Can offline stores drive online sales?

Kitty Wang and Avi Goldfarb[*]

September 2016

Abstract

We use evidence from store openings by a bricks-and-clicks retailer to examine the drivers of substitution and complementarity between online and offline retail channels. Our evidence supports the coexistence of substitution across channels and complementarity in demand. In places where the retailer has a strong presence, the opening of an offline store is associated with a decrease in online sales and search; however, in places where the retailer does not have a strong presence, the opening of an offline store is associated with an increase in online sales and search. Our evidence suggests that while online and offline may be substitutes in distribution, they are complements in marketing communications. Specifically, the type of marketing communication driving complementarity seems to be information about the existence of the brand. For example, we see a large increase in new customer acquisition and sales, and little difference between fit and feel products and other products. Thus, it is the presence of the store, rather than information about the attributes of the particular products in the store, that drives complementarity.

[*] Wang: Bauer Collage of Business, University of Houston. Goldfarb: Rotman School of Management, University of Toronto. Anna Shen provided excellent research assistance. We thank WCAI for providing the data. The paper benefited from helpful comments from the review team, seminar participants at Marketing Science 2014, the University of Toronto, McGill University, Temple University, George Washington University, Boston University, the University of Houston, and UC Riverside, and Elea Feit, Bryan Bollinger, Scott Shriver, and Catherine Tucker.

1. Introduction

An increasing fraction of consumers shop both online and offline. We often see marketing practitioners claim "multichannel marketing is a perfect storm of synergies" (Beck 2013), and firms believe that online and offline channels complement one another:

> "…the online channel is not an alternative to the offline distribution channel, but is complementary. Even as online buying increases over time, the offline channels of distribution currently in vogue will also grow." (Agarwal, 2012).

However, much of the existing literature has focused on online and offline competitors. By focusing on competitors, this literature has emphasized that online retailers and their offline competitors are substitutes (e.g., Brynjolfsson, Hu, and Rahman 2009, Forman, Ghose, and Goldfarb 2009, Choi and Bell 2011, Sinai and Waldfogel 2004), with the degree of consumer substitution depending on local demographic characteristics, product type, and proximity to physical store locations.[1]

One possibility is that the practitioners are wrong. They misinterpret demand shocks that hit both online and offline channels as evidence of complementarity. Another possibility is that the academics are wrong. The studies mentioned above do not capture the relevant attributes of complementarity between online and offline because they look across, rather than within, firms. Furthermore, those few academic studies that look within firms (e.g. Chintagunta, Chu, and Cebollada 2012) may not capture the benefits properly.

Recent work by Avery et al (2012) provides a framework to reconcile these approaches. Drawing on the conceptual work of Alba et al. (1997), Balasubramanian, Raghunathan, and

[1] Consistent with prior literature, we define substitution to mean that when an offline store opens, online sales decrease. Hence, the offline channel takes sales away from the online channel. We define complementarity as the opposite: When an offline store opens, it causes online sales to increase.

Mahajan (2005), and Ansari, Mela, and Neslin (2008), they argue that whether substitution or complementarity dominates depends on whether conspicuous or experiential capabilities have a larger effect on the purchase process. In particular, "conspicuous capabilities" (such as immediate gratification and no shipping fees) should lead substitution to dominate as the offline store will dominate the online store in these dimensions for many consumers. In contrast, "experiential capabilities" (such as a pleasurable shopping experience and building a relationship) provide a "living billboard" that should lead complementarity to dominate over the long run. The idea is that the in-store experience enhances the brand equity across both channels. They find evidence consistent with their framework: substitution in the short run (particularly in the catalog channel) and complementarity in the long run (particularly in the internet channel).

Like Avery et al. (2012), in this paper we look within a single firm and provide evidence of both substitution and complementarity. Our primary contribution is to provide quasi-experimental evidence of the existence of a particular mechanism under which marketing communications drive complementarity: A straight billboard effect of offline stores leading to an increase in online sales. By "straight billboard effect", we mean that the store serves as informative advertising about the existence (rather than the attributes) of the brand. In identifying this particular mechanism, we aim to help understand a source of the "synergies" claimed by practitioners.

In particular, we use data from July 2010 to June 2012 on purchases by 42,000 customers of three different bricks-and-clicks retailers owned by the same firm. We first use this data set to examine what happens to online activities when the company opens a store locally. On average, we find that when a store opens offline, online sales increase (with marginal significance). This provides weak support for complementarity, on balance. However, after splitting locations by prior brand presence, we observe stronger evidence for both substitution and complementarity in demand. In locations where the brand has a prior presence before a store opens nearby, the opening of an offline store is associated with a decrease in local online sales and browsing. In locations where the brand does not have a strong prior presence, the opening of an offline store is associated with an increase in online sales and browsing. We measure presence by the pre-existing sales of a brand in a narrowly defined location (the census tract), and show substitution only occurs when presence is high.[2]

We then explore the reasons why sales might increase online in response to an offline store opening. We provide four types of evidence to support a straight awareness-driven billboard effect, rather than a living billboard that is complementary to existing brand equity. First, we show that the opening of a new offline store leads to an increase in the number of first-time customers from the surrounding areas. This suggests that the offline store provides information about the retailer. Second, we demonstrate that the increase in online sales in an area after store opening is driven by these new customers and this increase persists for several months after the store opening. At the same time, the online sales from existing customers acquired before a local store opening do not appear to change after the store opens. This is true regardless of whether the customer resides in an area with a prior brand presence or not. Again, this suggests that the store provides information to these new customers. Third, we explore whether the complementarity is generated through the provision of information about the fit and feel of the products. We find no qualitative difference between apparel and other products. In this way, we do not find evidence to support the idea that the offline stores provide this important type of product attribute information. Fourth, while online product returns do fall after a store opens, we find no

[2] In contrast to our results, Avery et al find complementarity to be stronger when the retailer already has stores operating locally. Below, we hypothesize that the difference is driven by the increased size of the internet channel between the time of their study (1997 to 2006) and our study (2010 to 2012).

difference between places with a brand presence and places without such a presence, again suggesting that product attribute information does not drive the results.

Therefore, we argue that the marketing communication impact of the offline store we identified in areas without a brand presence is more likely to be information about the existence of the brand in general rather than information about product attributes. This analysis allows us to reframe the structure of Avery et al. in the familiar language of the core marketing framework: As distribution channels, online and offline stores appear to be substitutes, but the offline store generates a particular complementarity in marketing communications.

The retail brands in our data conducted almost no TV or print advertising. There was some online search and display advertising, as well as direct email and catalog campaigns to existing customers. Unlike the more established brand in Avery et al., this means that the dominant offline marketing communications channel may be the store itself. The benefit we propose therefore relates more directly to awareness: When an offline store opens locally, some consumers become more aware of the brand and its products. Consumers who were previously unfamiliar with the retailer become more likely to purchase through both channels.

Bell, Gallino, and Moreno-Garcia (2014) also identify an information-based mechanism. They document that offline showrooms increase sales for a previously online-only retailer by providing product information through sampling. Our results suggest that the communication happens through increasing brand awareness whereas they find evidence that for their product category (eyeglasses), offline stores do communicate information about product attributes Nevertheless, building on the framework of Avery et al., the two papers use different data and retail settings to come to similar conclusions: The offline channel serves as an informative marketing communications tool for the online channel.

More generally, a rich literature examines the informative role of advertising. We emphasize the role of stores in providing information about a brand's existence. This advertising mechanism is identified in Telser (1964), and is consistent with results on high effectiveness of search advertising for less-known brands (Narayanan and Kalyanam 2015, Simonov, Nosko, and Rao 2015). Advertising can also provide information about product attributes (e.g. Narayannan, Manchanda, and Chintagunta 2005; Ching and Ishihara 2012), which is similar to the experiential learning emphasized in Avery et al., and to the product sampling emphasized in Bell, Gallino, and Moreno-Garcia. Recent work on "showrooming" emphasizes the informative role that stores play in communicating product attributes (e.g. Rapp et al 2015, Wu, Wang, and Zhu 2015).

