Cross-channel competition and complementarities in US retail*

Hiroki Saruya Michael Sullivan

Department of Economics, Yale University

October 4, 2022

Abstract

We estimate the effects of brick-and-mortar retail stores on nearby consumers' online spending using data on US store locations and internet usage. Our empirical approach uses a rich set of internet usage variables to proxy for unobserved taste characteristics. It also exploits variables characterizing the demographic profiles of consumers' neighbourhoods to address endogeneity problems arising from systematic differences in shopping tastes across regions. Our estimates for the 2007–2008 time period imply that a multichannel retailer's online sales decrease by 1.1–3.8% on average across categories with the addition of a rival offline store, whereas these sales increase by 7.1–32.3% when the retailer adds a store of its own. We attribute this finding to the presence of cross-channel complementarities that exceed standard cannibalization effects. Additionally, our estimated effects of offline locations on rivals' online sales vary across categories. We suggest showrooming effects as a possible source of this heterogeneity.

*Email addresses: hiroki.saruya@yale.edu and m.r.sullivan@yale.edu. We thank Katja Seim and participants at the Yale School of Management Quantitative Marketing internal seminar series for helpful feedback on this project. Michael Sullivan is supported in part by funding from the Social Sciences and Humanities Research Council.

1 Introduction

Contemporary retail markets feature competition between retailers that operate exclusively offline, exclusively online, and both offline and online. The merits of these channel choices depend on the relationship between the offline retail environment—i.e., retailers' brick-and-mortar storefronts—and online spending within a market. This relationship is multi-dimensional: offline stores could present heterogeneous benefits and harms to online stores across different pairs of offline and online retailers and across different types of products. Our study estimates the effects of offline stores on online spending across retailers and retailing categories.

We identify several relevant effects of offline stores on online spending.[1] First, offline stores may lower rival retailers' online sales on account of *competitive* effects. The strength of these effects in part determines the extent to which online and offline retail markets are connected and, consequently, the strategic importance of offline stores in e-commerce. Cannibalization is a particularly notable competitive effect: a multichannel retailer's offline store may steal business from its online store, thereby weakening the retailer's incentive to open brick-and-mortar stores. Cannibalization may be offset by various positive effects of a retailer's offline stores on its online sales. An offline store, for example, may bolster its retailer's online sales by improving awareness or public opinion of the retailer. Additionally, an offline store may complement its corresponding online store by offering in-person customer service for products purchased online, allowing in-person returns of items purchased online, by offering pick-up of products purchased online, or by strengthening the retailer's logistical operations in the offline store's region. We collectively refer to effects that make a retailer's offline stores more favourable for its online sales than rival retailers' offline stores as *cross-channel complementarities*. Cross-channel complementarities and cannibalization are competing forces in a horse race determining whether a retailer's offline store increases its online sales.

Despite our emphasis on competitive effects above, offline stores need not exert solely negative effects on rivals' online sales; offline stores may expand the market in a retailing category and ultimately help rival online retailers on account of showrooming effects. The term *showrooming effect* describes a benefit that a retailer derives from informative or promotional services offered by rival retailers. To illustrate, a bookstore may invest in informative services by installing product displays, allowing visitors to read unpurchased books in the store, and training staff to educate visitors about new book releases. This investment benefits online book retailers whose product offerings overlap with those of the offline retailer. An online retailer may choose to contain its

[1] We interchangeably use the terms "offline stores" and "brick-and-mortar stores" in our paper.

costs by investing little in informative services while freeriding on the offline retailer's informative services and undercutting the offline retailer's prices. This behaviour discourages the offline retailer from offering informative services to begin with. The informative services that we associate with the showrooming effect are those that are broadly relevant to a product category (e.g., campaigns promoting reading or books sold by multiple retailers) or that are relevant to products sold by multiple competing retailers. Informative services that narrowly affect a retailer's exclusive product line (e.g., fitting services for a clothing item sold exclusively by the retailer) are instead captured by our concept of cross-channel complementarities.

The empirical relevance of showrooming effects is important to understand due to their implications for retail strategy and policy. Strong showrooming effects encourage retailers to develop exclusive product lines for which they can provide informative services without being undercut on price by freeriding competitors. The showrooming effect is also a popular motivation for minimum resale price maintenance (RPM), by which a manufacturer prohibits resale of its products below a certain price. Minimum RPM allows a retailer to invest in informative services without fear that rivals who free-ride on these services will undercut the retailer's prices.

Our paper empirically assesses the relationship between a consumer's offline retail environment and the consumer's online spending. We particularly focus on the roles played by competitive effects, cross-channel complementarities , and showrooming effects. The data that we use include online purchasing records for a large panel of US consumers in 2007–2018 and data characterizing the universe of US business locations in the same time period. Our inquiry begins with a discussion of empirical relationships between the offline retail environments facing consumers and these consumers' online expenditures. These empirical relationships need not reflect causal relationships between the offline retail environment and online purchasing decisions due to the presence of unobservable consumer, retailer, and market characteristics that induce spurious correlations between the offline stores nearby a consumer and the consumer's expenditures on e-commerce. Bookstores, for example, may choose to open locations near consumers who enjoy reading. This would induce a correlation between store locations and consumers' online spending even in the absence of a causal relationship between offline bookstores and online spending on books. Also, contextual network effects—that is, effects of the local population's characteristics on a consumer's spending—induce an endogeneity problem; store counts and online spending may correlate partly on account of the fact that book retailers, for example, open stores in regions with many active readers, and that people in these regions buy more books to fit in with their peers. One of our paper's primary contributions is the development of a framework for estimating the causal relationship between the offline retail environment and online spending in the presence of the endogeneity problems enumerated above. Our approach involves (i) using a rich set of consumer characteristics that proxy for unobserved taste characteristics and (ii) modelling region-level unobservables using a combination of fixed effects and the local demographic profile, which both shapes contextual network effects and controls for unobserved shopping tastes of regions' residents. Although the relationship between the offline retail environment and online spending has implications for the structure of retail markets, we leave the investigation of these implications to future research.

We now summarize our main findings. First, rival stores generally decrease spending at a retailer's online store. We summarize the effects of rival brick-and-mortar stores on a retailer's online sales using the average percentage change in spending at the retailer's store when a rival store opens within 20km of the consumer. In the 2007–2008 time period, these rival effects range from 2.1% to -3.0% across retailing categories when Amazon is included in the average and from -1.1% and -3.8% when Amazon is excluded. In general, we find that Amazon's sales are less negatively affected by rivals' offline stores than multichannel retailers' online sales. This pattern could be explained by showrooming: since Amazon does not offer offline informative services, it can sell products at prices below its multichannel competitors while freeriding on these competitors' offline informative services. More nearby offline stores could thereby boost Amazon's sales. This effect does not apply to other multichannel retailers if they do not charge lower prices for online purchases than for in-store purchases. We find that offline bookstores have an especially large positive effect on book sales on Amazon, which aligns with our expectation that the books category is especially prone to showrooming effects. The fact that Amazon has relatively positive rival effects could also indicate that multichannel retailers are closer substitutes to each other than they are to Amazon.

A robust pattern in our findings across retailers is that a multichannel retailer's own offline stores boost its online sales. We measure effects of a multichannel retailer's own offline stores on its online sales as the average percentage change in spending at the retailer's online store from placing an additional one of the retailer's own offline stores in the consumer's vicinity. Our estimated measures for 2007–2008 range from 7.1% to 32.3% across categories. Our results for 2017–2018 are qualitatively similar to those for 2007–2008, but they are less precise in part because of the decreased coverage of our internet usage panel in the 2017–2018 time period.

