalance (both on individual variables and jointly) between the treated and control groups by manually coarsening the pretreatment variables into bins and thereby ex-ante control the error in *ATT* estimates (see Iacus et al. 2009 for details). If *ATT* estimates from several different options of manual coarsening of pretreatment variables remain qualitatively similar to that obtained from the propensity score matching, it provides clear evidence for the robustness of our causal estimates.

In CEM method, we compute the multivariate imbalance statistic ($\mathcal{L}_1$), which indicates the imbalance of the full multivariate histogram of pretreatment variables for the treated and control groups and includes their interactions and nonlinearities (Blackwell et al. 2009, Iacus et al. 2009). The $\mathcal{L}_1$ statistic value for our unmatched full sample of affected and unaffected customers in case of type-1 store openings was 0.97364. We first applied the automatic coarsening algorithm (available in the CEM package in STATA) that matches the affected and unaffected customers based on equal sized bins of pretreatment variables. With this matching, the $\mathcal{L}_1$ statistic value reduced to 0.958106. However, inspection of cutoff points for these bins of pretreatment variables (reported in Table 5) revealed a highly skewed allocation of cutoff points towards the higher values of variables. The width of a bin for a variable in automatic coarsening algorithm is determined by dividing the maximum value of the variable by the total number of bins. For example, in case of cumulative total purchase quantities, the size of the bin was computed as maximum value/number of bins (7204/17 = 423.8). Accordingly, the cutoff points of the 17 bins for cumulative total purchase quantities were (0, 423.8, 847.7, …….. 7402). However, the distribution of cumulative total purchase quantities in the sample was (10th percentile value = 2, 25th percentile value = 5, 50th percentile value = 14, 75th percentile value = 41, 90th percentile value = 101, and 99th percentile value

= 391). Therefore, we find that the automatic coarsening method bunches more than 99 percent of the customers in the first bin and divides less than one percent of high purchasing customers into the remaining 16 bins. We found similar patterns of bin formation in other variables as shown in the second column of Table 5.

Table 5: Cutoff points for variables from different coarsening methods

Variables used for matching	Cutoff points	
	Automatic coarsening (17 equal sized bins)	Manual coarsening
Cumulative purchase variables		
Total No. of transactions	(0, 39.9, 79.9, …….. 679)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000)
No. of store transactions	(0, 33.6, 67.2, ……….. 571)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000)
No. of online transactions	(0, 5.8, 11.5, …….. 98)	(0, 2, 5, 10, 20, 40, 75, 125)
Total quantity	(0, 423.8, 847.5, ……….. 7204)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000, 2000, 5000, 10000)
Store quantity	(0, 384.6, 769.2, ………. 6538)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000, 2000, 5000, 10000)
Online quantity	(0, 17.2, 34.5, …………. 293)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500)
Total revenue (US$)	(0, 31881.7, 63763.3, ………….. 541988.3)	(0, 50, 100, 250, 500, 1000, 2000, 5000, 10000, 25000, 50000, 100000, 600000)
Store revenue (US$)	(0, 27628.9, 55257.7, ………….. 469690.5)	(0, 50, 100, 250, 500, 1000, 2000, 5000, 10000, 25000, 50000, 100000, 600000)
Online revenue (US$)	(0, 1045.2, 2090.4, ……….. 17768.5)	(0, 50, 100, 250, 500, 1000, 2000, 5000, 10000)
Cumulative return variables		
Total No. of transactions	(0, 23.9, 47.9, ………….. 407)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500)
No. of store transactions	(0, 17.2, 34.4, …………. 293)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500)
No. of online transactions	(0, 2.2, 4.4, ……………. 37)	(0, 2, 5, 10, 20, 40, 75)
Total quantity	(0, 70.5, 141.1, ………….. 1199)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000, 2000, 5000)
Store quantity	(0, 41.2, 82.4, ……………. 700)	(0, 2, 5, 10, 20, 40, 75, 125, 250, 500, 1000)
Online quantity	(0, 5.9, 11.9, …………. 101)	(0, 2, 5, 10, 20, 40, 75, 125)
Total revenue (US$)	(0, 9235.1, 18470.3, ……………… 156997.1)	(0, 50, 100, 250, 500, 1000, 2000, 5000, 10000, 25000, 50000, 100000, 600000)
Store revenue (US$)	(0, 4549.5, 9099.0, ……………… 77341.5)	(0, 50, 100, 250, 500, 1000, 2000, 5000, 10000, 25000, 50000, 100000)
Online revenue (US$)	(0, 839.7, 1679.4, ………………. 14274.8)	(0, 50, 100, 250, 500, 1000, 2000, 5000)
Demographic variables		
Income category	(1, 1.29, 1.59, ………... 6)	(0, 1, 2, 3, 4, 5, 6)
Head of household age category	(1, 1.35, 1.71, ……….. 7)	(0, 1, 2, 3, 4, 5, 6, 7)

Distance from store in pre-period (kilometers)	(62.8, 81.2, 99.5, ………………….. 374.8)	(60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 375)

To correctly create the bins for matching variables in our context, we held discussions with the representatives of the retailer to understand what ranges of different variables would indicate similar customer purchase/return behavior. For instance, cumulative store purchase transactions of less than 2 would indicate occasional customers, between 2 to 5 store purchase transactions would indicate the next category of customers and so on. Accordingly, we manually created unequal sized bins for each variable in our data such that (1) the range of variable values in a bin indicate similar customers' purchase/return behavior and (2) the cutoff points for bins of a variable are distributed, as far as possible, over all quartiles of distribution of that variable. We report the cutoffs of manually created bins of variables in the third column of Table 5 to show their comparison with corresponding bins created from automatic coarsening method.