The role of physical stores in enhancing brand image is well-established in the ethnographic marketing literature and in the popular press. Kozinets et al (2002, p. 17) discuss the role of flagship retail stores that are "operated with the intention of building or reinforcing the impact of the brand rather than operating to sell product at a profit." Walter Isaacson (2011), in his biography of Steve Jobs, emphasizes the brand building role of the Apple Stores. Similarly, when Microsoft stores started opening, the press often emphasized that the stores were showrooms (e.g. Manjoo 2012). In our study, we argue that local offline stores help create brand awareness, and consequently, help with acquisition of new customers from the area. The newly acquired customers then in turn buy more, online and offline. Prior literature has also explored how information in one channel leads to purchases in another channel (e.g. Joo et al 2014; Wiesel, Pauwels, and Arts 2011; Ansari, Mela, and Neslin 2008), and how store redesigns are particularly effective for attracting and retaining profitable new customers (Dagger and Danaher 2014). Other literature has also explored channel substitution across retail store type (e.g. Qian, Anderson, and Simester 2013) and the channels of integrating online and offline channel experiences (as summarized in Verhoef, Kannan, and Inman 2015).

We conclude our introduction by emphasizing that we interpret the estimated correlations between the store opening and changes in online sales and other activity as likely to be causal. We show that the timing of the change in online sales is closely tied to the store openings. Our key identification assumption is that store openings are not correlated with unobservable changes in demand that are differentially trending in locations where the retailer has a presence and locations where the retailer has no presence. While it is possible that a retailer might open offline stores in locations with anticipated increased demand, our results show that opening an offline store sometimes precedes higher online sales and sometimes precedes lower online sales, depending on the degree of prior brand presence. We next describe the data and setting for our study.

2. Data on Offline Store Openings and Customer Behavior

We use data provided by a U.S. based specialty retailer to the Wharton Customer Analytics Initiative to investigate the effect of offline store openings on customers' online purchases, offline purchases, and online search behavior.[3] The company owns three sub-brands operating in the same industry. The firm is vertically integrated: the products are designed by the brands' in-house team, and are sold in their own retail stores. The retailers sell a variety of products including clothes, shoes, housewares, and accessories. Some of the products sold by the company have non-digital attributes (such as fit and feel), which suggests that consumers may have a tendency to prefer browsing and purchasing through the offline store rather than through

[3] Bollinger and Shriver (2013) use the same data set in their structural model of cross-channel revenue effects. Danaher and van Heerde (2014) use the same data set in their study of the impact of direct marketing on online and offline sales. Soysal and Zentner (2014) use the same data set to study differences in product popularity across channels.

the online channel. We investigate what happens to local online sales, offline sales, web browsing, and new customer acquisition when the company opens an offline store nearby.

The data set contains purchases made by 42,000 randomly sampled US customers between July 2010 and June 2012. We use the first three months as a pre-period to measure brand presence, and they are subsequently dropped in the main analysis. Our analysis is therefore on data from the 21 months from October 2010 to June 2012. All three brands had established both online and offline stores far before the beginning of the data period. For these 42,000 customers, we have information on where they are located (tracked by credit card and shipping address), where each of the purchases are made (if from an offline store, where this store is located), and the quantity purchased in dollars. In addition to purchase history, the data contain customers' online browsing behavior on the company's own websites.[4] We use data from all three brands. Table 1 provides descriptive statistics.

We aggregate our data to the brand-census tract-month level. Although we have data on each sale by each individual customer over the two-year period, we aggregate to monthly sales by census tract. Because the key covariate of interest (offline store entry) varies by location (not by customer), we aggregate to the relevant level of observation. Aggregating to the location level does not lead to a loss of information relevant to the analysis. This is not to say that we ignore heterogeneities in sales across customers. In contrast, analysis at the tract level enables measurement of the arrival of new customers from that area, hence we are able to study how customers acquired after a physical store first enters a local market purchase differently from existing customers from the same area. This is a parsimonious way to control for location-specific differences between new and old customers.

[4] For two of the three brands, it was possible to order via mail or phone from the catalog. A company representative said that catalog sales are so small that they do not consider it a meaningful sales channel. Therefore, they did not provide us with such data and we do not think this missing data will affect our results in a meaningful way.

This aggregation means that, while we define our variables at the location level, our measures are all based on the random sample of 42,000 customers that we observe. Sales in a location are therefore a randomly sampled measure of sales from that location. Our interpretation relies on having a large number of customers and a large number of tracts. The analysis will be consistent, but not as precisely measured as it would be if we had all customers. When measuring a location's brand presence, we focus on splitting locations into two types: "no brand presence" and "with prior brand presence". "No brand presence" means no customers from our random sample purchased in that location, either online or offline, in the first three months of the data. However, our results do not rely on this particular discretization of brand presence, and in the online appendix, we show robustness to several alternatives.

We measure a variety of outcomes that we use as dependent variables. Specifically, for each brand-census tract-month, we measure online sales revenue, total sales revenue, number of sessions on the website, number of customers browsing the website, and number of new customers. These measurements are the totals for all customers in our sample with a home address in that tract.[5] We measure new customers using the company's identifier for "first purchase". While this would not include some cash purchases, it does not depend on the time span of our data set.

The census tract is a geographically small unit of analysis: the 42,000 customers in the sample are located in a total of 19,076 tracts, resulting in an average of 2.20 customers in each tract. This yields 1,201,596 observations at the brand-tract-month level. In much of our analysis, we use fixed effect Poisson regression, which drops any brand-tracts with only zero values of the dependent variable, reducing the total number of observations.

[5] Offline sales are aggregations of purchases made by customers living in each tract. For instance, if a customer who lives in Chicago shops once in a store in New York City, this purchase would count towards total sales in the tract in Chicago, not New York City.

Our core covariate of interest is whether there is an offline store within 25 miles. We sometimes refer to this variable as "store opening" because our identification comes off the addition of new stores. At the beginning of the data period, there were a total of 403 physical stores in the U.S. for three brands (166, 165, and 72 respectively from brands A, B, and C). During the 21-month period (after dropping the 3-month training period), there were 88 store openings (25, 27 and 36 respectively). There were just 3 store closings (2, 1, and 0 respectively) and so we do not use store closings in our analysis. We use physical proximity of each tract to stores to measure offline accessibility. To do so, we first use the address of each store to determine its latitude and longitude. We then use these geographic coordinates to measure the straight line distance between the center of each census tract and each retail store. This calculation is updated every month. As a result, when a store opens, it can affect the distance measures of all tracts. For each census tract, we then use the distance measures to construct the dummy variables that indicate if this is the first store from a particular brand within 25 miles. We report results using the 25-mile measure, though results are robust if we use 10 miles or a continuous measure of distance instead.

After these calculations, we define the store opening variable as equal to one for all months after the first time a tract is within 25 miles of a store by that brand. It is equal to zero otherwise. Therefore it is zero before the store opens and for tracts in which no stores open during the data period. If more than one store opens, only the first store opening is registered.

In several specifications, we include a number of controls. First, and perhaps most importantly, we control for direct marketing communications sent out by the firm to customers in the sample. We measure direct marketing effort by counting the total number of both online and offline direct mail sent out by each brand to customers in a tract each month. The offline direct mail includes catalogs sent by the company.[6] Second, we collected the store opening data on six major competitors to control for the competition effect. The six major competitor brands were identified by the firm as "brands that share a common customer base". These competitors have broad clothing lines that compete with all three brands. The openings were collected directly from the competitor websites and from press releases mentioning store openings. We use competitor store within 10 and 25 miles as controls. We do not model the three brands owned by the firm in our data as competitors with each other.[7] Third, we interact the census tract population with monthly time dummies to control for the impact of population size over time. We obtain the population data from the 2010 U.S. census. Fourth and finally, time and brand-tract fixed effects are also included in our analysis.

3. Empirical Strategy and Results

3.1 Empirical Strategy

We explore how the opening of an offline store affects local online sales, local total sales, local online browsing, and local new customer acquisition. In this section, we first discuss the main effects when all census tracts are treated homogenously. Because stores open gradually over the 21-month period in different locations, we are able to use a fixed effects difference-in-difference identification strategy to estimate the average effect of store openings. Specifically, when the first new store opens within 25 miles of a census tract, the covariate defining store openings changes from value 0 to 1. When examining the average effect, the control group includes census tracts that either never had a store opening within 25 miles during the sample period or already had a store within 25 miles at the beginning of the sample period. The

[6] In the online appendix, we show that the main results are robust to separating the effects of email and catalog mailings. These email and catalog campaigns were rarely about store openings. As shown at the bottom of Table 1, of 11,600,454 emails sent, just 1,198 contained content about store openings.