The primary contribution of our paper is the evaluation of heterogeneity in the effects of offline stores on online spending across categories and retailers. Most papers in the extant literature on online/offline retail interactions focus on a single retailer or do not distinguish effects between retailers. To the best of our knowledge, our work is the first to empirically document heterogeneous effects of offline store presence on the sales of own (positive) and rival (negative) online channel. Our paper also demonstrates the importance of addressing endogeneity problems introduced by consumer and retailer location decisions and by contextual network effects. Descriptive analyses of our data suggest that our approaches to addressing these problems correct bias in the direction that we predict from a consideration of economic theory.

1.1 Related literature

Our paper joins a literature analyzing the relationship between offline retail and online sales. We distinguish our study from others in several ways. First, we study competition between multichannel retailers rather than focus on a single retailer's offline and online operations, as is done by much of the extant literature. We also use several empirical techniques for estimating the relationship between offline retail presence and online spending that are novel in the context of the literature on online/offline retail substitution.

Earlier studies on cross-channel competition (e.g., Goolsbee [2001], Sinai and Waldfogel [2004], Forman et al. [2009], Brynjolfsson et al. [2009]) document evidence for channel substitution, the degree of which depends on the nature of products, local market characteristics, and customers' proximity to store locations. These studies do not, however, distinguish multichannel retailers from single-channel retailers.[2] More recent papers study substitution between a particular retailer's online and offline retail channels in the context of apparel and home furnishings (Avery et al. 2012, Wang and Goldfarb 2017, Shriver and Bollinger 2020), eyewear (Bell et al. 2018), and groceries (Chintagunta et al. 2012, Pozzi 2013). Many of these papers emphasize both cannibalization and cross-channel complementarities (e.g., offline stores' effects on retailer awareness and their provision of information about fit-and-feel product attributes), and find that the latter effect is economically significant. Our study complements those listed above by analyzing consumer spending across multiple online stores and product categories to separately evaluate various effects of offline stores on online sales. Data on multiple retailers and multiple product categories are useful here because, first, a rival's offline store has a showrooming effect but not a brand awareness effect on a retailer's online sales. Additionally, the scope for showrooming effects differs across product categories.

Dolfen et al. [2019] estimate the relationship between offline retail presence (in their

[2]Prince [2007] measures the elasticity of demand for computers at online retailers with respect to offline price and argues that the cross-price elasticity increased following the rise in multichannel operations.

case, distance to offline stores) and online spending using payment card data in analyzing e-commerce's welfare effects. Our study differs from theirs in that (i) we study absolute effects of offline stores on online spending as opposed to the share of spending that occurs online; (ii) we focus on endogeneity concerns related to retailer location choices and unobserved regional taste shocks; and (iii) we emphasize differential effects across stores owned by different firms and across retailing categories. Quan and Williams [2018] also investigate the welfare effects of online retail and find that the welfare gains crucially depend on the heterogeneity of consumer tastes and offline retail environment across geographies.

Our paper also relates to the literature on showrooming. The literature on the economic theory of showrooming—which includes Jing [2018], Kuksov and Liao [2018], and Mehra et al. [2018]—studies the effect of showrooming on brick-and-mortar retailers' profits but pays little attention to the online sales of multichannel retailers. Carlton and Chevalier [2001] find evidence that manufacturers internalize freeriding by online retailers in their distribution and pricing strategies. More recently, Goetz et al. [2020] find that bookstore closures in Germany in the 2010s were associated with decreases in overall book sales. Our study complements Goetz et al. [2020] by comparing the extent of the showrooming effect in the book category with its extent in other retail categories. See MacKay and Smith [2014] for a summary of the debate surrounding minimum resale price maintenance, which has emphasized the showrooming effect and retailers' incentives to invest in demand-enhancing services. MacKay and Smith [2014] also provide empirical evidence on the effects of minimum resale price maintenance.

Our study also relates to a wider literature on the rise of e-commerce as a leading retail channel. We refer the reader to Hortaçsu and Syverson [2015] and Dolfen et al. [2019] for analysis of online retail's evolution in the United States.

We conclude our literature review by comparing our approach to endogeneity problems to related approaches. Our approach involves exploiting demographic variation across localities, and is related to the Waldfogel instruments defined by Berry and Haile [2016] after Waldfogel [2003]. The main idea underlying the Waldfogel instruments is that local demographics often shape consumers' choice sets.[3] This approach is a novelty in the literature on online/offline retail substitution, with the exception of Goetz et al. [2020]'s use of local population as an instrument for the number of offline bookstores in a regression with sales as the outcome. Our approach differs from that of Goetz et al. [2020] in that we use local demographics as controls rather than instruments. Last, we account for unobserved individual preferences using an approach similar to the proxy

[3] This idea also appears in Waldfogel [2008], which examines the relationship between local demographic profiles and the sorts of restaurants that operate in areas of the US.

or "replacement functions" approach—see Heckman and Vytlacil [2007]—which has extensively been used for estimating production functions; see, for example, Olley and Pakes [1996], Levinsohn and Petrin [2003], Ackerberg et al. [2015], and Gandhi et al. [2020].

2 Data

Our main two data sources are the Comscore Web Behavior Database and the Data Axle (formerly Infogroup) database, both of which we access through Wharton Research Data Services (WRDS). The Comscore data are a panel of US consumers' online browsing and transactions records. We have access to four years of these data: 2007, 2008, 2017, and 2018. Throughout the paper, we divide our sample into two periods: 2007–2008 and 2017–2018. Because there is limited overlap in the identities of the web users in the Comscore panel across distinct years of the data, we define a panelist as an individual/year pair. The sample we study includes 147,852 panelists in the 2007–2008 time period and 172,615 panelists in the 2017–2018 time period. As discussed later in this section, we focus on the 2007–2008 time period because the panel's coverage of online transactions seems to have decreased between our two time periods. We observe various personal characteristics of each panelist, the web domains that each panelist visits over the course of the year, and the online transactions made by the panelist. The observed demographic characteristics include: the age, education, and race of the head of panelist's household; the presence of a child in the panelist's household; household income; household size; and the ZIP code of the household's residence. For each transaction, we observe the name of the product that the panelist purchased; the website on which the transaction occurred; the listed price of the product; the total dollar value of the panelist's transaction, including taxes and discounts; and the category of the purchased product (e.g., books, electronics, office supplies, videos, etc.). De Los Santos et al. [2012] find that the Comscore Web Behavior Database was largely representative of online buyers in the United States via a comparison of Comscore's 2002 and 2004 datasets with the Internet and Computer Use Supplement of the Current Population Survey (CPS) and an additional survey conducted by a market research company. In Online Appendix O.1, we compare demographic profile of panelists included in the 2007 year of the Comscore data with CPS estimates and arrive at the same conclusion.[4]

[4]We find that the Comscore panel features a similar distribution of age and broadband internet adoption as the population of United States internet users, and a similar distribution of household income and region of residence as the overall population of United States households. Like De Los Santos et al. [2012], we find that Comscore over-samples Hispanic people relative to the share of Hispanic internet users reported by the CPS Computer and Internet Supplement. We additionally find that Comscore over-samples white people and under-samples Asian people.

We limit our attention to large cross-category retailers and several specialized retailing categories that are well represented in our sample. The large cross-category retailers that we study are Walmart, Costco, Target, and Amazon, and the specialized categories we analyze are books, office supplies, and electronics.[5] Within each category, we analyze retailer-specific sales and store counts for a few large retailers while aggregating together other retailers. Table 1 lists the large offline retailers on which our analysis focuses. We analyze offline stores not indicated by Table 1 together as a grouping of "other" stores within each category. We chose the offline retailers listed by Table 1 based on an inspection of retailers' national store counts, which Appendix A reports; in particular, we chose retailers that appear within the top three retailers in their respective categories for at least one year, excluding retailers that specialize in a narrow subcategories of their respective retailing categories (e.g., we do not include a large religious book retailer, Family Christian Book Store, in our analysis). Although Office Depot and Office Max merged in 2013, the merged firm—which goes by the name Office Depot—continued operating offline stores under each of the Office Depot and Office Max brands. We consider these brands as distinct offline retailers in analyzing the office supplies category in 2017–2018.