Table 6: Comparison of CEM and Propensity Score matching estimators

	Propensity score matching		Automatic Coarsening		Manual Coarsening	
Number of matched (total) affected customers	Not applicable		8456 (8878)		3388 (8878)	
Number of matched (total) unaffected customers	Not applicable		73931 (84911)		10782 (84911)	
Overall imbalance ($\mathcal{L}_1$ statistics)	Not applicable		0.958106		0.60279	
Change in Variables from Pre- to Post-period	ATT	t-value	ATT	t-value	ATT	t-value
Purchase transactions						
Total No. of transactions	0.247***	3.85	0.263***	5.27	0.239***	3.53
No. of store transactions	0.171***	3.01	0.198***	4.45	0.183***	2.96
No. of online transactions	0.047***	3.27	0.045***	4.92	0.039***	2.91
Total quantity	0.739***	2.87	0.867***	4.5	0.792***	3.45
Store quantity	0.538**	2.28	0.686***	3.82	0.643***	2.98
Online quantity	0.111**	2.08	0.106***	4.06	0.089**	2.45
Total revenue (US$)	44.39***	2.79	45.98***	4.05	53.11***	4.56
Store revenue (US$)	37.34**	2.58	36.87***	3.53	45.89***	4.26
Online revenue (US$)	4.00*	1.94	6.07***	3.61	4.62*	1.88
Return transactions						
Total No. of transactions	0.078***	2.84	0.074***	3.86	0.051***	2.60
No. of store transactions	0.048**	2.58	0.043***	3.64	0.026**	1.98
No. of online transactions	0.001	0.22	0.003	0.99	0.002	0.45
Total quantity	0.189***	2.75	0.137***	3.24	0.071*	1.79

Store quantity	0.127^{**}	2.52	0.093^{***}	3.05	0.023^{**}	1.97
Online quantity	0.005	0.53	0.007	1.04	0.008	1.02
Total revenue (US$)	15.24^{**}	2.46	8.18^{**}	2.22	10.38^{***}	3.53
Store revenue (US$)	13.55^{***}	2.66	6.53^{**}	2.37	4.39^{**}	1.93
Online revenue (US$)	1.20	0.33	0.706	1.31	0.531	0.7
Net transactions						
Total purchase quantity	0.573^{**}	2.55	0.73^{***}	4.23	0.721^{***}	3.42
Total revenue	33.31^{**}	2.57	37.79^{***}	3.89	42.73^{***}	4.11
Nature of purchase transactions						
Store purchase diversity	0.009^{*}	1.95	0.01^{*}	1.66	0.012^{**}	2.12
Online purchase diversity	0.018^{***}	3.69	0.02^{***}	4.23	0.022^{***}	3.14
Avg. store product price (US$)	1.12	1.10	1.57^{*}	1.92	2.24^{*}	1.79
Avg. online product price (UD$)	1.56^{***}	3.13	1.59^{***}	3.53	1.48^{**}	2.25

$^{***, **, *}$ = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

Based on such manually created bins, we computed the overall imbalance $\mathcal{L}_1$ statistics value and the number of matched treated and control customers out of the total customers in each category and report it in Table 6. First, we find that manual coarsening results in lesser number of matched customers in each category and a significantly lower overall imbalance $\mathcal{L}_1$ statistic value (0.60279 as compared to 0.958106 in automatic coarsening).[12] This suggests that the manual coarsening of variables resulted in a higher extent of exact matching between the affected and unaffected customers and thus more precise estimation of *ATT*. We computed the *ATT* for change in purchase and return variables from prior- to post-period for case of type-1 store openings based on the bins for different variables created from the automatic and manual coarsening and report it in Table 6 along with the *ATT* estimates from the propensity score matching. We found qualitatively similar estimates of *ATT* from the two coarsening methods that suggest that sign and significance of our *ATT* estimates are robust to the variations in the overall imbalance of the treated and control customers.[13] We further find that the *ATT* estimates from the propensity score matching methods are qualitatively similar to their corresponding values obtained from CEM methods. This reassures us of the validity of out propensity score matching estimates.

4.3 Differential Impact of Store Openings for Far and Near Customers

[12] A large percentage of total customers were matched in each group the automatic coarsening because more than 90 percent of all customers were bunched in the first bin for all variables. When we corrected this bin formation in our manual coarsening, a much higher percentage of customers were pruned due to lack of match and we obtained smaller number of matched customers in each group and consequently a smaller imbalance between the two groups of customers.

[13] We also tried several other variations of manual coarsening of variable bins and find qualitatively similar results.

Next, we analyzed the variations in the effect of store openings across the affected customers based on their distance from the nearest store after the store opening. We compared the change in purchase/return variables for the two subcategories of affected customers – customers in first subcategory have their post-period distances from the nearest store in the first (lowest) quartile (hereafter near customers) and the customers in second subcategory have their post-period distances in the fourth (highest) quartile of distribution of distance values (hereafter far customers). As per Table 1, the distance from the nearest store for near and far customers in case of type-1 store openings is less than 10.5 kilometers and more than 94.9 kilometers, respectively, and in case of type-2 store openings is less than 7.4 kilometers and more than 23 kilometers, respectively. Once again, before computing the *ATT* we ensured that the overlap assumption is satisfied (see Figure A2 in Appendix A).

In Table 7, we report the propensity score based *ATT* estimates for near and far customers for the two types of store openings.[14] The results in Table 7 reveals a significant effect of store openings on the store and online purchase behaviors of near customers but insignificant effect for far customers in case of type-1 store openings. We also find insignificant store opening effects in case of type-2 store openings for both near and far customers, which may be due to the fact that the change in distance from the nearest store is relatively insignificant for customers in type-2 store openings. For return transactions, we find that the store returns increase significantly for the near customers but not for far customers in case of type-1 store openings. The online returns, however, remain statistically similar for both subcategories of customers in case of type-1 store openings. All returns remain statistically similar for both subcategories of customers in case of type-2 store openings. Overall, we find that the net total purchase quantity and revenue significantly increase for the near customers due to type-1 store openings. We also find that the purchase diversity and average product price on the online channel increase only for near customers in case of type-1 store openings, which indicates that easier store access causes the change in nature of

[14] The CEM based *ATT* estimates for far and near customers are qualitatively similar and are available on demand from the authors.

online purchases. These results provide further evidence that the observed change in customers' purchase and return behaviors are driven by the reduction in their store access costs.