[7] In the online appendix, we show robustness of the main results to including store openings from the company's other brands.

treatment group includes tracts that had a first-store opening within 25 miles during the sample period.

We then split the census tracts into groups based on local brand presence in the months prior to the beginning of the estimation sample, and examine if the opening of a local store has the same impact in areas with versus without brand presence. In this set of regressions, we employ a difference-in-difference-in-difference identification strategy. For the last "difference" we compare the impact of store opening between tracts without a prior brand presence and those with a prior brand presence. In order to measure brand presence, we use the first 3 months as a pre-period, and they are subsequently dropped. Therefore a total of 21 months in the data are used in the analysis.

The unit of observation of our analysis is the brand-tract-month level. We present results using a conditional quasi-maximum likelihood fixed effects Poisson regression specification developed by Hausman, Hall, and Griliches (1984), showing robustness to a wide variety of other specifications in the online appendix. The regressions are implemented using Stata's xtpoisson, fe function.

We emphasize fixed effects Poisson because it is particularly useful for non-negative but skewed data (Azoulay, Zaff-Griven, and Wang 2010). This has made it one of the core models in research on innovation, where distributions of outcomes are highly skewed. In particular, the Poisson model is in the linear exponential family, meaning that the coefficient estimates remain consistent as long as the mean of the dependent variable is correctly specified (Gourieroux, Monfort, and Trognon 1984). In other words, as Wooldridge (2002, p. 675) notes, the estimates of the parameters for the conditional mean are consistent even if the mean and variance do not have the same value and "there can be overdispersion or underdispersion in the latent variable model". This estimator can be used for any non-negative dependent variables, whether continuous or whole number (Santos Silva and Tenreyro 2006). Furthermore, the quasi-maximum likelihood (robust) standard errors are consistent even if the underlying data-generating process is not Poisson.

The fixed effects Poisson model has several other advantages for our context, relative to other non-linear models. As in a linear panel model with many fixed effects, the brand-tract fixed effects are conditioned out rather than estimated (Wooldridge 2002, p. 675), overcoming the incidental parameters problem that appears in many other non-linear fixed effects models (Lancaster 2000). The quasi-maximum likelihood standard errors are robust to arbitrary patterns of serial correlation (Wooldridge 1997) and hence do not require clustering as recommended for difference-in-difference estimation in linear models by Bertrand, Duflo, and Mullainathan (2004). Wooldridge (2002) contains a lengthy discussion of this method.

The main cost of using the fixed effects Poisson model is efficiency. For example, a negative binomial regression may be more efficient under some assumptions. It might also be possible to add power by using Seemingly Unrelated Regression (Wooldridge 2002, p. 163), though this is not straightforward in the fixed effects Poisson context. We view the choice of consistency over potential efficiency as the more conservative option. Moreover, given that we find statistical significance, this choice seems to have little impact in terms of statistical inference.

We proceed by first examining the relationship between store openings and online sales, total sales, and online browsing activity. We then compare places with and without prior brand presence. In section 3.4, we present our core contribution exploring the underlying mechanism and documenting the marketing communication role of the offline store.

3.2 Main effects

We identify the main effect of local store openings by comparing the change in outcomes in the same location before and after a store opens locally to the change in outcomes in the same time period in other locations that did not have local store openings. Therefore, in this subsection, the main identifying assumption is that there was no location-level differential trend in locations with stores opening relative to locations without stores opening. Clearly, this is a strong assumption. We add controls to partially alleviate this concern, and to demonstrate that the most obvious controls do not change the estimated coefficients much.[8] Specifically, we estimate the impact on sales and online browsing using the following fixed effect Poisson model:[9]

(1) $Outcome_{blt} \sim Poisson(\mu_{bl} \exp(\alpha Store25_{blt} + X_{blt}\theta + \tau_t))$

$$b = A, B, C, \quad l = 1 \ldots, L, \quad t = 1,2, \ldots, 21,$$

The outcome variables, for each brand (b)-location (l)-month (t) include online sales, total sales, the number of sessions on the website, and the number of unique visitors to the website. $Store25_{blt}$ captures whether there is a store within 25 miles and so α is the main coefficient of interest; X_{blt} is a vector of controls, including direct marketing communications, store openings by competitors, and interactions between tract population and each month of the data; τ_t captures the month fixed effects for each of the 21 months in the data (excluding a base month); and μ_{bl} captures brand-tract fixed effects.

Table 2 presents the results for separately regressing online sales, total sales, number of website sessions, and number of customers browsing the website on store openings and the controls. In the online appendix, we show robustness to a propensity score matching

[8] Our core results in section 3.2 below, however, rely on the weaker assumption of different trends when store openings occur in places with a brand presence relative to when store openings occur in places without such a presence. Furthermore, in the online appendix, we show that email and catalog marketing do not significantly change in the aftermath of a store opening.

[9] We show robustness to a number of other models including linear, log linear, and tobit models.

specification, a specification without the marketing message and competition controls, and a fixed effects linear specification. Column (1) shows that online sales (weakly) rise when a store opens. This suggests that there might be complementarity between the online and offline stores. The opening of the offline store seems to generate online sales.

Column (2) shows that there is a sharp increase in total sales when a store opens. This is not surprising since sales from the new store are counted toward total sales. As shown by Pauwels and Neslin (2015), when more people have access to a product, it sells more. Furthermore, back-of-the-envelope calculation suggests that our estimated magnitude is similar to their estimate of a net increase in revenues of 20%.

Columns (3) and (4) look at whether new offline stores are correlated with increased traffic to the website, even if sales do not rise much. Column (3) measures the effect on website traffic with the number of distinct online sessions by sampled customers. Column (4) measures the effect on website traffic with the number of unique customers in the sample to the website. The consistently positive and statistically significant coefficients suggest that traffic does increase.[10]

Table 2 suggests neither strong substitution nor strong complementarity between the two channels, though overall there does seem to be more complementarity than substitution. Thus, on average, in this data set, the practitioner's emphasis on complementarities seems to dominate. Next, we conduct further analysis to attempt to reconcile the practitioner and academic literatures on complementarities and substitution, building on the framework of Avery et al. (2012). Finally, we present our core contribution which is careful identification of the mechanism underlying the evidence of complementarities.

[10] Compared to the results without controls in the online appendix, the coefficient magnitudes are similar. This suggests that including two additional types of controls as covariates does not have a substantial impact on the results. Thus the impact of unobservables would have to be relatively large relative to the impact of these observables for omitted variables to generate a substantial change in our qualitative results (Altonji, Elder, and Taber 2005; Oster 2014).

3.3 Complementarity vs. substitution moderated by brand presence

Table 2 documents that, when treating all census tracts homogeneously, we observe something that looks more like complementarity than substitution, though the effects are not particularly strong. Next, we investigate how prior brand presence changes the way these two channels interact and how they affect customers' purchase and browsing behavior. We show evidence of substitution in places where the brand already had a strong presence and complementarity in other places. We argue that this is suggestive of a marketing communications role for offline stores. Below, in section 3.3, we provide further evidence to support this argument.

We define a brand to have a presence in a census tract when there are positive sales either online or offline in this tract in the three months before the data period of the analysis. Otherwise, a brand has no presence in a census tract. We determine the brand presence for each brand separately. For example, if brand A has a presence in tract number 109, it does not mean brand B also has a brand presence in this area. We then create a dummy variable $BrandPresence_{bl}$ that takes the value of 1 if the brand b has a presence in tract l, and 0 if the brand has no presence in tract l. The presence is therefore defined by the small number of observed customers in the tract and relates directly to the customers in our sample.

We therefore add the interaction of brand presence and store openings to our estimating equation. Because $Presence_{bl}$ is defined at the beginning of the sample and does not change over time, the main effect of $Presence_{bl}$ drops out due to the brand-tract fixed effects:

$$(2)\ Outcome_{blt} \sim Poisson(\mu_{bl} \exp(\alpha Store25_{blt} + \beta Store25_{blt} \times Presence_{bl} + X_{blt}\theta + \tau_t))$$

In interpreting the coefficient on the interaction term in the Poisson model, we checked the marginal effects at the median values yielded similar qualitative results to the signs of the coefficients (Ai and Norton 2003).