We include the online stores associated with each offline retailer in our analysis (e.g., we analyze sales at `staples.com` because we include Staples as an offline retail chain). We also study online sales at Amazon within each retailing category.[6] We only include sales within the product category in question in our analysis. When studying electronics stores, for example, we do not include products other than electronics (e.g., computer bags or office supplies). We include sales of products in all categories when studying cross-category retailers.

Tables 2 and 3 describe our category-specific transactions data. A comparison of the first panels in these tables shows that spending at Amazon increases almost threefold between our two time periods. Although Amazon held a strongest market position in the book and electronics categories during the first time period, it faced competitors that boasted sales of the same order of magnitude as its own in these categories. By 2017–2018, however, Amazon's sales far exceeded its rivals across categories.

[5]We identify offline stores in the book and office supplies categories by querying the Infogroup data for business locations with the following six-digit Standard Industrial Classification (SIC) codes: 594201 ("Book Stores") for books and 594301 ("Office Supplies") for office supplies. For electronics, we query for stores with the following four-digit SIC codes: 5731 ("Radio, Television, and Consumer Electronics Stores"), 5734 ("Computer and Computer Software Stores"), and 5946 ("Camera and Photographic Supply Stores").

[6]In the books category, we do not analyze online stores for Borders and Waldenbooks because our data include no online sales for these retailers in 2007. Similarly, we exclude `booksamillion.com` from our analysis of the 2017–2018 data because these data include no transactions at `booksamillion.com`. We exclude `radioshack.com` from our analysis of the 2017–2018 period for the same reason. Office Depot ceased operating the `officemax.com` online store following its merger with Office Max. This explains why `officemax.com` does not appear in our analysis of the 2017–2018 time period.

Table 1: Large offline retailers by retailing category

Category	Retailers 2007–2008	Retailers 2017–2018
Cross-category	Walmart Target Costco	Walmart Target Costco
Books	Barnes & Noble Books-a-Million Waldenbooks Borders	Barnes & Noble Books-a-Million
Office supplies	Staples Office Depot Office Max	Staples Office Depot Office Max
Electronics	Best Buy Circuit City Radio Shack Apple	Best Buy Radio Shack Apple

Note: Borders, Waldenbooks, and Circuit City each closed all of their brick-and-mortar locations between the 2007–2008 and 2017–2018 time periods. We therefore exclude these retailers from our analysis of the 2017–2018 data.

We focus on the 2007–2008 time period because the Comscore panel's coverage of transactions seems to have decreased between 2008 and 2017. In 2017 and 2018, for example, there were respectively 92 and 101 million members of Amazon Prime (Amazon's premium subscription service) in the United States.[7] Additionally, in an October 2017 survey, 92% of Prime members reported ordering from Amazon at least once a month, as did 61% of survey respondents who did not subscribe to Prime.[8] Despite these facts, we find that only 18.3% of Comscore panelists made a transaction at Amazon in the 2017–2018 time period. Considering together the facts that (i) most United States consumers used Amazon in 2017–2018, (ii) Amazon reported only 81 million active accounts (Prime or non-Prime) in 2008,[9] and (iii) the share of Comscore panelists using Amazon increased only from 14.3% in 2007–2008 to 18.3% in 2017–2018 constitute evidence that the Comscore panel's coverage of e-commerce transactions markedly decreased between our time periods. A leading hypothesis for why Comscore's coverage decreased is that internet usage shifted from personal computers, which are tracked by Comscore, to other sorts of devices (e.g., smartphones and tablets) that are not tracked by the Comscore panel. Our findings for 2017–2018 are qualitatively similar to those for 2007–2008, although our estimates for the later time

[7]See here: https://www.digitalcommerce360.com/2019/07/11/82-of-us-households-have-a-amazon-prime-membership/.

[8]See here: https://www.forbes.com/sites/louiscolumbus/2018/03/04/10-charts-that-will-change-your-perspective-of-amazon-primes-growth/?sh=42b198b43fee.

[9]See here: https://www.nytimes.com/2008/10/12/business/12giants.html.

period are typically less precise. This reduction in precision reflects that observed spending at multichannel retailers was lower in this period due in part to reduced coverage of the Comscore panel.

Table 2: Summary of consumer panel, 2007–2008

(a) Cross-category retailers

Store	**Unconditional** Spending per panelist	Pct. spending	**Conditional on positive spending** Average spending	Median spending	95th percentile spending
amazon	18.51	14.32	129	53	461
costco.com	3.27	0.54	607	280	1989
target.com	3.84	3.47	111	60	360
walmart.com	7.16	6.07	118	50	424

(b) Bookstores

Store	**Unconditional** Spending per panelist	Pct. spending	**Conditional on positive spending** Average spending	Median spending	95th percentile spending
amazon	6.90	8.28	83	39	290
barnesandnoble.com	1.08	1.83	59	30	191
booksamillion.com	0.07	0.13	55	32	182

(c) Electronics

Store	**Unconditional** Spending per panelist	Pct. spending	**Conditional on positive spending** Average spending	Median spending	95th percentile spending
amazon	4.12	1.83	226	109	755
bestbuy.com	2.75	0.88	311	180	1000
circuitcity.com	2.47	0.76	323	180	1255

(d) Office supplies

Store	**Unconditional** Spending per panelist	Pct. spending	**Conditional on positive spending** Average spending	Median spending	95th percentile spending
amazon	0.08	0.10	83	41	330
officedepot.com	4.36	0.57	768	279	3277
officemax.com	0.39	0.11	350	169	1543
staples.com	5.47	0.84	653	233	2888

Note: The "Spending per consumer" column reports the total dollar amount spent at the store in the panel divided by the number of panelists. The "Pct. spending" column reports the percentage of panelists who make at least one purchase from the indicated store. The "Average spending," "Median spending," and "95th percentile spending" columns describe the distribution of spending among panelists who make at least one purchase from the indicated store. When Amazon is included as an online store in panels (b) onward, only its transactions within the indicated category are included in the analysis.

Table 3: Summary of consumer panel, 2017–2018

(a) Cross-category retailers

Store	Unconditional		Conditional on positive spending		
	Spending per consumer	Pct. spending	Average spending	Median spending	95th percentile spending
amazon	52.47	18.30	287	109	1123
costco.com	2.23	0.54	414	140	1500
target.com	1.20	1.06	114	59	353
walmart.com	6.06	4.41	137	60	478

(b) Bookstores

Store	Unconditional		Conditional on positive spending		
	Spending per consumer	Pct. spending	Average spending	Median spending	95th percentile spending
amazon	4.68	4.95	95	45	339
barnesandnoble.com	0.15	0.30	52	29	187

(c) Electronics

Store	Unconditional		Conditional on positive spending		
	Spending per consumer	Pct. spending	Average spending	Median spending	95th percentile spending
amazon	15.00	7.96	188	76	698
bestbuy.com	2.74	0.67	408	200	1460

(d) Office supplies

Store	Unconditional		Conditional on positive spending		
	Spending per consumer	Pct. spending	Average spending	Median spending	95th percentile spending
amazon	2.09	1.85	113	56	390
officedepot.com	0.28	0.10	289	93	1151
staples.com	0.35	0.13	269	85	918

Note: See the notes for Table 2.