Table 7: Variations in *ATT* with post-period distance from nearest store

Change in Variables from Pre- to Post-period	Type-1 Store Openings				Type-2 Store Openings			
	First quartile (2234)		Fourth quartile (2169)		First quartile (2011)		Fourth quartile (1906)	
	ATT	t-value	*ATT*	t-value	*ATT*	t-value	*ATT*	t-value
Purchase transactions								
Total No. of transactions	0.479***	3.65	0.158	1.31	-0.23*	1.75	0.225*	1.71
No. of store transactions	0.328***	2.87	0.131	1.22	-0.20*	1.68	0.228	1.65
No. of online transactions	0.111***	3.08	0.016	0.65	0.00	0.02	0.013	0.56
Total quantity	1.57***	2.84	0.507	1.02	-0.405	0.75	-0.009	0.02
Store quantity	1.12**	2.26	0.396	0.89	-0.356	0.71	0.018	0.04
Online quantity	0.376**	2.30	-0.01	0.14	-0.019	0.21	0.028	0.41
Total revenue (US$)	61.27**	1.98	35.09	1.16	9.99	0.30	-1.69	0.05
Store revenue (US$)	46.34**	1.96	30.87	1.14	3.78	0.12	8.75	0.25
Online revenue (US$)	16.05**	2.13	-2.19	0.43	3.22	0.60	-1.77	0.43
Return transactions								
Total No. of transactions	0.137**	2.23	0.038	0.68	0.052	0.99	-0.048	0.89
No. of store transactions	0.089**	2.05	0.054	1.48	-0.008	0.21	-0.036	0.97
No. of online transactions	0.012	1.38	-0.013	1.5	0.008	0.77	-0.003	0.26
Total quantity	0.331**	2.17	0.199	1.36	-0.184	1.37	-0.20	1.65
Store quantity	0.217**	1.96	0.196	1.65	-0.114	1.12	-0.121	1.35
Online quantity	0.033	1.51	-0.022	1.11	-0.009	0.44	-0.003	0.16
Total revenue (US$)	12.87*	1.71	12.24	1.06	-9.68	0.65	-9.23	0.83
Store revenue (US$)	8.89*	1.92	14.42	1.55	-3.49	0.28	-2.57	0.28
Online revenue (US$)	0.789	0.55	-1.76	1.25	-0.286	0.19	0.528	0.31
Net Transactions								
Total purchase quantity	1.32***	2.79	0.34	0.78	-0.248	0.52	0.126	0.27
Total revenue	51.27**	1.97	27.89	1.12	15.35	0.55	7.45	0.27
Nature of purchase transactions								
Store purchase diversity	0.03**	2.50	0.004	0.34	0.003	0.23	-0.01	0.80
Online purchase diversity	0.027***	2.89	0.007	0.76	-0.014	1.31	-0.002	0.16
Avg. store product price (US$)	5.00***	2.80	-0.543	0.30	0.965	0.59	-1.69	0.43
Avg. online product price (US$)	2.28***	3.53	-0.002	0.01	-1.18	1.31	-0.706	0.65

***, **, * = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

4.4 Variations in Customers' Online Purchase Behavior with their Store Visits

Thus far, we have shown that the reduction in customers' store access costs due to opening of a store near them affects the quantity and nature of their online purchases besides, as expected, affecting (increasing) their store purchases. In our conceptual framework, we proposed two mechanisms, *information effect* and *return effect*, through which higher customers' store visits due to easier store access may affect the quantity and nature of their online purchases. In our data, we observe only the store visits in which customers either purchase or return products.[15] A higher store purchase transactions in the post period points to customers receiving the *information effect* and a higher store return transactions in the post period indicates customers getting the *return effect*. Thus, we provide empirical evidence of these mechanisms by showing that the quantity, diversity, and average product price of online purchases are increased for only those affected customers who make higher store visits, either for purchase or return transactions, in the post-period as compared to prior-period.

Table 8: *ATT* estimates for online purchase behavior with change in store purchase and returns

Change in Online Purchase Variables from Pre- to Post-period	More post-period store purchase transactions		Less or equal post-period store purchase transactions		More post-period store return transactions		Less or equal post-period store return transactions	
	ATT	*t-value*	*ATT*	*t-value*	*ATT*	*t-value*	*ATT*	*t-value*
All Customers								
No. of affected customers	3428		5450		1474		7404	
No. of transactions	0.119***	6.81	-0.011	0.62	0.212***	4.44	0.011	0.80
Purchase quantity	0.284***	5.17	0.009	0.11	0.602***	2.60	0.035	0.85
Purchase revenue (US$)	14.66***	4.14	-3.61	1.01	26.75***	2.82	0.739	0.28
Purchase diversity	0.053***	8.71	-0.004	0.69	0.068***	7.10	0.005	0.93
Avg. product price (US$)	4.65***	7.06	-0.799	1.33	6.20***	6.09	0.105	0.2
Near Customers								
No. of affected customers	901		1333		406		1828	
No. of transactions	0.138***	3.86	0.054	1.02	0.342**	2.32	0.049*	1.80
Purchase quantity	0.313***	2.64	0.321	1.26	1.29*	1.93	0.128	1.52
Purchase revenue (US$)	14.29*	1.93	12.29	1.21	47.66**	1.98	7.74	1.52
Purchase diversity	0.057***	4.64	-0.003	0.29	0.09***	4.86	0.013	1.27
Avg. product price (US$)	4.95***	3.81	0.087	0.07	7.22***	3.74	1.31	1.28

***, **, * = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

[15] Customers may make visits the retailer's store just to check out products without either purchasing or returning products. Such visits may have the information effect on customers but are not observed in our data.

We divided the samples of all affected customers in case of type-1 store openings into two subcategories, customers in the first subcategory made more and in the second subcategory made less or equal store purchase/return transactions in the post-period as compared to the prior-period. Once again, before computing the *ATT* we ensured that the overlap assumptions are satisfied in each case (see Figure A3 and A4 in Appendix A). Then, we separately estimated the *ATT* for online purchase behavior for these subcategories of affected customers with their propensity score matched samples of unaffected customers and report it in Table 8.[16] The top half of Table 8 indicates positive and significant *ATT* estimates for quantitative online purchase variables (number of transactions, quantity, and revenue) as well as qualitative online purchase variables (diversity and average product price) only for the first subcategory of all customers based on purchase and returns, i.e., customers who make higher store purchase or return transactions in the post-period as compared to the prior-period.

In the previous section, we showed that the store openings affect the online purchase behavior of only near customers, as the store access costs reduced significantly for near customers only. Now, we show that even among near customers, the online purchase behavior changes only for those near customers who make more store transactions in the post-period as compared to the prior-period. To show this, we performed the analysis described in the previous paragraph on the sample of near customers and report the *ATT* estimates in the bottom half of Table 8. We find positive and significant *ATT* estimates for only first subcategory of near customers who make more store transactions, either for purchase or return, in the post-period. These results suggest that the reduction in customers' store access costs does not by itself affect their online purchases but higher store visits, either for purchase or for return transactions, are the main cause for the change in quantity and nature of their online purchases.