Compared to section 3.2, the main identifying assumption we use here is weaker. To accept the estimation results in the remainder of the paper, we assume that there is no differential trend in sales and online search between locations with and without a prior brand presence. Similar to the estimation of equation (1), we address this concern in the online appendix by adding and taking away control covariates, and show that the magnitudes and significance are comparable with and without controls.

Table 3 shows estimation results, mirroring the structure of table 2.[11] Column (1) shows a stark difference in online sales after a store opens offline between places with and without a brand presence. In places where the brand did not have strong sales prior to the offline store opening, the first row shows that sales rose substantially. The linear specification in the online appendix suggests a back-of-the-envelope marginal impact of 96.6% (2.825/2.925). Thus, in places where the brand did not have a presence at the beginning of the sample, offline store openings seem to help the online channel, suggesting some kind of complementarity.

In contrast, adding the first and second rows together suggests that adding the offline channel reduces online sales in places that had a brand presence prior to a store opening. The sum of the coefficients is significantly negative and large in magnitude: -35.6% ((2.825-9.907)/19.915) in the linear specification. This is consistent with substitution.

[11] In the online appendix, we show robustness of table 2, and especially of the main result in column (1), to a variety of alternative specifications including propensity score, linear regression, dropping outliers, dropping the controls for marketing variables and competition, separating email and catalog marketing messages, a state-by-state Tobit specification, various log-linear specifications, different brand presence definitions including by Metropolitan Statistical Area rather than by distance, adding a lag between the time of the brand presence definition and the start of the data analysis, and brand-by-brand results.

Thus, in terms of the effect of offline stores on online sales, the results suggest the presence of both complementarity and substitution: It depends on the pre-existing presence of the brand in that location.

Column (2) suggests a large increase in total sales in places without a brand presence, and a positive but smaller effect in places with a brand presence.[12] Similarly, Columns (3) and (4) show a sharp increase in both browsing sessions and unique customers for places without a strong brand presence at the beginning of the sample. The second row suggests that the positive effect goes away for places with a strong presence. Overall, we interpret Table 3 as to suggest the possibility that two different forces are at play and we see both substitution and complementarity, depending on the prior presence of the brand.

Figure 1a repeats the analysis of Table 3 column (1) on online sales, but at a finer level of detail over time. Specifically, it splits the key covariates into a sequence of dummy variables for the months before and after a store opens (also interacted with prior brand presence). The base is more than four months before a store opens. We graph the coefficients associated with these dummy variables to show how activities change in accordance to the timing of the store opening. The solid line shows the coefficients on the effect of store openings in locations without a prior brand presence. It shows online sales are higher after the store opens than before, with a big increase between the month before and after opening. Prior to the store opening, the estimated coefficients are near zero and generally flat, suggesting no substantive increase in the four months prior to opening. There is a small and insignificant increase between two and one months before opening, meaning that we cannot reject the possibility that information about the

[12] The estimation result on total sales likely serves as a lower bound for the true effect because cash transactions made in store are not included because they cannot be mapped to a particular customer and census tract. Therefore, the results will underestimate the overall revenue impact of store openings. Partly for this reason, we do not emphasize the offline sales data nor do we interpret the results to be informative about the profitability of opening new stores.

forthcoming opening led to online sales. Still, we interpret this figure to suggest that it is unlikely that the results are driven by reverse causality: expected increase in demand causing store openings.

The dashed line shows the estimates for store openings in locations with a prior brand presence (the equivalent of the sum of the first two rows of table 3). Here, it is hard to identify any clear change at the time of the store opening, though there is a slight decrease in the coefficient size for the first few months and a sharp decrease four or more months after opening. Below, we will provide evidence of sharper effects on browsing for locations with a prior brand presence.

Figure 1b repeats the analysis of Table 3 column (2) on total sales, again splitting the key covariates into a sequence of month dummies with interactions. The solid line shows that, unsurprisingly, total sales sharply rose in places without a prior brand presence and the dashed line shows that total sales also rose in locations with a prior brand presence, but only weakly.

Overall, these results provide evidence for both channel substitution and complementarity. They also suggest the circumstances under which each of these scenarios happen: When the brand is already known locally, the online and offline channels serve as substitutes. In these areas they are simply two alternative distribution channels. In contrast, when the brand is not known locally, the online and offline channels seem to serve as complements.

It is important to recognize that the results in this subsection differ from the results of Avery et al (2014). In particular, Avery et al find that a prior presence led to an even larger increase in online sales. Given the anonymity of the retailer in Avery et al, any discussion of the reason for the difference is necessarily speculative. Still, we believe the most likely explanation has to do with the time period rather than the retailer. Avery et al studies a period from the late

1990s to 2006. During this period, the online channel was relatively small and so drawing in new customers to the store might draw people into the nascent online channel. In contrast, our data studies a period where the online channel is more mature (2010-2012), and so the cannibalization effect dominated when a second store opened. In other words, we speculate that there were not enough online sales to cannibalize in the Avery et al time period. Consistent with this hypothesis, Avery et al find cannibalization in the catalog channel, at least over several years.[13] Because of their different results, Avery et al emphasize experiential learning through store openings. We believe an alternative mechanism is at play in our setting: A billboard effect as defined above.

Next we explore whether this is driven by the potential for offline stores to act as marketing communications channels that inform consumers of the brand.

3.4 Mechanism: A marketing communications role for the offline store

The existence of both positive and negative effects of offline store openings on online sales suggests that two different forces are at play. We argue that the positive effect is due to a marketing communications role and the negative effect is driven by online and offline as competing distribution channels.

As communications channels, information gathered offline can positively affect the online channel. Offline stores can enhance the brand perception (Kozinets et al. 2002; Avery et al. 2012) or act as a billboard for the existence of the brand (Avery et al. 2012). Offline stores can also provide information about the products offered, and the literature has emphasized that uncertainties about fit and feel are much more easily resolved in the store (e.g. Ward and Morganosky 2002; Lieber and Syverson 2012; Bakos 2001). Generally, a core challenge in

[13] An alternative explanation relates to the definition of brand and brand presence; however we believe the evidence is not consistent with this explanation. In particular, they use a 60 minute drive to define pre-existing stores and we emphasize a 25 mile radius. In addition, their stores are primarily in malls and ours are a mix of malls and neighborhoods. We do not believe this explanation is likely because our results hold for defining presence by city (or even state).

online retail is information asymmetry: The customer does not have as much information about that particular product as in the offline environment. This matters more for non-digital attribute products (Bakos 2001; Borenstein and Saloner 2001; Waldfogel and Chen 2006). Thus there are two distinct marketing communications roles that offline stores may serve: Brand awareness and resolving product information uncertainty.

As distribution channels, the online and offline channels substitute for one another. Models by Balasubramanian (1998) and Zhang (2009) emphasize that the online channel provides a substitute for the offline channel, and the empirical literature has largely supported this perspective, albeit with a focus on substitution between competitors (Forman, Ghose, and Goldfarb 2009; Brynjolfsson, Hu, and Rahman 2009). Our result of a negative effect of offline openings on online sales in places with brand presence is consistent with this emphasis: When the new store provides little new information, we see online sales fall.

In the remainder of this section, we explore which type of marketing communication mechanism is behind the increase in online sales when stores open in places without a brand presence. We argue that it is driven by the store serving as a billboard that provides information about the existence of the brand. We first demonstrate that offline store openings led to an increase in new customers from that location. Then we show that the new customers had a lasting effect on sales beyond the first few months after a store opening, even online. Third, we show that the effect does not seem to be about fitting or trying on clothes offline before buying online. Together, we argue that this suggests a billboard-like informative marketing communications role for the offline store.

The first evidence appears in Figure 2. It examines what happens to the number of new customers acquired before and after a store opens locally, in areas with and without prior brand presence. As mentioned above, new customers are defined using a "first purchase" flag in the company's dataset. As in Figure 1, we regress the number of newly acquired customers on month-by-month dummies for whether or not there is a store within 25 miles and the interaction of store openings with brand presence, and the control covariates. Newly acquired customers in each brand-tract-month are those whose first ever purchase with a brand are made in that particular month. The figure shows that the number of new customers acquired increases after a local store opens, and this increase in new customers is particularly prominent in areas without prior brand presence. There is a sharp increase in new customers after opening. The increase in new customers is largest the first month after a store opening for places with a brand presence and weakens substantially over time. In places without a brand presence, the effect persists beyond three months. This suggests that offline store openings led to a persistent increase in new customers. These results do not separate whether this increase is due to increased brand awareness or increased product information. However, before we show our evidence on separating these effects below, we first show that the overall increase in online sales is driven by these new customers.