Although our main interest is in online spending, we use the Comscore browsing data to construct variables that we use as controls. The Comscore browsing data reports each web domain that a panelist visits and information about this visit including the time of the visit, the visit's duration, and the number of pages that the panelist visits within the domain. We use the Comscore browsing data to construct panelist-level variables that report the number of times that the panelist visits a web domain in each of several categories within the course of a year. The categories that we specify include, among others, information, sports, news, gaming, social media, weather, and finance.[10] Table 4 describes these control variables. Note that the Comscore data

[10]The complete list of categories is: adult, advertising services (e.g., domains of digital adver-

capture web browsing on personal computers and not mobile devices; the increase in mobile usage between our time periods may explain the decrease in the overall number of visits between our time periods. Additionally, some specialized domains (e.g., weather websites) experienced a decrease in usage that is partly attributable to the increase in significance of large digital platforms that offer a wide breadth of online services (e.g., Google and Apple offer weather services).

We also construct measures of the demographic profiles of the areas surrounding panelists. We construct these measures by computing, for each Comscore panelist, the average of demographic variables among Comscore panelists within 20km of this panelist. [11] The demographic variables that we use to construct our measures of local demographics are: an indicator for the panelist's household income exceeding $75,000; indicators for the panelist's racial background being white and being black; indicators for the panelist's head-of-household being less than 40 years of age and falling between 40 and 54 years of age; the panelist's household size; an indicator for the presence of children in the household; an indicator for the panelist being Hispanic; an indicator for the panelist having broadband internet access; and an indicator for the panelist being a college graduate. Note that we use consumer's own values of these demographic variables in addition to our measures of local demographics as controls in our empirical analysis.

Our other data source is the Data Axle (formerly Infogroup) database of firm locations. This dataset reports the locations of US businesses at an annual level for the years 1997 to 2018. Variables include business name, geographical coordinates, and an identifier of the business location's parent company. We use these data to compute the number of a retailer's locations within 20km of each Comscore panelist and panelists' minimum

tising platforms that a user may transiently visit upon clicking certain display ads), career (e.g., online recruitment services), finance (e.g., online banking and brokerages), gaming, government (i.e., websites of federal and state governments or governmental agencies), information (i.e., web domains that provide information to users including, for example, `wikipedia.org` and `mayoclinic.com`), malware, media, portals (e.g., Yahoo, Google, and MSN), retail, social media, video streaming, weather, webservices (i.e., domains for firms that provide hosting and design services to websites, e.g., `geocities.com`), dating websites, internet/wireless services' websites (e.g., `att.com`), news, sports, travel, career, download services (e.g., `limewire.com`, which was used for torrenting), and directories (e.g., `yelp.com`). Our procedure for categorizing websites begins by identifying the most popular sites in the Comscore browsing data by visits and unique users. In particular, for each of the time periods (i) January to February 2007, (ii) November to December 2008, (iii) January to February 2017, and (iv) November to December 2018, we construct a list of the top 500 sites by number of visits in the Comscore browsing data as well as a list of the top 500 sites by number of unique visitors in these data. We then concatenate these lists and drop duplicated sites from the combined list. For each site, we manually determine in which of the aforementioned categories the site belongs. Some sites are not well described by any of these categories, and we do not place these sites in any category. Additionally, we do not place sites that are not in our list in any category.

[11] Recall that the Comscore data reports each panelist's ZIP code of residence; we compute the measures described by the preceding sentence by averaging over Comscore panelists living in a ZIP code tabulation area whose centroid is within 20km of that in which the focal panelist resides.

Table 4: Number of sessions including web browsing in various categories, 2007–2008

	2007–2008		2017–2018	
Category	Median	90th percentile	Median	90th percentile
Adult	2	66	0	36
Advert	9	80	0	1
Career	1	15	0	5
Dating	0	5	0	0
Directory	0	1	0	2
Downloads	0	7	0	2
Finance	6	97	1	39
Gaming	2	106	0	41
Government	2	19	0	6
Info	19	93	3	39
Internet/wireless	4	58	0	10
Malware	6	150	0	4
Media	10	95	1	16
News	3	29	1	15
Portal	393	1577	88	697
Retail	42	223	7	98
Social Media	33	585	5	107
Sports	1	21	0	3
Travel	2	24	0	9
Video	8	99	6	133
Weather	2	67	0	2
Webservice	41	320	0	9
Total	2228	8063	476	3464

Note: This table reports the cross-panelist distribution of the number of times that a panelist visited a website in various categories. The “Total” row describes the distribution of the number of distinct visits to web domains made by panelists. Recall that a panelist is defined as a pair of a web user in the Comscore data and a calendar year.

distances to locations operated by each of the offline retailers that we analyze. Table 5 describes these variables. We see that, between 2007–2008 and 2017–2018, large cross-category retailers increased their offline retail presences; the number of bookstores near consumers fell;and Radio Shack’s offline retail presence dramatically decreased.

Table 5: Description of offline retail presence variables

(a) 2007–2008

Category	Store	# stores (20km)					Min. distance					# stores
		Mean	0.25	0.50	0.75	0.95	Mean	0.25	0.50	0.75	0.95	(total - 2007)
Cross-category	Costco	2.34	0	0	4	10	75.13	8.54	25.77	106.10	302.47	374
Cross-category	Target	7.48	0	4	12	24	33.41	3.89	7.44	22.96	84.41	1446
Cross-category	Walmart	8.33	2	6	12	23	8.66	3.06	5.34	10.02	25.82	3411
Books	Barnes	5.66	0	2	8	18	26.37	5.43	11.10	32.80	96.77	832
Books	Books-a-Million	0.51	0	0	0	3	474.97	32.06	134.02	504.68	2093.03	178
Books	Borders	4.35	0	2	6	19	45.23	6.30	14.29	56.97	186.00	660
Books	Waldenbooks	1.73	0	1	3	7	53.92	9.31	23.64	56.56	153.75	464
Books	Other	71.42	6	23	67	299	7.08	1.40	3.04	7.50	28.01	11251
Electronics	Apple	1.46	0	0	2	7	86.19	13.35	36.70	117.27	308.64	169
Electronics	Best Buy	4.95	0	2	8	17	23.75	5.20	9.59	27.47	89.86	851
Electronics	Circuit City	4.16	0	2	6	16	33.75	5.63	10.79	33.66	119.36	685
Electronics	Radio Shack	24.41	4	12	30	90	7.01	1.78	3.27	6.85	26.70	5095
Electronics	Other	157.68	10	47	161	694	4.90	0.92	1.94	4.57	20.80	23011
Office Supplies	Office Depot	7.11	0	4	11	28	26.09	4.24	9.31	31.15	96.82	1262
Office Supplies	Office Max	4.73	0	2	8	18	29.18	5.25	12.19	39.27	103.81	982
Office Supplies	Staples	10.63	0	2	11	44	41.76	3.93	9.05	33.37	127.92	1486
Office Supplies	Other	29.31	3	11	33	115	9.56	2.32	4.79	11.65	33.92	5543

(b) 2017–2018

Category	Store	# stores (20km)					Min. distance					# stores
		Mean	0.25	0.50	0.75	0.95	Mean	0.25	0.50	0.75	0.95	(total - 2017)
Cross-category	Costco	3.46	0	2	6	13	40.11	6.52	13.65	50.63	161.63	510
Cross-category	Target	10.89	2	6	16	37	15.11	3.16	5.54	13.35	66.92	1809
Cross-category	Walmart	12.72	4	10	18	35	6.74	2.49	4.16	7.37	19.94	4318
Books	Barnes	5.91	0	4	8	20	21.48	5.07	9.69	23.65	78.89	953
Books	Books-a-Million	0.57	0	0	0	3	190.94	21.52	48.98	149.75	803.83	238
Books	Other	64.96	6	24	65	239	7.76	1.59	3.36	7.83	31.02	8957
Electronics	Apple	2.62	0	1	4	11	53.67	9.45	19.97	65.44	209.02	274
Electronics	Best Buy	6.32	2	4	9	21	18.28	4.43	7.91	17.52	72.86	1024
Electronics	Radio Shack	2.56	0	1	3	11	47.08	10.55	26.39	53.14	132.64	1124
Electronics	Other	166.01	13	61	196	724	4.88	0.89	1.84	4.11	21.06	21188
Office Supplies	Office Depot	5.36	0	2	8	23	31.83	4.88	11.86	36.61	123.91	968
Office Supplies	Office Max	2.54	0	1	4	10	41.24	7.78	20.34	56.88	145.96	622
Office Supplies	Staples	9.74	0	4	9	37	36.65	4.05	8.27	24.43	104.62	1380
Office Supplies	Other	26.20	3	10	27	105	10.76	2.62	5.24	12.37	38.78	4313

3 Descriptive evidence

Before developing our approach to estimating the effects of offline stores on online spending, we describe empirical relationships between offline retail presence, online shopping, and the auxiliary variables that we use to address our endogeneity problems. In doing so, we aim to show that these empirical relationships admit multiple interpretations that we require econometric assumptions to disentangle. We additionally suggest how our auxiliary variables, which describe web browsing behaviour and the demographics of consumers' vicinities, help us to address our endogeneity problems.