Customers who make higher purchase transactions are also likely to make higher return transactions at the store channel. Therefore, the results shown in Table 8 could be driven by only one of these effects and does not necessarily mean that the information and return effects separately influence

[16] The corresponding CEM based *ATT* estimates are qualitatively similar and are available on demand from the authors.

customers' online purchases. To separately tease out the impact of these two effects, we divided the full sample of affected customers into following four subsamples based on whether they make more/(less or equal) purchase/return transactions at the store channel in the post-period as compared to the prior-period: (1) those who make more store purchase and return transactions, (2) those who make more store purchase transactions but less or equal store return transactions, (3) those who make more store return transactions but less or equal store purchase transactions, and (4) those who make less or equal store purchase and return transactions in the post-period. Then, we separately estimated the *ATT* for online purchase behavior for these subsamples of affected customers with their propensity score matched samples of unaffected customers and report it in Table 9.[17]

Table 9: *ATT* estimates for online purchases with interaction of change in store purchase and returns

	Change in Online Purchase Variables from Pre- to Post-period	More post-period store purchase transactions		Less or equal post-period store purchase transactions	
		ATT	*t-value*	*ATT*	*t-value*
More post-period store return transactions	No. of affected customers	1084		390	
	No. of transactions	0.153***	4.67	0.382**	2.44
	Purchase quantity	0.223**	2.17	1.59*	1.91
	Purchase revenue (US$)	11.11**	2.12	68.68**	2.19
	Purchase diversity	0.072***	6.65	0.048**	2.54
	Avg. product price (US$)	6.00***	5.08	6.95***	3.39
Less or equal post-period store return transactions	No. of affected customers	2344		5081	
	No. of transactions	0.086***	4.13	-0.046**	2.55
	Purchase quantity	0.219***	3.38	-0.115**	2.39
	Purchase revenue (US$)	10.64**	2.47	-7.09**	2.39
	Purchase diversity	0.041***	5.70	-0.011*	1.88
	Avg. product price (US$)	3.46***	4.42	-1.32**	2.12

***, **, * = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

From Table 9, we find positive and significant *ATT* estimates for quantitative online purchase variables (number of transactions, quantities, and revenue) and qualitative online purchase variables (purchase diversity and average product price) for only customers in the first three subsamples but find a negative and significant *ATT* estimates for these variables for customers in the fourth subsample. This provides empirical support to the fact that both effects, information effect captured by higher store

[17] The CEM based *ATT* estimates are qualitatively similar and are available on demand from the authors.

purchase transactions and return effect captured by higher store return transactions, separately affect the quantity and nature of customers' online purchases.

To summarize, in this section we show that the reduction in customers' distance from the nearest store does not by itself leads to the increase in their online purchases. But when such reduction in distance leads to information and return effects due to their higher store visits, customers make higher quantities, higher priced, and more diverse purchases on the online channel.

4.5 Other Robustness Checks

Our positive and significant *ATT* estimates from our diff-in-diff design may come due to differential pre-existing trends in purchase behaviors between the affected and unaffected customers. For example, if the retailer opened the store in an area based on the increasing trends in purchase behaviors of the customers living around that area, our diff-in-diff design will simply capture the higher difference in purchases of affected customers from prior- to post-period as compared to the corresponding value for unaffected customers due to increasing pre-existing trends in purchase behaviors on affected customers and not due to store openings. To control for the differential trends in purchase behavior of the two groups of customers, we extended our experimental design to diff-in-diff-in-diff design as described in Appendix B. In this analysis, we control for the differential pre-existing trends in purchases between the affected and unaffected customers to find the change in purchases caused by store openings. We found qualitatively similar *ATT* estimates (in magnitude, sign, and significance) for the effect of store openings from this analysis indicating that our estimates are robust to the possibility differential trends in the purchase behaviors of two groups of customers.

Thus far we have analyzed the effect of store openings based on the data pertaining to two type-1 stores and two type-2 stores opened by the retailer in fall 2003. One may argue that our estimated effects may be limited to something specific to either the time period of fall 2003 or the stores opened in that duration. To address this concern, we performed the whole analysis on data pertaining to a type-1 store openings by the retailer in spring 2004. The results are reported in Appendix C. We find that the effect of spring 2004 store openings is very similar (in magnitude and significance) to the effect of fall 2003 store openings.

5. Conclusions and Future Work

We designed a quasi-natural experiment and used customer-level data to estimate the causal effect of store openings by a fashion retailer in the US on the purchase and return behavior of its existing customers. We showed that reduction in customers' store access costs due to store openings increases not only their store purchases and returns but also their online purchases. We also found that easier store access with the store openings results in customers purchasing more diverse products on the store and online channels and higher priced products on the online channel. Drawing from a diverse literature on customer multichannel behavior, we argue that higher store visits may affect the quantity and nature of customers' online purchases by two mechanisms, information effect and return effect. We provide empirical evidence of these effects by showing that the quantity, diversity, and average product price of online purchases increase for only those customers, who make higher number of store transactions, either for purchase or return, after the store openings. Overall, we show that easier store access leads to an increase in customers' net total purchases with the retailer.

The main findings of this paper is how facilitating store access increases customers' store transactions, either for purchase or returns, which, in turn, boosts their store and online purchases. The managers can utilize this insight to create events to attract higher store traffic and design customer friendly return policies to allow them to return products at low costs on any channel of their choice. These findings also inform managers about the benefits of integrating the capabilities of the online and store channels to meet the customer needs at different stages of the purchase process – information acquisition, consideration set formation, purchase, and post-purchase transactions. Moreover, the result of increased customers' online purchases by facilitating their store access informs managers of a way to make them multichannel and thus increase their value for the retailer. Overall, findings in the paper inform managers of the preeminence of the store channel, at least in the fashion apparel industry, in driving sales on the online channel. Therefore, there is a need to build up synergies across the store and online channels rather than treating them as competing channels of sales.