Specifically, Table 4 examines whether the persistent increase in online sales is driven by customers acquired before a local store opens ("old customers") or customers acquired after a local store opens ("new customers"). We use Table 4 to further investigate whether offline stores act as a communication device (either through brand awareness or product information) rather than through another mechanism. The intuition is that if offline stores complement the online channel through enhancing brand awareness, then it should be the customers who are not previously familiar with the brand that will be affected the most after a local store opening, compared to the customers who are already aware of the brand and its attributes.

Table 4 replicates the analysis in the first column of Table 3 but splits the sales into those acquired before the store opened and those acquired after opening. Column (1) uses online sales by old customers as the dependent variable and column (2) uses online sales by new customers as the dependent variable.[14]

Table 4 shows that after a store opens nearby, we see no increase in purchases made by old customers. In other words, regardless of whether the customers are in locations with high or low brand presence, we see no increase in sales from customers who were acquired prior to the opening of the local store. In contrast, we see an increase in sales made by new customers. The increase in new customers' sales is not surprising given that, by definition, they generate zero sales prior to opening. Still, we include the new customer results in column (2) because we think they provide a useful contrast to the results on old customers in column (1).

Overall, we interpret the results of Figure 2 and Table 4 to suggest that opening the offline store generates sales by new customers to the online store. Although this is what we would expect if the role of the offline store is to communicate the existence of the brand to potential customers, it is also what we would expect if the role of the offline store is to provide information about product attributes such as fit and feel. It suggests informative marketing communication, but does not identify the particular type of informative marketing communication, whether about the existence of the brand or about the product attributes.

Therefore, we next examine whether the increase in online sales attributed to offline stores in places without a brand presence is driven by attribute information about the fit and feel of products. If so, this is a marketing communications role related to communicating specific information about the match of the product with a particular customer, rather than

[14] In the online appendix, we show robustness to a propensity score method, to a linear specification, and to dropping the marketing and competition controls.

communicating general information about the existence of the brand. We explore this hypothesis by comparing products for which fit and feel are likely to be important with products for which they are less likely to be important.

Table 5 examines the role of product fit. Columns (1) to (6) show that the result identified in Table 3 above holds for fit and feel products and for other products. We used three external coders to code each of the 76 product categories provided by the company into fit and feel, and non fit and feel products. Out of the three coders, two were Ph.D students, and one was an undergraduate student. Two coders first coded the product categories, then the third coder served as a tie-breaker in the few situations for which the first two coders disagreed. For robustness, we then estimate the same equation using two alternative definitions of fit and feel and non fit and feel products (columns (3) to (6)).[15] Without a prior brand presence, online sales of both types of products rise. With a brand presence, they (weakly) fall. We interpret this to reject the hypothesis that the impact of opening offline stores is primarily about telling potential customers how the products fit. The strong impact on non-fit and feel products suggests product attribute information is not the primary driver of the results. Therefore, we emphasize a general brand awareness effect.[16]

Related to information about the match of the product to the customer, Table 5 column (7) looks at how returns on online purchases are affected by store openings. We find that the percentage of returns to total online sales revenue decreases post store opening. The decrease of proportion of returns is consistent with findings in Bell, Gallino, and Moreno-Garcia (2014);

[15] In these definitions, we did the assignment to fit and feel ourselves. We assigned categories defined by bottoms, tops, apparel, and dresses to fit and feel and categories defined by housewares, accessories, gifts, furniture, intimates, bed/bath, holiday, and plants to not fit and feel. Under definition 1, shoes are fit and feel. Under definition 2, shoes are not fit and feel.

[16] In the online appendix, we show robustness to a variety of alternative specifications including propensity score, log linear, and dropping outliers.

however, this decrease is true in locations with and without prior brand presence, and therefore does not explain the differences in the impact of store openings on online sales across locations.

In other words, while we do not have direct measures of brand awareness, we interpret Tables 4 and 6 to suggest that increased brand awareness drives the increase in online sales after a store opens in areas without a strong prior brand presence. As in much other work on branding (e.g. Simon and Sullivan 1993; Kamakura and Russell (1993)), we do not observe brand awareness and generate our interpretation by eliminating other explanations and by looking for suggestive evidence. We are comfortable with this interpretation because it is consistent with our collective results: (1) offline stores (weakly) increase online activity controlling for marketing activities (Table 2), (2) this effect is driven by locations without a prior brand presence and the sign often reverses in other locations (Table 3; Figure 1), (3) this effect is driven by new customers (Table 4; Figure 2), (4) this effect is not stronger for fit and feel products relative to non-fit and feel products (Table 5), and (5) there is no difference in changes in product returns for places with and without a brand presence (Table 5).

4. Conclusions

In this paper, we use data of store openings from three different bricks-and-clicks retailers owned by the same firm to reconcile the industry perception of complementarity with academic research findings of substitution between online and offline retail channels. We build on the framework of Avery et al. (2012) and investigate what happens to online and offline activities when a company opens an offline store locally. We find that when treating all areas equally, our data suggest neither strong substitution nor complementarity, though on balance the evidence suggests online sales likely rose. However, splitting areas by brand presence, we find that online sales and online browsing increase only in areas without a prior brand presence after the company opens a store locally, and decrease in areas that already had a brand presence prior to a store opening.

We interpret our findings to suggest that when viewed as retail channels, online and offline are likely substitutes; however, we argue that the complementarities between online and offline channels are created through informative marketing communications generated by the mere presence of offline stores.[17] We then show evidence consistent with the store serving as a billboard that provides information about the existence of the brand being the most likely marketing communications role of the offline stores in our setting.

Our findings are in many ways consistent with Avery et al. (2012) and Bell, Gallino, and Moreno-Garcia (2014). Together, we believe these three papers provide compelling evidence of complementarity between online and offline retail channels, perhaps reducing doubt of the validity of the empirical finding in any one of the papers (Meyer 2015). In each paper, the mechanism is related to marketing communications. We emphasize an awareness-focused billboard effect, while Bell, Gallino, and Moreno-Garcia (2014) emphasize quality and fit information and Avery et al. (2012), though generally more agnostic about the mechanism, emphasize a brand-building billboard effect. Future work (both theoretical and empirical) could look across a variety of retail settings to further unpack these mechanisms and provide a unifying framework for these three related papers.

There are several limitations to this research. First, it is important to note that several of the hypotheses we test have been speculated previously. Our contribution is in providing new quasi-experimental evidence supporting these hypotheses rather than hypothesis generation per se. Second, we only have sales and online browsing data from one company, therefore we are not

[17] This hypothesized mechanism is different from the "mere exposure effect" in which familiarity generates fluency (e.g. Fang, Singh, and Ahluwalia 2007). We emphasize informative, rather than persuasive, marketing communications.

able to examine online and offline substitution across many different firms. Third, we lack data on competitors' marketing activities that can affect the sales and online search from the company we examine. Fourth, as with all treatment effects analysis, our results only measure the local average treatment effect for places that experienced store openings. Thus, our results are most informative about places that are similar to the places in our data that experienced store openings during our sample. Fifth, while we believe our evidence points to a billboard effect (new customers, not fit and feel, etc.), we cannot directly observe whether customers showroomed: visiting the offline store and then buying online. Sixth, we have a sample of customers rather than the full set of customers. This means that our measures of "brand presence" and our use of the language "locations with brand presence" are specific to our sample. It is possible that there are other consumers who buy in that location. This does not change our interpretation in terms of the relevant margin for our analysis: purchasers of our sample. It does mean that we cannot infer social effects or local spillovers from our results. In addition, we do not have direct measures of brand awareness. Instead, we infer brand awareness based on sales, browsing, the lack of difference between fit and feel products and other products, and the lack of difference in returns between places with and without a brand presence.

Nonetheless, we believe our results show evidence for both substitution and complementarity between online and offline channels. In particular, in addition to providing an additional distribution channel, opening a store serves a marketing communication purpose through what appears to be enhanced brand awareness.