To assess the empirical relationship between online spending and the offline retail environment, we estimate Nadaraya-Watson kernel regressions of online spending measures on the distance of a consumer from the nearest location of a particular retailer. Our estimands in these regressions are the nonparametric functions m_{sj} defined by

$$m_{sj}(d_{ij}) = \mathbb{E}[y_{is} \mid d_{ij}],$$

where y_{is} is a measure of consumer i's spending at the online store s and d_{ij} is consumer i's distance from a location of chain retailer j. In practice, we use an indicator for whether the consumer made any transaction at the store s in a particular category as our spending measure y_{is}, and we present results for a subset of online store/offline retailer pairs. We use the standard normal density as our kernel with a bandwidth selected to minimize the sum of squared prediction errors in leave-one-out cross validation. The dotted red bands in our figures display a 95% pointwise confidence interval around our estimates.[12]

We now turn to a subset of our results from the 2007–2008 time period that highlight the problems in drawing conclusions about the effects of offline stores on online sales from empirical relationships between these variables. For each plot, we display the estimated m_{sj} function over an interval ranging from 0–20km. Online Appendix O.2 provides results for additional pairs of stores and for additional categories. Table 5, which describes the distributions of the d_{ij} variables, is useful in interpreting these results. We first consider a regression of spending at Walmart's online store on the consumer's distance from the nearest Walmart store. Figure 1a displays the results; it shows that, among panelists within 10km of a Walmart store, panelists who are further away from a brick-and-mortar Walmart location tend to spend less at Walmart's online store. A consideration of this relationship raises several of the empirical challenges that we face in assessing the effect of Walmart's offline stores on its online

[12] We construct these confidence intervals using the $\hat{v}_{n,1}(x)$ asymptotic variance estimator analyzed in Chu et al. [2020].

sales. Indeed, we cannot naively conclude that Walmart's offline stores do not cannibalize its online sales based on Figure 1a because the relationship plotted in this figure reflects Walmart's choice of store locations, which may target areas in which people have high propensities for shopping at Walmart through both of the retailer's channels, in addition to the cross-channel complementarities that we discussed in our introduction. Although about 75% of panelists live within 10km of a Walmart location (see Table 5), the fact that our estimated relationship between distance and spending becomes upward sloping after 10km suggests the presence of effects of varying signs that mediate the overall empirical relationship between distance and spending. Our preferred empirical approach accounts for these effects in pursuit of the effect of an exogenous store opening on online sales.

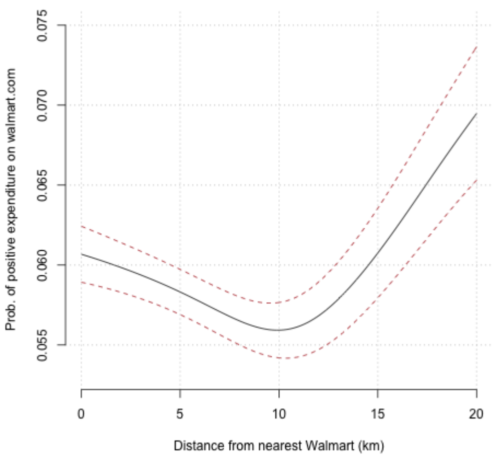
Our next regression evaluates the empirical relationship between spending at Walmart's online store and the consumer's distance from a location of Costco, which is one of Walmart's principal competitors. Figure 1b displays the results. We see that consumers who are further away from a Costco are more likely to purchase from Walmart's online store, which suggests a negative competitive effect of Costco's stores on Walmart's sales. This relationship, however, also depends on Costco's store location decisions—which depend on characteristics of nearby consumers that are relevant to their online shopping behaviour—in addition to potential showrooming effects.

Last, Figures 1c and 1d plot the results of regressions of Barnes and Noble's and of Amazon's online store's book sales, respectively, on the consumer's distance from a Barnes & Noble bookstore. These figures show a negative relationship between the consumer's distance from a Barnes & Noble store and each of (i) own and (ii) rival online sales; that is, Amazon has higher book sales among consumers who live nearby a Barnes & Noble store. This contrasts with the relationship plotted in Figure 1b, which shows that consumers living further from Costco are more likely to shop at Walmart's online store. The negative relationship displayed by Figure 1d could owe to a strong showrooming effect or the location of Barnes & Noble stores near consumers with high propensities to purchase books online.

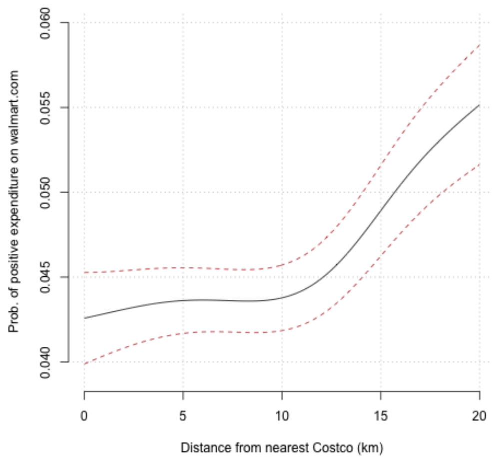
The evidence presented thus far admits several interpretations regarding the contributions of cross-channel complementarity, showrooming effects, competitive effects, and consumers' and retailers' location choices to the relationship between offline store presence and online shopping behaviour. Inference about the effects enumerated above therefore requires an alternative empirical approach.

The alternative empirical approach that we develop in the subsequent section exploits control variables that proxy unobserved components of consumers' tastes. One of our principal endogeneity problems involves dependencies between unobserved consumer tastes and the offline retail environment faced by the consumer. These dependencies

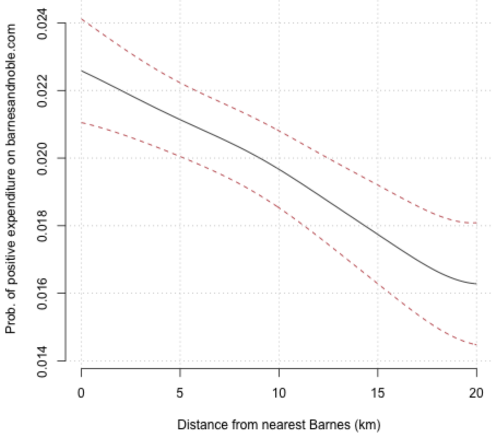
Figure 1: Selected regressions of online spending on distance from retailer, 2007–2008