Our paper has few limitations that offers opportunity for addressing them in future research. First, we assumed that the counterfactual purchase and return behavior of the affected customers can be estimated by the purchase and return behavior of the unaffected customers matched on observable demographic characteristics and past purchase and return behavior. Although, it is a reasonable assumption in our context but there is an outside possibility that the two groups of customers are systematically different on unobserved characteristics that may differentially influence their purchase and return behaviors. Second, the results in this paper are obtained for sensory products such as fashion apparel, accessories and home products and they may not be generalizable to other products. However, our findings remain important for the business world, as these product categories comprise an economically significant portion of present Ecommerce. Third, the effect of customers' store transactions on their online purchases from the same retailer is dependent on the extent of competition between the online and multichannel retailers. If there is a fierce price competition between retailers in a product category, the showrooming effect of a multichannel retailer may result in customers purchasing more from the online channel of a pure online retailer or of other multichannel retailers. The estimated effects in the present paper could also be due to higher differentiation or branding effect of the retailer and may not be generalizable in other competitive settings. Further research is required to study the impact of store openings by retailers in different product categories and different competitive settings on the online sales of the multichannel retailers and the pure-play online retailers. Fourth, this paper examines the causal effect of facilitating customers' access on store channel on their purchase and return behaviors on store and online channels. It will interesting to examine the effect in opposite direction, i.e., the effect of facilitating customers' access to the online channel, such as by reducing shipping fee, on their store and online purchase behavior.

References

Anderson Eric T., K Hansen, D. Simester (2009). The option value of return: Theory and Empirical Evidence. *Marketing Science* 28(3) 405-423.

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.

Bell, D., S. Gallino, and A. Moreno (2013). Inventory Showrooms and Customer Migration in Omni Channel Retail: The Effect of Product Information. *Working Paper*.

Biyalogorsky, Eyal and Prasad Naik (2003), "Clicks and Mortar: The Effect of Online Activities on Offline Sales," *Marketing Letters*, 14 (1), 21–32.

Blackwell, M., S. M. Iacus, G. king, G. Porro (2009). CEM: Coarsened Exact Matching in Stata. *The Stata Journal* 9(1) 524-546

Brynjolfsson, E., Y. Hu, and M. S. Rehman. (2009). Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition. *Management Science* 55(11) 1755-1765.

Cook, Alan (2014). "With Sales Falling at Department Stores, Nordstrom Plans Bigger Online Push". Geekwire, February 22, 2014 (http://www.geekwire.com/2014/sales-falling-department-stores-nordstrom-plans-bigger-online-push/)

Danaher, P. J., I. W. Wilson, and R. A. Davis (2003). A Comparison of Online and Offline Consumer Brand Loyalty. *Marketing Science* 22(4) 461-476.

Deleersnyder, B., I. Geyskens, K. Gielens, and M. G. Dekimpe (2002). How Cannibalistic is the Internet Channel? A Study of the Newspaper Industry in the United Kingdom and the Netherlands. *International Journal of Research in Marketing* 19(4) 337-348.

Dholakia, U. M., B. E. Kahn, R. Reeves, A. Rindfleisch, D. Sewart and E. Taylor. (2010). Consumer Behavior in a Multichannel, Multimedia Retailing Environment. *Journal of Interactive Marketing* 24 86-95.

Forman, C., A. Ghose, and A. Goldfarb (2009). Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live. *Management Science* 55(1) 47-57.

Gensler, S., P. C. Verhoef and M. Bohm. (2012). Understanding Consumer's Multichannel Choices across the Different Stages of the Buying Process. *Marketing Letters* 23 987-1003.

Geykens, I., K. Gielens, and M. G. Dekimpe (2002). The Market Valuation of Internet Channel Additions. *Journal of Marketing* 66(2) 102-119.

Iacus, S. M., G. King, G. Porro (2009). CEM: Software for Coarsened Exact Matching. *Journal of Statistical Software* 30(9) 1-27

Inman, J. J., V. Shankar, and R. Ferraro (2004). The Roles of Channel-Category Associations and Geodemographics in Channel Patronage Decision. *Journal of Marketing* 68(2) 51-71.

Knox, G. A. H. (2007). Modeling and Managing Customers in a Multichannel Setting. Unpublished Dissertation. University of Pennsylvania.

Lynch, P. D., R. J. Kent and S. S. Srinivasan. (2001). The Global Internet Shopper: Evidence from Shopping Tasks in Twelve Countries. *Journal of Advertising Research.* May-June 15-23.

McIntyre, D. A., A. E. M. Hess (2014). “Nine Retailers Closing the Most Stores”. 24/7 Wall Street, March 12, 2014 (http://www.usatoday.com/story/money/business/2014/03/12/retailers-store-closings/6333865/).

Neslin, S. A. and V. Shankar. (2009). Key Issues in Multichannel Customer Management: Current Knowledge and Future Directions. *Journal of Interactive Marketing* 23 70-81.

Neslin, S.A., D. Grewal, R. Leghorn, V. Shankar, M. L. Teerling, J. S. Thomas, and P. C. Verhoef (2006). Challenges and Opportunities in Multichannel Customer Management. *Journal of Service Research* 9(2) 95-112.

Pauwels, K., P. S. H. Leeflang, M. L. Teerling and K. R. E. Huizingh. (2011). Does Online Information Drive Offline Revenues? Only for Specific Products and Consumer Segments! *Journal of Retailing* 87(1) 1-17.

Ratchford, B. T., M. Lee and D. Talukdar. The Impact of the Internet on Information Search for Automobiles. *Journal of Marketing Research*. 40 (2) 193-209.

Shankar, V., A. K. Smith, and A. Rangaswamy (2003). The Relationship between Customer Satisfaction and Loyalty in Online and Offline Environments. *International Journal of Research in Marketing* 20(2) 153-175.

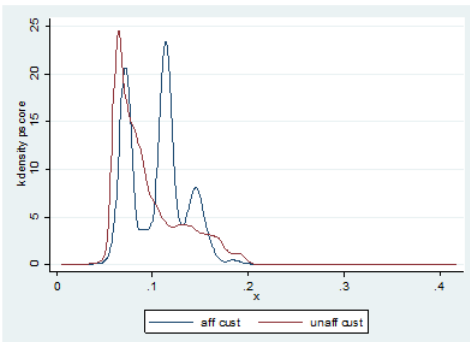
Verhoef, P. C., S. A. Neslin and B. Vroomen. (2007). Multichannel Customer Management: Understanding the Research-Shopper Phenomenon. *International Journal of Research in Marketing*. 24 129-148.