References

Agarwal, S. 2012. Online Channel would Complement Existing Distribution. *Live Mint & Wall Street Journal*, Nov 21, 2012

Ai, Chunrong, and Edward C. Norton. 2003. Interaction terms in logit and probit models. *Economics Letters* 80: 123-129.

Alba, Joseph, John Lynch, Barton Weitz, Chris Janisewski, Richard Lutz, Alan Sawyer, and Stacy Wood. 1997. Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Markets. *Journal of Marketing* 61(4), 38-53.

Altonji, Joseph, G., Todd Elder, and Christopher R. Taber. 2005. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy* 113(1): 151-184.

Ansari, Asim, Carl Mela, and Scott Neslin. 2008. Customer Channel Migration. *Journal of Marketing Research* 45(1), 60-76.

Avery, Jill, Thomas Steenburgh, John Deighton, and Mary Caravella. 2012. Adding Bricks to Clicks: Predicting the Patterns of Cross-Channel Elasticities Over Time. *Journal of Marketing*, 76(3), 96-111.

Azoulay, Pierre, Jonathn Graff Zivin, and Jialan Wang. 2010. Superstar Extinction. *Quarterly Journal of Economics* 125(2), 549-589.

Bakos, J. Yannos 2001. The emerging landscape for retail e-commerce. *Journal of Economic Perspectives.* 15(1), 69-80.

Balasubramanian, Sridhar. 1998. Mail versus mall: a strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science*, 17(3), 181-195.

Balasubramanian, Sridhar, Rajagopal Raghunathan, and Vijay Mahajan. 2005. Consumers in a Multichannel Environment: Product Utility, Process Utility, and Channel Choice. *Journal of Interactive Marketing* 19(2), 12-30.

Beck, Jennifer. 2013. Multichannel Marketing is a Perfect Storm of Synergies. http://blogs.gartner.com/jennifer-beck/multichannel-marketing-is-a-perfect-storm-of-synergies-2/, July 3.

Bell, D., S. Gallino, and A. Moreno. 2014. Inventory Showrooms and Customer Migration in Omni-channel Retail: The Effect of Product Information. Working Paper, Wharton School.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1), 249-275.

Bollinger, B., and S. Shriver. 2013. Retail Entry in a Multi-Brand Environment: Empirical Analysis of Cross-Channel Revenue Effects. Working paper, Columbia University.

Borenstein, Severin, and Garth Saloner. 2001. Economics and Electronic Commerce. *Journal of Economic Perspectives.* 15(1)

Brynjolfsson, E., Y. Hu, M. Rahman. 2009. Battle of retail channels: How product selection and geography drive cross-channel competition. *Management Science* 55(11), 1755-1765.

Ching, Andrew, and Masakazu Ishihara. 2012. Measuring the Informative and Persuasive Roles of Detailing on Prescribing Decisions. *Management Science* 58(7), 1374-1387.

Chintagunta, Pradeep, Junhong Chu, and Javier Cebollada. 2012. Quantifying Transaction Costs in Online/Off-line Grocery Channel Choice. *Marketing Science* 31(1), 96-114.

Choi, Jeonghye and David Bell. 2011. Preference Minorities and the Internet. *Journal of Marketing Research*, 58, 670 - 682.

Dagger, Tracey S., and Peter J. Danaher. 2014. Comparing the Effect of Store Remodeling on New and Existing Customers. *Journal of Marketing* 78(3), 62-80.

Danaher, Peter, and Harald van Heerde. 2014. Comparing the Effectiveness of Multiple Direct Marketing Efforts in Offline and Online Environments. Presentation at the 2014 INFORMS Marketing Science Conference, Atlanta.

Fang, Xiang, Surendra Singh, and Rohini Ahluwalia. 2007. An Examination of Different Explanations for the Mere Exposure Effect. *Journal of Consumer Research* 34, 97-103.

Forman, Chris, Anindya Ghose, and Avi Goldfarb. 2009. Competition between Local and Electronic Markets: How the benefit of buying online depends on where you live. *Management Science* 54(1), 47-57.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica* 52(3), 701-720.

Isaacson, Walter. 2009. *Steve Jobs*. Simon & Schuster, New York.

Joo, Mingyu, Kenneth C. Wilbur, Bo Cowgill, and Yi Zhu. 2014. Television Advertising and Online Search. *Management Science* 60(1), 56-73.

Kamakura, Wagner, and Gary Russell. 1993. Measuring brand value with scanner data. *International Journal of Marketing Research* 10, 9-22.

Kozinets, Robert V., John F. Sherry, Benet DeBerry-Spence, Adam Duhachek, Krittinee Nuttavuthisit, and Diana Storm. 2002. Themed flagship brand stores in the new millennium: theory, practice, prospects. *Journal of Retailing* 78, 17-29.

Lancaster, Tony. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391-413.

Lieber, Ethan and Chad Syverson. 2012. Online vs. Offline Competition. *Oxford Handbook of the Digital Economy*. Eds. M. Peitz and J. Waldfogel. Oxford Handbooks. Oxford & New York: Oxford University Press, 2012, 189-223.

Manjoo, Farhad. 2012. Welcome to the Microsoft Store. Slate.com. http://www.slate.com/articles/technology/technology/2012/04/microsoft_store_it_s_a_blatant_rip_off_of_the_apple_store_and_it_just_might_save_the_company_.html. Accessed June 4 2014.

Meyer, Robert. 2015. Editorial: A Field Guide to Publishing in an Era of Doubt. *Journal of Marketing Research* 52 (5), 577-579.

Narayanan, Sridhar, and Kirthi Kalyanam. 2015. Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. *Marketing Science* 34(3), 388-407.

Narayanan, Sridhar, Puneet Manchanda, and Pradeep Chintagunta. 2005. Temporal Differences in the Role of Marketing Communication in New Product Categories. *Journal of Marketing Research* 42(3), 278-290.

Oster, Emily. (2014). Unobservable selection and coefficient stability: Theory and evidence. Working paper, University of Chicago Booth School of Business.

Pauwels, Koen, Peter S.H. Leeflang, Marije L. Teerling, and K.R. Eelko Huizingh. 2011. Does online information drive offline revenues? Only for specific products and consumer segments. *Journal of Retailing.* 87(1), 1-17.

Pauwels, Koen, and Scott Neslin. 2015. Building with bricks and mortar: The revenue impact of opening physical stores in a multichannel environment. *Journal of Retailing* 91(2), 182-197.

Rapp, Adam, Thomas L. Baker, Daniel G. Bachrach, Jessica Ogilvie, and Lauren Skinner Beitelspacher. 2015. Perceived customer showrooming behavior and the effect of retail salesperson self-efficacy and performance. *Journal of Retailing* 91(2), 358-369.

Qian, Yi, Eric Anderson, and Duncan Simester. 2013. Multichannel Spillovers from a Factory Store. NBER Working Paper # 19176, Cambridge MA.

Santos Silva, J.M.C., and Silvana Tenreyro. 2006. The Log of Gravity. *Review of Economics and Statistics.* 88(4), 641-658.

Simon, Carol, and Mary Sullivan. 1993. The measurement and determinants of brand equity: A financial approach. *Marketing Science* 12, 28-52.

Sinai, Todd, and Joel Waldfogel. 2004. Geography and the Internet: Is the Internet a substitute or a complement for cities? *Journal of Urban Economics* 56, 1–24.

Simonov, Andrey, Chris Nosko, and Justin Rao. 2015. Competition and Crowd-Out for Brand Keywords in Sponsored Search. Working paper, University of Chicago.

Soysal, Gonca, and Alejandro Zentner. 2014. Measuring E-Commerce Concentration Effects When Product Popularity is Channel-Specific. Working paper, University of Texas at Dallas.

Telser, Lester G. 1964. Advertising and Competition. *Journal of Political Economy* 72(6), 537-562.

Verhoef, Peter C., P.K. Kannan, and J. Jeffrey Inman. 2015. From Multi-Channel to Omni-Channel Retailing: Introduction to the Special Issue on Multi-Channel Retailing. *Journal of Retailing* 91(2), 174-181.

Verhoef, Peter C., Scott Neslin, and Bjorn Vroomen. 2007. Multichannel customer management: Understanding the research-shopper phenomenon. *International Journal of Research in Marketing* 24, 129-148.

Waldfogel, J, Chen, L. 2006. Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers. *Journal of Industrial Economics*, 54(4), 425-449.