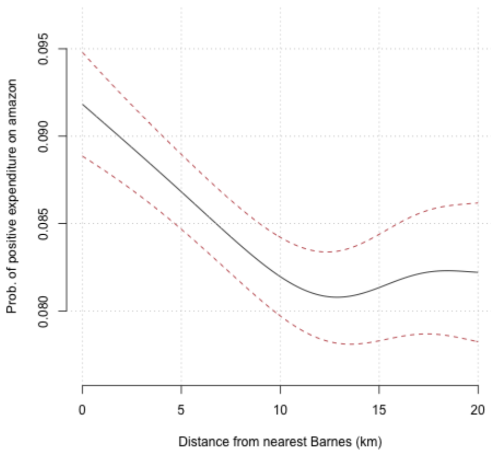
(a) Probability of transaction on `walmart.com` by distance from Walmart



(b) Probability of transaction on `walmart.com` by distance from Costco



(c) Probability of transaction on `barnesandnoble.com` by distance from Barnes & Noble



(d) Probability of transaction on Amazon by distance from Barnes & Noble

owe to both (i) the dependence of retailers' store location decisions on the geographical distribution of unobserved consumer tastes (e.g., retailers opening offline stores nearby consumers who would like to shop from those retailers) and (ii) the dependence of consumers' residential location decisions on the geographical distribution of stores (e.g., consumers may choose to live close to stores at which they like to shop). Both endogeneity problems confound an interpretation of empirical relationships such as those displayed by Figure 1 because they introduce unobserved consumer tastes as a shifter of both online spending and the offline retail presence in the consumer's vicinity. That is, the endogeneity problems mentioned above together constitute an omitted variables bias problem in which the omitted variables are unobserved consumer tastes. We attempt to overcome our problem by controlling for a rich set of variables that proxy unobserved aspects of consumer preferences.

The variables that we use to proxy for unobserved tastes are the internet usage variables described in Section 2. The idea underlying our use of these controls is that consumers reveal aspects of their interests, aesthetic preferences, and personalities through their choices of visited websites. To fix ideas, consider an endogeneity problem that arises when bookstores may open in areas that are known to have intellectually inclined populations. To the econometrician, intellectual inclination is a component of unobserved tastes. Intellectual inclination may reflect itself in a consumer's web browsing behaviour, e.g., through visits to websites with informational or literary content. Thus, controlling for a rich set of variables characterizing internet usage may amount to controlling for intellectual inclination, thereby resolving the endogeneity problem.

To determine whether our internet usage controls may help in overcoming our endogeneity problems, we regress indicators for whether a panelist made an online book purchase on the number of offline bookstores within 20km of the consumer, our internet usage variables, and various consumer characteristics. Table 6 provides the results. We note that, first the internet usage variables predict online book spending in reasonable ways. Visits to websites in the information and news categories—which are categories offering informative texts to read, just as books do—predict online book shopping. We also see that websites in the adult and gaming categories, which offer relatively little informative and literary content, are associated with less online book purchasing. Table 6 also reports a large change in estimated coefficient for the number of stores variable when the internet usage variables are omitted from the regression. This coefficient becomes markedly more positive. This suggests that including internet usage controls removes upward bias in the estimates of offline stores' effects on online spending stemming from the fact that bookstores locate near people who like buying books, whether online or offline.

We additionally take concern that consumers with the same unobserved taste characteristics may have different propensities for online shopping across regions due to region-level taste shifters. One reason for this concern is that regional differences in demographics may shift region-level tastes for retailers in the presence of network effects. Consumer tastes for retailers may depend on the characteristics of people in their neighbourhoods when consumption is driven by a desire to impress or fit in with one's peers; a person living in a wealthy neighbourhood may feel a stronger impulse to buy a designer jacket as a status symbol, for example, than a person living in

Table 6: Regression of online book shopping indicator on nearby stores and web use controls, 2007–2008

		Controls included		Controls excluded	
Variable type	Note	Estimate	t-statistic	Estimate	t-statistic
N. stores (log-transformed)		-0.05	-0.96	0.30	5.50
N. visits (hundreds)	Adult	-1.11	-11.02		
N. visits (hundreds)	Advert	-1.17	-7.75		
N. visits (hundreds)	Career	-1.23	-3.14		
N. visits (hundreds)	Finance	2.27	19.45		
N. visits (hundreds)	Gaming	-0.43	-7.17		
N. visits (hundreds)	Government	1.13	4.49		
N. visits (hundreds)	Info	4.85	28.32		
N. visits (hundreds)	Malware	-0.32	-7.58		
N. visits (hundreds)	Media	0.39	4.30		
N. visits (hundreds)	Other	0.02	4.60		
N. visits (hundreds)	Portal	0.21	17.19		
N. visits (hundreds)	Retail	3.04	48.89		
N. visits (hundreds)	Social Media	-0.09	-3.00		
N. visits (hundreds)	Video	-0.71	-6.71		
N. visits (hundreds)	Weather	0.15	1.99		
N. visits (hundreds)	Webservice	-0.56	-9.98		
N. visits (hundreds)	Dating	-0.20	-0.96		
N. visits (hundreds)	Internet Wireless	0.50	6.65		
N. visits (hundreds)	News	1.16	13.99		
N. visits (hundreds)	Sports	-0.21	-1.75		
N. visits (hundreds)	Travel	5.51	18.10		
N. visits (hundreds)	Downloads	-0.55	-4.82		
N. visits (hundreds)	Directory	9.88	2.40		

Note: The dependent variable is an indicator for whether the panelist ever purchased a book online. "N. stores (log-transformed)" is the number of stores within 20km of the panelist transformed by the mapping $x \mapsto \log(x+1)$. Both regressions also include the following variables as regressors, whose estimated coefficients are omitted from the table: an indicator for the panelist's household income exceeding $75,000, indicators for white and black racial backgrounds, indicators for the panelist being under 40 years of age and for being between 40 and 55 years of age, household size, an indicator for the presence of children in the panelist's household, an indicator for the panelist being Hispanic, an indicator for the panelist having broadband internet, an indicator for the panelist having a university degree, and year fixed effects. We run the regression on our full sample of panelists for the time period 2007–2008.

a deprived neighbourhood. Following the peer effects literature,[13] we use the term *contextual network effects* to refer to effects on a consumer's purchasing behaviour of the characteristics of nearby consumers. Given that the tastes of other people in the consumer's vicinity also affect retailers' decisions of whether to open offline stores in the consumer's vicinity, network effects introduce a spurious dependency between online spending and offline retail presence. The demographic profile of a consumer's

[13]We refer the reader to Section 7.3.1 in Jullien et al. [2021] for a discussion of contextual network effects.

neighbourhood could also predict a consumer's online shopping because the consumer's choice of neighbourhood reflects unobserved aspects of the consumer's tastes. Dependence between unobserved consumer tastes and local demographic profile introduce the same sort of endogeneity problem as contextual network effects.

We address the endogeneity problems posed by region-level taste shifters by controlling for the local demographic profile and by using census region fixed effects. We find strong evidence that the characteristics of consumers' neighbours influence their online purchasing behaviour, which suggests that either contextual network effects, or a relationship between unobserved consumer tastes for online shopping and neighbourhood choice are relevant. Our approach of controlling for regional demographics addresses the endogeneity problems posed by these phenomena.

We now consider evidence that local demographics influence online shopping behaviour. This evidence comprises results from regressions of spending at each of three large retailers' online stores on various demographic variables, on the number of the retailers' offline stores within 20km of the consumer, and the share of the population within 20km of the consumer with household incomes exceeding \$75,000. We run this regression on both our 2007–2008 dataset and our 2017–2018 dataset; in each case, we limit the influence of outliers by trimming observations with spending at the retailer's online store exceeding the 98th percentile of spending at this retailer conditional on positive spending. Table 7 provides the results. We see that, conditional on the number of offline stores operated by a retailer and consumer demographics, consumers in higher-income areas spend significantly more at Costco's online store, moderately more at Target's online store, and less at Walmart's online store. We offer two interpretations of these results. Target and especially Costco target consumers with higher socio-economic status relative to Walmart, and thus the effect of the high-income share of consumers' localities could reflect that consumers in higher-income areas may prefer shopping at stores that are more upscale than Walmart to impress or fit in with their well-to-do neighbours. These consumers' preference for higher-income areas could also relate to their unobserved shopping tastes for reasons other than network externalities (i.e., a consumer's preference for a glamorous aesthetic may lead the consumer to both live in a high-income area and to shop at relatively upscale retailers).