Wallace, D. W. Gliese, J. L. and J. L. Johnson. (2004). Customer Retailer Loyalty in the Context of Multiple Channel Strategies. *Journal of Retailing* 80(4) 249-263.

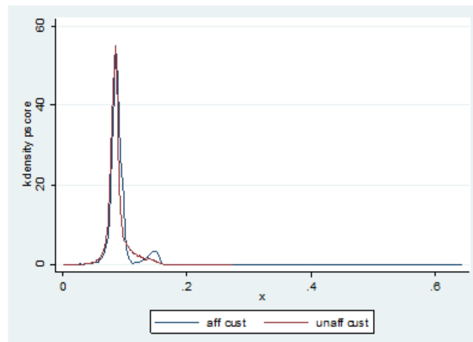
ONLINE APPENDIX

Appendix A: Testing for Overlap assumption in propensity score matching

In this appendix, we show the propensity score distributions for the samples of affected and unaffected customers in different analyses conducted in this paper. The objective is to visually show that the overlap assumption for estimating treatment effect is satisfied in these cases.

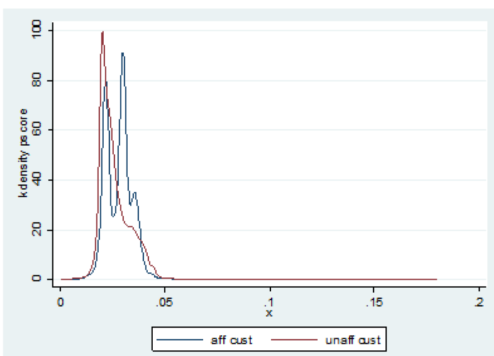


Type-1 store opening

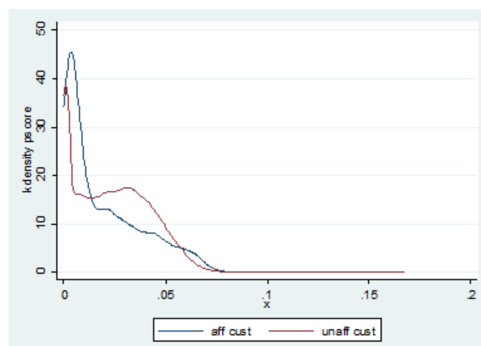


Type-2 store opening

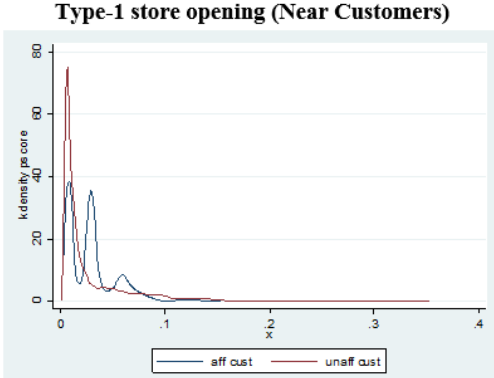
Figure A1: Propensity score distribution for Full sample of customers



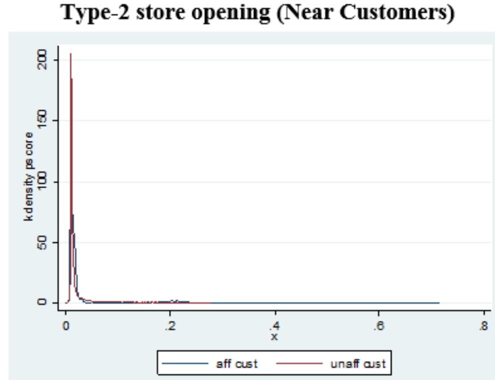
Type-1 store opening (Near Customers)



Type-2 store opening (Near Customers)

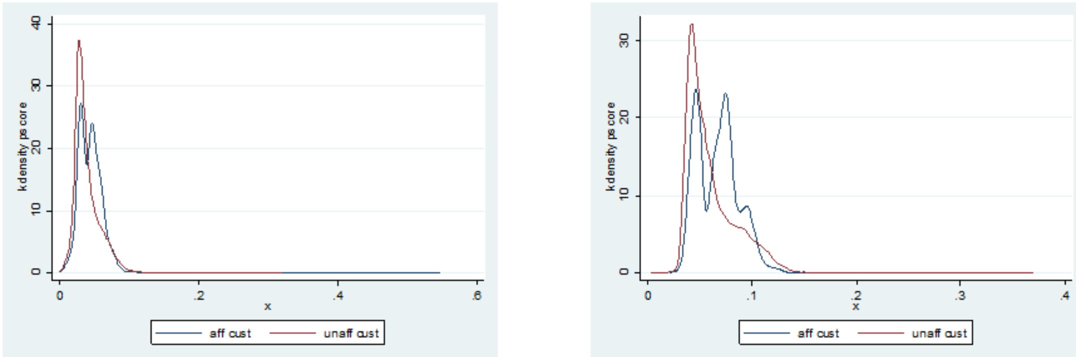


Type-1 store opening (Far Customers)



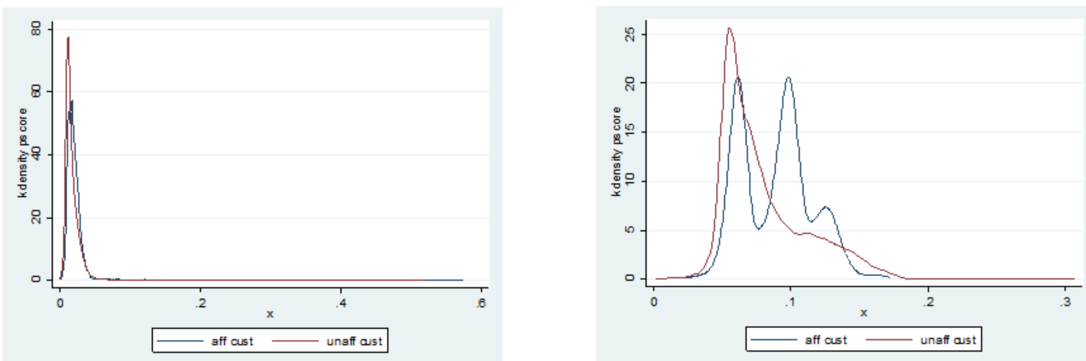
Type-2 store opening (Far Customers)