Ward, Michael R., and Michelle Morganosky. 2002. Consumer Acquisition of Product Information and Subsequent Purchase Channel Decisions. *Advances in Applied Microeconomics*, ISSN Book Series, 231-255.

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

Wooldridge, Jeffrey. 1997. Quasi-Likelihood Methods for Count Data. In *Handbook of Applied Econometrics*. Vol. 2. Eds. M.H. Pesaran and P. Schmidt. Oxford UK, Blackwell, 352-406.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, Cambridge MA.

Wu, Chunhua, Kangkang Wang, and Ting Zhu. 2015. Can Price Matching Defeat Showrooming? Working paper, University of British Columbia.

Zhang, Xubing. 2009. Retailers' Multichannel and Price Advertising Strategies. *Marketing Science* 28, 1080-1094.

Table 1: Summary statistics by brand-tract-month

	Observations	Mean	Std. Dev.	Min	Max
Dependent Variables					
Online sales ($)	462,922	26.04	127.67	0	11,419.9
Total sales ($)	748,312	50.48	172.02	0	11,419.9
# of sessions	525,943	3.699	12.44	0	649
# of customers browse	525,943	0.413	0.58	0	7
Number of new customers	380,477	0.058	0.236	0	3
For tracts with a first store opening in 25 miles during the data period, inclusive of tracts with all zero outcomes					
Online ($) before store opening	36,999	4.607	47.441	0	3,366
Online ($) after store opening	52,482	7.534	74.806	0	5,199.45
Online ($) before store opening – Without brand presence	33,336	2.925	41.218	0	3,366
Online ($) after store opening – Without brand presence	46,212	5.143	68.460	0	5,199.45
Online ($) before store opening – With brand presence	3,663	19.915	83.746	0	1,632
Online ($) after store opening – With brand presence	6,270	25.160	109.299	0	2369.8
Total ($) before store opening	36,999	7.646	57.193	0	3,366
Total ($) after store opening	52,482	22.848	117.810	0	5,244.5
# of sessions before store opening	36,999	.965	6.105	0	207
# of sessions after store opening	52,482	1.309	8.077	0	271
# of customers browse before store opening	36,999	0.117	0.349	0	4
# of customers browse after store opening	52,482	0.141	0.383	0	4
Online ($) by customers acquired before store opening	317,539	30.468	138.489	0	11,419.9
Online ($) by customers acquired after store opening	127,974	12.814	86.608	0	9,918
Fit and feel online sales – type 1	340,174	61.439	18.205	0	5,203.34
Other online sales – type 1	379,361	15.455	91.912	0	10,659.9
Fit and feel online sales – type 2	299,623	15.551	73.81	0	4,722
Other online sales – type 2	409,580	18.058	98.491	0	10,937.9
Covariates					
# of direct marketing activities	1,201,596	6.460	11.068	0	225
Store within 25 miles	1,201,596	0.644	0.479	0	1
Census tract population	1,201,596	4783.426	2139.535	0	37,452
Competitor store open in 10 miles	1,201,596	0.238	0.690	0	6
Competitor store open in 10 - 25 miles	1,201,596	0.453	0.963	0	7
Online sales ($): Brand A	152,922	37.927	176.501	0	11,420
Online sales ($): Brand B	106,470	23.885	124.725	0	5614.96
Online sales ($): Brand C	203,530	18.244	73.229	0	8731
Per-head online ($) by customers acquired before store open	316,065	20.810	103.973	0	11419.9
Per-head online ($) by customers acquired after store open	79,328	17.945	99.564	0	9918
Campaign content					
# of catalog campaigns	7,275				
# of email campaigns	24,652				
# of email campaigns about store openings	54				
# of emails sent	11,600,454				
# of email sent about store openings	1,198				

Table 2: Store openings and customer actions

	(1)	(2)	(3)	(4)
Dependent variable →	**Online sales**	**Total sales**	**# of sessions**	**# of customers browse website**
Store open within 25 miles	0.202*	0.807***	0.257***	0.126***
	(0.107)	(0.068)	(0.073)	(0.031)
# of direct marketing messages	0.042***	0.017***	0.020***	0.019***
	(0.001)	(0.001)	(0.001)	(0.0004)
Competition store open within 10 miles	-0.023	-0.016	-0.043**	-0.030***
	(0.026)	(0.012)	(0.022)	(0.008)
Competition store open within 10-25 miles	-0.023	-0.011	-0.008	-0.014**
	(0.022)	(0.009)	(0.015)	(0.006)
# of observations	462,922	748,312	525,943	525,943
# of tracts	22,044	35,634	25,045	25,045
Log pseudolikelihood	-19,170,413	-37,985,207	-1,474,750	-291,499

Unit of observation is the census brand-tract-month. Fixed effects Poisson regressions shown here. Robustness to various linear and non-linear specifications shown in the online appendix. Regressions include brand-location fixed effects, monthly fixed effects, and interactions between population and the month fixed effects. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 3: Store openings, brand presence, and customer activity

	(1)	(2)	(3)	(4)
Dependent variable →	**Online sales**	**Total sales**	**# of sessions**	**# of customers browse website**
Store open within 25 miles	0.486***	1.187***	0.546***	0.309***
	(0.141)	(0.093)	(0.094)	(0.042)
Store open within 25 miles x prior brand presence	-0.731***	-0.921***	-0.609***	-0.487***
	(0.194)	(0.123)	(0.142)	(0.060)
# of direct marketing messages	0.042***	0.017***	0.020***	0.019***
	(0.001)	(0.0007)	(0.001)	(0.0004)
Competition store open within 10 miles	-0.023	-0.015	-0.044**	-0.034***
	(0.026)	(0.012)	(0.022)	(0.008)
Competition store open within 10-25 miles	-0.022	-0.010	-0.007	-0.014**
	(0.022)	(0.009)	(0.015)	(0.006)
# of observations	462,922	748,312	525,943	525,943
# of tracts	22,044	35,634	25,045	25,045
Log pseudolikelihood	-18,496,230	-37,605,611	-1,474,018	-291,447
Sum of first two rows is significant with 95% confidence	Yes-Negative	Yes-Positive	No	Yes-Negative

Unit of observation is the census brand-tract-month. Fixed effects Poisson regressions shown here. Robustness to various linear and non-linear specifications shown in the online appendix. Regressions include brand-location fixed effects, monthly fixed effects, and interactions between population and the month fixed effects. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 4: Customers acquired after local store opening drive online sales increase

Dependent variable →	(1) Online sales by (old) customers acquired *before* store opening	(2) Online sales by (new) customers acquired *after* store opening
Store open within 25 miles	-0.047	23.639***
	(0.161)	(0.180)
Store open within 25 miles x prior brand presence	-0.139	-0.454
	(0.210)	(0.305)
# of direct marketing messages	0.037***	0.052***
	(0.001)	(0.003)
Competition store open within 10 miles	-0.023	-0.010
	(0.028)	(0.051)
Competition store open within 10-25 miles	-0.031	0.052
	(0.024)	(0.044)
# of Observations	317,539	127,974
# of tracts	15,121	6,094
Log pseudolikelihood	-14,144,186	-3,273,943
Sum of first two rows is significant with 95% confidence	No	Yes – Positive

Unit of observation is the census brand-tract-month. Fixed effects Poisson regressions shown. Robustness in the online appendix. Regressions include brand-location fixed effects, monthly fixed effects, and interactions between population and the month fixed effects. Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: In-store behavior does not seem to drive the observed changes in online behavior