Table 7: Regressions of spending at cross-category spending on high income share

	Spending					
	costco.com: 2007/08	target.com: 2007/08	walmart.com: 2007/08	costco.com: 2017/18	target.com: 2017/18	walmart.com: 2017/
	(1)	(2)	(3)	(4)	(5)	(6)
N. stores (log-transformed)	2.551***	0.040	−1.319***	0.797***	0.162***	−0.585***
	(0.159)	(0.059)	(0.113)	(0.105)	(0.030)	(0.094)
High income	0.659*	0.620***	−0.237	1.154***	0.277***	0.309
	(0.352)	(0.165)	(0.248)	(0.238)	(0.082)	(0.210)
Race: White	0.710	0.825**	1.140*	−0.497**	0.443***	1.898***
	(0.863)	(0.401)	(0.597)	(0.251)	(0.087)	(0.220)
Race: Black	−0.633	−0.181	0.297	−1.473***	−0.290***	−0.465*
	(0.995)	(0.464)	(0.697)	(0.300)	(0.104)	(0.265)
Hispanic	−0.713**	−1.362***	−1.806***	−0.360	−0.044	−0.048
	(0.318)	(0.149)	(0.223)	(0.266)	(0.092)	(0.234)
Broadband internet	1.166**	1.590***	2.372***	1.385*	1.039***	4.333***
	(0.455)	(0.214)	(0.321)	(0.712)	(0.246)	(0.629)
College graduate	0.436	0.178	−0.711*	2.643***	0.624***	1.211***
	(0.540)	(0.254)	(0.381)	(0.275)	(0.095)	(0.242)
High income (average)	2.667***	0.632**	−0.769*	1.541***	0.449**	−1.279***
	(0.638)	(0.300)	(0.447)	(0.509)	(0.176)	(0.444)
Mean dep. var.	2.51	3.19	5.70	1.76	0.98	4.76
Observations	147,836	147,749	147,673	172,596	172,578	172,462

Note: “N. stores (log-transformed)” is the log of one plus the number of offline stores operated by the retailer indicated by the column header within 20km of the consumer. “High income (average)” is the share of people within 20km of the consumer that have household incomes exceeding $75,000. We include year fixed effects; indicators for the panelist’s head of household being under 40 years of age and between 40 and 54 years of age; an an indicator for the presence of children in the panelist’s household. We omit, however, these regressors’ estimated coefficients from the table

Table 8: Dependence of offline retail environment on local demographics, 2007–2008

	Costco	Target	Walmart
	(1)	(2)	(3)
High income (average)	0.793***	1.015***	0.536***
	(0.008)	(0.011)	(0.009)
Race: White (average)	−3.059***	−2.634***	−0.799***
	(0.023)	(0.029)	(0.023)
Race: Black (average)	−2.282***	−1.132***	0.115***
	(0.026)	(0.033)	(0.026)
Age: Under 40 (average)	0.258***	0.416***	0.316***
	(0.014)	(0.017)	(0.014)
Age: 40 to 54 (average)	0.043***	0.066***	0.034***
	(0.012)	(0.016)	(0.013)
Household size (average)	−0.084***	−0.047***	0.006
	(0.005)	(0.006)	(0.005)
Children in household (average)	−0.193***	−0.372***	−0.244***
	(0.014)	(0.017)	(0.014)
Hispanic (average)	0.606***	0.726***	0.419***
	(0.012)	(0.015)	(0.012)
Broadband internet (average)	0.911***	1.682***	1.322***
	(0.013)	(0.017)	(0.014)
College graduate (average)	0.146***	0.326***	0.148***
	(0.022)	(0.027)	(0.022)
Observations	147,852	147,852	147,852
R^2	0.223	0.225	0.137

Omitting measures of local demographics from our regressions of online spending on offline store counts biases our estimates of offline stores' effects when local demographics shift both online spending and offline store counts. Whereas Table 7 describes the relationship between online spending and a measure of local demographics (the high-income share), Table 8 describes the relationship between local demographics and the offline retail environment. In particular, we regress the number of offline stores within 20km of a consumer on the shares of the population within 20km of the consumer that (i) have household incomes exceeding $75,000, (ii) belong to each of the white and black racial categories, (iii) are under 40 years of age, (iv) are between 40 and 54 years of age, (v) live in a household with children, (vi) are Hispanic, (vii) have access to broadband internet, (viii) and are college graduates. We additionally include the average household size among people within 20km of the consumer as a regressor, and our specification features year fixed effects. We run the regression on our 2007–2008 dataset, and we use the counts of offline Costco, Target, and Walmart locations as dependent variables. These demographic measures have substantial power in predicting the presence of offline stores, with the R^2 of the regressions for Costco and Target exceeding 0.20. Additionally, each retailer has more offline stores in places with populations that are higher income, are younger, do not have children, and are college educated. This table, along with Table 7, suggests that controlling for local demographics removes omitted variables bias from our estimates of offline stores' effects on online spending.

4 Model of online shopping

As argued by Section 3, empirical relationships between measures of offline retail presence and online spending do not necessarily reflect the causal relationship between the offline retail environment and online purchasing behaviour. We now formalize this argument in the context of the model of online shopping on which we base our primary empirical analysis. This model is described by the equation

$$y_i = h(n_i)'\alpha + z_i'\beta + \omega_i. \tag{1}$$

In (1), y_i is a measure of consumer i's online spending, n_i is a vector of counts of retailers' offline stores in consumer i's region, z_i are characteristics of consumer i (e.g., income and age), and the vector-valued function h allows for the possibility of transformations of the raw store counts n_i affecting consumer i's online spending. Last, ω_i is an unobservable shifter of the spending measure y_i, whose components we spend much of this section analyzing. The principal determinants of ω_i are i's tastes for online spending that are not observed by the econometrician, but that are known to the consumer and possibly to retailers. These tastes may be correlated within a geographical region because of, e.g., correlation in unobserved taste characteristics within a region. To the extent that our measure of consumer spending y_i suffers from mismeasurement, its measurement error will enter ω_i.

Additional consideration of the determinants of ω_i reveals the reasons that the empirical relationship between y_i and n_i does not generally represent the effect of changes in n_i on shopping behaviour. The determinants on which we focus are unobservable taste characteristics and unobservable region-level taste shifters, which relate to the two endogeneity problems outlined in Section 3. First, ω_i includes any unobserved taste characteristic that determines online shopping behaviour, which captures the endogeneity problem associated with the simultaneous determination of the distribution of taste characteristics within a region and the region's offline retail environment. We model unobserved taste characteristics of this sort as components of a vector ξ_i, which we assume is finite dimensional. The vector ξ_i enters ω_i through the following component structure on ω_i:

$$\omega_i = \xi_i' \psi + \tilde{\omega}_i.$$

The $\xi_i' \psi$ term represents the contribution of i's taste characteristics to i's online shopping behaviour. The $\tilde{\omega}_i$ random variable represents all other drivers of online spending that vary across individuals. Examples of components of ξ_i include i's enjoyment of reading, i's personality traits, and i's aesthetic preferences. We take these characteristics to be fixed across time. The $\tilde{\omega}_i$ unobservable captures idiosyncratic determinants of online spending that are not reflective of the taste characteristics in ξ_i. The $\tilde{\omega}_i$ random variable captures two aspects of taste heterogeneity. First, if the finite-dimensional ξ_i excludes any of consumer i's taste characteristics, then these characteristics will be captured by $\tilde{\omega}_i$. This is relevant because we use the internet usage variables described by Sections 2 and 3 to proxy for ξ_i. If our internet usage variables fail to proxy for an unobserved taste characteristic, then $\tilde{\omega}_i$ will include this taste characteristic.