Figure A2: Propensity score distribution for near and far customers



More post period store purchase **Less or equal post period store purchase**

Figure A3: Propensity score distribution for all unaffected customers and affected customers who make more/less or equal store purchase transactions in the post-period



More post period store returns **Less or equal post period store returns**

Figure A4: Propensity score distribution for all unaffected customers and affected customers who make more/less or equal store return transactions in the post-period

Appendix B: Controlling Differential Time Trends with Doubly Robust Estimator

We may get a positive and significant *ATT* in our diff-in-diff design merely due to a higher time trends in purchase behavior of the affected customers as compared to the unaffected customers in the period prior to store opening. This may also be a reasons due to which the retailer opened a store in the area near affected customers. To control for differences in the rate of change in purchase behaviors with time between the two groups of customers, we add data on purchase and return behaviors for the two groups of customers two year prior to the store openings in our existing analysis and run the following weighted least square specification with diff-in-diff-in-diff experimental design

$$\Delta Y_{it} = \beta_0 + \beta_1\, Treat_i + \beta_2\, Post_t + \beta_3\, Treat \times Post_{it} + \varepsilon_{it}\ , \qquad \text{------------- (B1)}$$

where, i denotes the customers and t denotes the two time period changes: from two year prior to one year prior, and one year prior to one year after the store openings. *Post* is an indicator variable equal to one for the second time period change and zero for the first time period change, *Treat* is an indicator variable equal to one for affected customers and is zero otherwise, and *Treat x Post* is the interaction of these two variables. The dependent variable ΔY in specification (BI) denotes the change in purchase and return variables for the customers in two time period changes. We assign weights to the purchase and return behavior of a customer (i) based on her propensity score $p(X_i)$ such that $Weight_i = [\{Treat_i / p(X_i)\} + \{(1-Treat_i)/ (1- p(X_i))\}]$, and X_i denote the observed pretreatment control variables of customer (i) used for computing the propensity scores. Such OLS specifications with inverse probability weighting on propensity scores is referred to as the Doubly Robust Estimators and is widely used in the economics literature (Hirano and Imbens 2001).

Table B1: Doubly robust treatment effect estimates

Change in Variables in a time period change (ΔY)	Doubly Robust Estimator			
	β_3	t-value	β_3	t-value
	Purchase transactions		**Return transactions**	
Total No. of transactions	0.186^{***}	7.31	0.06^{***}	5.48
No. of store transactions	0.137^{***}	6.09	0.042^{***}	5.62
No. of online transactions	0.039^{***}	6.85	0.004^{*}	1.85
Total quantities	0.573^{***}	5.37	0.186^{***}	6.71
Store quantities	0.411^{***}	4.21	0.122^{***}	6.00
Online quantities	0.104^{**}	4.97	0.016^{***}	3.47
Total revenue (US$)	34.29^{***}	5.48	12.33^{***}	5.30
Store revenue (US$)	26.69^{***}	5.20	11.34^{***}	6.31
Online revenue (US$)	3.21^{***}	2.89	0.929^{***}	3.02
	Net transactions			
Net total quantity purchased	0.387^{***}		4.12	
Net total revenue	21.97^{***}		4.21	
	Nature of purchase transactions			
Store purchase diversity	0.009		0.51	
Online purchase diversity	0.019^{***}		3.14	
Avg. store purchase price	-0.437		1.00	
Avg. online purchase price	1.86^{***}		8.80	

$^{***, **, *}$ = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

In specification (BI), coefficient β_1 and β_2, respectively, capture the differential time trends for the affected customers as compared to unaffected customers and the second time period change as compared to first time period change. Our coefficient of interest, β_3, captures the treatment effect of store openings on the change in purchase and return behaviors of affected customers after controlling for the differential time trends in purchase behaviors for the two groups of customers and the two time periods.

We report the estimates of β_3 for different purchase and return variables in Table B1. Table B1 reveals qualitatively similar values (in sign, and significance) for β_3 as compared to the corresponding *ATT* estimates obtained from the two matching methods. This provides further credibility to the robustness of our results.

Appendix C: Results for a different Store Openings in Spring 2004

In this appendix we show the *ATT* estimates for another store opened by the retailer in spring 2004. If the *ATT* estimates obtained from a different store opened at different time period are similar to the results of our main analysis, it would indicate that our results are generalizable for any store openings.

Table C1: Distribution of distance from the nearest store

Distance from the nearest store in kilometers	Affected customers (5696)		Unaffected customers (48462)
	Prior-period	Post-period	Prior/post period
Mean	194.7	60.8	184.7
Std. Dev.	35.5	52.0	44.70
1 percentile	118.2	0.0	121.4
25 percentile	178.3	12.8	147.2
50 percentile	190.8	35.8	175.1
75 percentile	210.6	116.4	217.9
99 percentile	275.0	148.1	278.1

For the store opened in spring 2004, we identified a sample of 5696 affected customers whose distance from the nearest store substantially reduced due to store openings. We then identified a sample of 48462 unaffected customers from the total population of customers such that their distribution of distances from the nearest store is similar to the distribution of distances for the sample of affected customers prior to store opening. We provide the details of distances from nearest store for the two samples of customers in Table C1.

In Table C2, we provide the summary statistics for change in purchase and return variables from prior- to post-period for the affected and unaffected customers. It is evident from Table C2 that the store purchases and returns (in number of transactions, quantity, and revenue) increased substantially more for the affected customers as compared to the unaffected customers. But we only find higher change in online purchase values for affected customers as compared to unaffected customers. We also find a higher change in purchase diversity and average purchase price on both channels for the affected customers. Overall, we find that the change in net total purchases by the affected customers are higher than unaffected customers.