	Fit and feel vs. other products		Fit and feel definition 2		Fit and feel definition 3		Returns
Dependent variable →	**(1) Fit and feel sales**	**(2) Other sales**	**(3) Fit and feel sales**	**(4) Other sales**	**(5) Fit and feel sales**	**(6) Other sales**	**(7) Return ($) / Online sales($)**
Store open within 25 miles	0.507***	0.432***	0.435**	0.539***	0.331*	0.587***	-0.377**
	(0.158)	(0.149)	(0.169)	(0.144)	(0.186)	(0.138)	(0.171)
Store open within 25 miles x prior brand presence	-0.726***	-0.729***	-0.606***	-0.870***	-0.670***	-0.760***	0.032
	(0.217)	(0.224)	(0.225)	(0.215)	(0.258)	(0.199)	(0.236)
# of direct marketing messages	0.043***	0.040***	0.042***	0.041***	0.044***	0.040***	0.004***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Competition store open within 10 miles	-0.044*	0.016	-0.053**	0.008	-0.066**	0.003	0.073***
	(0.025)	(0.044)	(0.026)	(0.035)	(0.028)	(0.031)	(0.027)
Competition store open within 10-25 miles	-0.021	-0.028	-0.017	-0.029	-0.018	-0.026	-0.015
	(0.023)	(0.031)	(0.023)	(0.026)	(0.024)	(0.024)	(0.019)
# of Observations	387,129	309,662	340,174	379,361	299,623	409,580	38,795
# of tracts	18,435	14,746	16,199	18,065	14,268	19,504	7,839
Log pseudolikelihood	-12,478,966	-8,272,775	-10,230,634	-10,570,793	-7,822,124	-12,841,255	-12,200.5
Sum of first two rows is significant with 95% confidence	Yes – Negative	Yes – Negative	No	No	Yes – Negative	No	Yes – Negative

Unit of observation is the census brand-tract-month. Fixed effects Poisson regressions shown here. Robustness of fit and feel results to various linear and non-linear specifications shown in the online appendix. Regressions include brand-location fixed effects, monthly fixed effects, and interactions between population and the month fixed effects. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

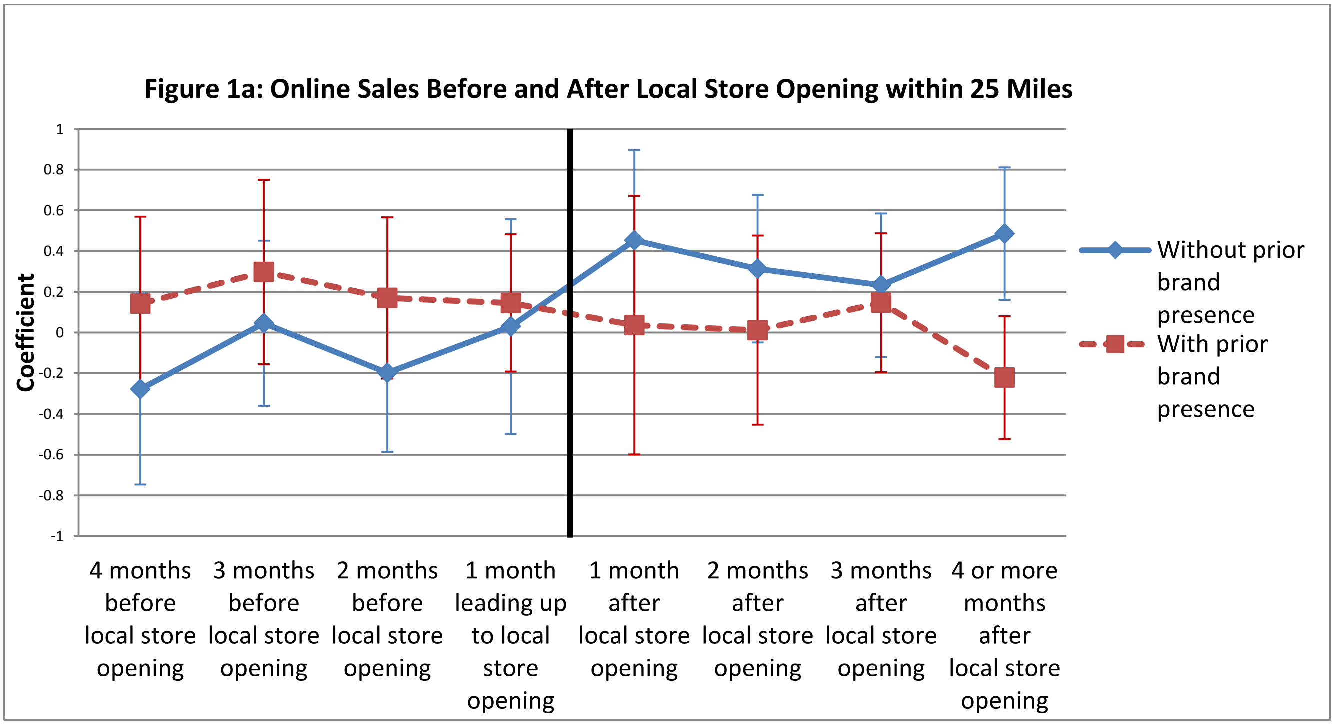
Figure 1a: Online Sales Before and After Local Store Opening within 25 Miles
Coefficient
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
4 months before local store opening
3 months before local store opening
2 months before local store opening
1 month leading up to local store opening
1 month after local store opening
2 months after local store opening
3 months after local store opening
4 or more months after local store opening
Without prior brand presence
With prior brand presence

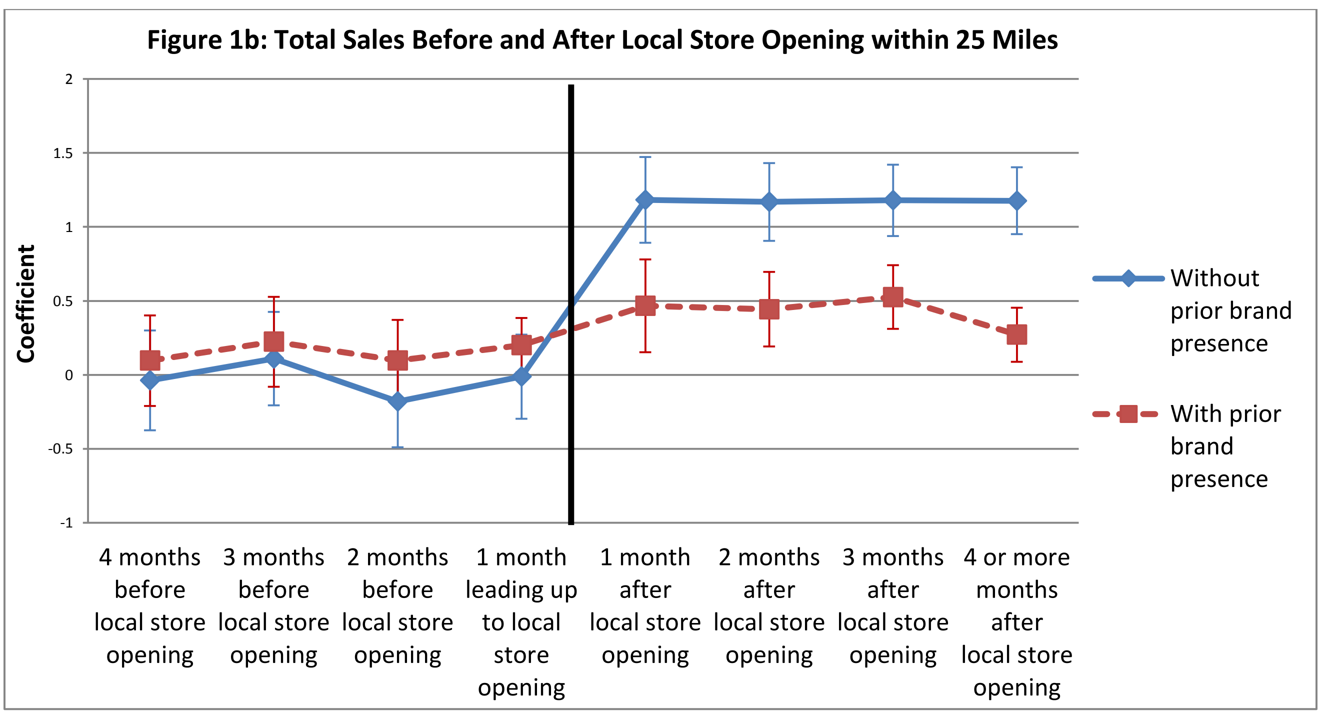
Figure 1b: Total Sales Before and After Local Store Opening within 25 Miles
Coefficient
2
1.5
1
0.5
0
-0.5
-1
4 months before local store opening
3 months before local store opening
2 months before local store opening
1 month leading up to local store opening
1 month after local store opening
2 months after local store opening
3 months after local store opening
4 or more months after local store opening
Without prior brand presence
With prior brand presence

Figure 2: Newly Acquired Customers Before and After Local Store Opening within 25 Miles