Table C2: Summary statistics for change in purchase/return behavior from prior- to post-period

Change in Variables from Pre- to Post-period	Affected Customers (5696)		Unaffected Customers (48462)	
	Mean	Std. Dev.	Mean	Std. Dev.
Purchase transactions				
Total No. of transactions	0.84	5.35	0.50	5.27
No. of store transactions	0.78	4.78	0.51	4.67
No. of online transactions	0.07	1.06	0.03	1.01
Total quantities	2.55	21.99	1.52	21.48
Store quantities	2.32	20.45	1.50	19.62
Online quantities	0.18	3.79	0.08	3.55
Total revenue (US$)	173.12	1308.23	115.28	1388.44
Store revenue (US$)	148.76	1157.80	103.69	1248.06
Online revenue (US$)	16.03	252.02	11.01	235.84
Return transactions				
Total No. of transactions	0.22	2.07	0.16	2.27
No. of store transactions	0.17	1.42	0.11	1.52
No. of online transactions	0.00	0.45	0.00	0.41
Total quantities	0.40	5.53	0.29	6.02
Store quantities	0.30	3.59	0.22	4.24
Online quantities	0.01	0.94	0.00	0.91
Total revenue (US$)	46.00	548.19	30.16	502.83
Store revenue (US$)	30.89	387.94	20.16	360.88
Online revenue (US$)	2.22	78.21	1.13	67.70
Nature of purchase transactions				
Store purchase diversity	0.000	0.485	-0.015	0.496
Online purchase diversity	0.005	0.387	-0.009	0.386
Avg. store purchase price	2.902	72.161	-0.487	81.713
Avg. online purchase price	1.985	42.526	0.981	41.826

Similar to our main analysis, we estimated the *ATT* for the effect of spring 2004 store openings on affected customers' purchase and return behaviors. We estimated *ATT* by both, the propensity score based and manual CEM based estimators and report the resulting estimates in Table C3. We find similar *ATT* estimates for the effect of store openings on customers' purchase and return behaviors from the two estimators supporting the claim of causal effect of store openings. Moreover, comparison of the effects of the fall 2003 and spring 2004 store openings revealed similar size, sign, and significance for the two *ATT* estimates. This further confirms that we have identified the general effect of store openings on customers' purchase and return behaviors on online and store channels.

Table C3: Comparison of propensity score matching and CEM Estimates

	Propensity score matching		Manual CEM	
Number of matched (total) affected customers	Not applicable		2036 (5696)	
Number of matched (total) unaffected customers	Not applicable		5703 (48462)	
Overall imbalance	Not applicable		0.668685	
Change in Variables from Pre- to Post-period	ATT	t-value	ATT	t-value
Purchase transactions				
Total No. of transactions	0.318^{***}	3.84	0.406^{***}	4.09
No. of store transactions	0.271^{***}	3.67	0.321^{***}	3.50
No. of online transactions	0.034^{**}	2.09	0.056^{***}	3.23
Total quantities	1.07^{***}	3.18	1.44^{***}	4.35
Store quantities	0.923^{***}	2.97	1.31^{***}	4.34
Online quantities	0.091^{*}	1.94	0.062^{*}	1.92
Total revenue (US$)	56.02^{***}	2.69	51.79^{***}	3.01
Store revenue (US$)	45.27^{**}	2.44	44.99^{***}	2.82
Online revenue (US$)	4.56^{*}	1.92	3.21^{**}	1.96
Return transactions				
Total No. of transactions	0.075^{**}	2.31	0.109^{***}	3.9
No. of store transactions	0.061^{***}	2.78	0.079^{***}	4.01
No. of online transactions	0.002	0.25	0.003	0.51
Total quantities	0.135^{*}	1.75	0.208^{***}	3.92
Store quantities	0.105^{*}	1.91	0.164^{***}	4.20
Online quantities	0.01	0.71	0.001	0.03
Total revenue (US$)	17.13^{**}	2.05	14.18^{***}	2.98
Store revenue (US$)	13.27^{**}	2.19	11.51^{***}	3.83
Online revenue (US$)	0.644	0.55	-0.611	0.64
Net transactions				
Net total quantity purchased	0.944^{***}	3.17	1.23^{***}	4.21

Net total revenue	38.28^{**}	2.27	37.6^{**}	2.49
Nature of purchase transactions				
Store purchase diversity	0.011	1.45	0.022	1.60
Online purchase diversity	0.012^{**}	2.05	0.019^{**}	1.98
Avg. store purchase price	3.890^{***}	3.30	2.89^{**}	1.98
Avg. online purchase price	0.855^{**}	1.99	1.22^{*}	1.94

$^{***}, ^{**}, ^{*}$ = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

Next, we compute the *ATT* for online purchase behavior of different subcategories of affected customers based on whether they make more store purchase/ return in the post-period as compared to prior-period. Similar to the analysis in the main paper, we then present the results in 2x2 matrix form in Table C4. Similar to the fall 2003 store opening results, we find that only affected customers who make either more store purchases or more store returns in the post-period or both make higher quantity (number of transactions, quantity, and revenue) and quality (purchase diversity and average product price) on the online channel. This suggests that the change in online purchases of customers is driven by higher store transactions either for purchase or for return.

Table C4: Change in online purchases with interaction of change in store purchase and returns

	Change in Online Purchase Variables from Pre- to Post-period	More post-period store purchase transactions		Less or equal post-period store purchase transactions	
		ATT	*t-value*	*ATT*	*t-value*
More post-period store return transactions	No. of affected customers	769		286	
	No. of transactions	0.173^{***}	4.25	0.191^{**}	2.56
	Purchase quantities	0.393^{**}	2.48	0.564^{**}	2.31
	Purchase revenue (US$)	35.64^{***}	2.90	31.02^{**}	2.26
	Purchase diversity	0.071^{***}	5.39	0.076^{***}	3.18
	Avg. purchase price	9.45^{***}	6.01	6.57^{**}	2.33
Less or equal post-period store return transactions	No. of affected customers	1602		3039	
	No. of transactions	0.118^{***}	4.52	-0.039^{*}	1.82
	Purchase quantities	0.302^{***}	3.50	-0.098	1.27
	Purchase revenue (US$)	15.64^{***}	2.94	-7.98	1.59
	Purchase diversity	0.047^{***}	5.12	-0.025^{***}	3.06
	Avg. purchase price	2.59^{***}	2.53	-2.48^{***}	2.78

$^{***}, ^{**}, ^{*}$ = statistically significant at $\alpha = 0.01$, $\alpha = 0.05$, and $\alpha = 0.10$ levels (two-sided test), respectively.

Overall, we found similar *ATT* estimates for overall sample as well as the different subsample of affected customers in case of stores opened in fall 2003 and spring 2004. Thus the estimated overall effect of facilitating customers' store access on their purchase and return behaviors as well as the mechanisms through which such easier store access affects their online purchase behavior are generalizable to any store openings by the retailer.