Suppose that the taste characteristics ξ_i were observable. Even in this favourable case, an endogeneity problem remains on account of the possibility that $\tilde{\omega}_i$ includes regional taste disturbances that correlate with the offline retail environment on account of, e.g., contextual network effects. There are several distinct sets of assumptions that we could place on our model to address this endogeneity problem. We consider two approaches, both of which involve decomposing the unobservable $\tilde{\omega}_i$ into a structural part $\rho_{r(i)}$ that depends on consumer i's region $r(i)$ and an idiosyncratic part v_i:

$$\tilde{\omega}_i = \rho_{r(i)} + v_i. \tag{2}$$

The $\rho_{r(i)}$ unobservable is the component of $\tilde{\omega}_i$ that is responsible for the endogeneity problem related to region-level taste shocks. The v_i unobservable, meanwhile, captures idiosyncratic disturbances in online spending that relate neither to the offline retail environment nor persistent aspects of the consumer's tastes for online spending. The v_i unobservable may reflect, for instance, the timing of the consumer's needs for retail items (e.g., the consumer's laptop computer breaks, leading the consumer to purchase a replacement) or transient shocks to the consumer's liquidity that shift the consumer's expenditures (e.g., an unexpected bill that decreases the consumer's online spending for a particular year, or a work bonus that increases this spending). Additionally, v_i will include any measurement error in y_i.

One approach to the endogeneity problem posed by region-level taste shocks is to specify region fixed effects. The assumption underlying the validity of this approach is that unobserved taste characteristics excluded from ξ_i vary across *but not within* regions whose boundaries are known to the researcher. One drawback of this approach is that it requires the researcher to take a stand on the level of geography (e.g., county, municipality, media market, etc.) across which tastes for online shopping differ. Another drawback concerns estimation: by controlling for $r(i)$, we restrict ourselves from using cross-region variation in estimating the effects of n_i on y_i. This limits the precision with which we can estimate these effects. In particular, if measurement error accounts for a larger portion of within-region variation in store counts than it accounts for their total (i.e., both within- and cross-region) variation, then restricting to within-region variation leads to larger bias in the estimates of the effects of store presence. The potentially worsened estimation bias due to an increased fraction of measurement error in the explanatory variable is the reason why researchers have argued against the fixed effects approach to production function estimation; see, e.g., Ackerberg et al. [2007]. In our context, because we only have two periods of data per panelist, within-region variation in store counts is likely much smaller and more error-driven than their total variation, and we do find that estimates from fixed effects regressions are imprecise, even though some of the qualitative results are similar to our main findings (see Appendix O.4). Last, this approach does not fully solve the endogeneity problem if there are unobserved taste characteristics that are omitted from ξ_i, that vary within regions, and that correlate with n_i.

We address the endogeneity problem caused by region-level taste unobservables by modelling the dependence of these unobservables on the local demographic profile. This approach is appropriate if contextual network effects affect online shopping behavior, or if consumers' unobserved shopping tastes correlate with factors that lead them to choose neighbourhoods based on their demographic composition. In this approach, which is compatible with the fixed-effects approach outlined above, we assume that

$$\rho_r = \hat{\rho}_r + g(w_r) \tag{3}$$

where w_r are average demographic characteristics in the region r, and $\hat{\rho}_r$ is a region fixed effect. Note that the regions r need not be the same regions across which the fixed effects vary—in practice, our $\hat{\rho}_r$ fixed effects are constant for all areas r within a census region—nor do they need to be disjoint. We specify $\hat{\rho}_r$ to vary only across coarsely defined regions (i.e., census regions) so that there remains ample variation in n_i within these regions for the estimation of α. Assumption (3) requires that systematic variation in tastes for shopping (i.e., variation that relates to local store counts) between areas within a census region depend only on the characteristics of people within these areas.

This allows local online shopping tastes to be shaped by contextual network effects, or for people with different shopping tastes sorting into neighbourhoods based on the characteristics of those neighbourhoods' residents.

The assumption made in (3) implies that

$$y_i = h(n_i)'\alpha + z_i'\beta + \xi_i'\psi + \hat{\rho}_r + g(w_{r(i)}) + v_i. \tag{4}$$

This equation shows how, under our maintained assumptions, controlling for w_r solves the endogeneity problem caused by region-level taste unobservables. Our approach differs from the use of local demographics as instruments as suggested by Berry and Haile [2016]. Although we could use the w_r variables to instrument for n_i, the exclusion restriction required of this approach is unlikely to hold in our setting. This exclusion restriction is

$$\mathbb{E}[\tilde{\omega}_i \mid w_{r(i)}, z_i, \xi_i] = 0. \tag{5}$$

This condition is unlikely to hold because consumer i's tastes for online shopping $\tilde{\omega}_i$ may depend on the characteristics of i's neighbours for reasons noted above, which is precisely the dependency that we intended to address in controlling for w_r.[14]

4.1 Proxy approach

In the discussion above, we treated the ξ_i as if they were observable characteristics despite the fact that they are unobservable. Our treatment of ξ_i as an observable control follows from our assumption that the internet usage variables discussed in Sections 2 and 3 proxy for these unobserved characteristics. Suppose that ξ_i has dimension d_ξ, and let q_i denote our proxy variables, which are of dimension d_q. We assume that our proxies are noisy measurements of the taste characteristics ξ_i:

$$q_i = \Pi\xi_i + \Lambda z_i + \eta_i,$$

[14] There are conditions under which it would be preferable to use the w_r variables as instruments, however. Suppose, first, that there are no contextual network effects and that consumer tastes do not vary across places with different demographic compositions conditional on these consumers' own observable demographic characteristics. Suppose also that there are unobserved taste characteristics that are not captured by ξ_i and are therefore included in $\tilde{\omega}_i$. In this case, condition (5) is plausible and our preferred approach may fail because unobserved taste characteristics excluded from ξ_i may be correlated with n_i. If the w_r variables are mean independent of these unobserved taste characteristics, then they will be appropriate instruments for n_i. Given the richness of our internet usage variables and the possibilities of contextual network effects and of unobserved taste characteristics that correlate with local demographic composition, however, we view the assumptions underpinning our control approach as more plausible than (5).

where η_i is independent of all other random variables. Assume that Π has full column rank d_ξ, which requires the order condition $d_q \geq d_\xi$. Under this assumption,

$$\xi_i = (\Pi'\Pi)^{-1}\Pi'(q_i - \Lambda z_i) + \tilde{\eta}_i,$$

where $\tilde{\eta}_i = -(\Pi'\Pi)^{-1}\Pi'\eta_i$. We can substitute this expression for ξ_i into the structural equation (4) to obtain

$$\begin{aligned} y_i &= h(n_i)'\alpha + z_i'\tilde{\beta} + q_i'(\Pi'\Pi)^{-1}\psi + \tilde{\eta}_i'\psi + \hat{\rho}_{r(i)} + g(w_{r(i)}) + v_i \\ &= h(n_i)'\alpha + z_i'\tilde{\beta} + q_i'\gamma + \hat{\rho}_{r(i)} + g(w_{r(i)}) + \varepsilon_i \end{aligned} \tag{6}$$

for the composite parameters $\tilde{\beta} = \beta - \psi'(\Pi'\Pi)^{-1}\Pi'\Lambda$ and $\gamma = \Pi(\Pi'\Pi)^{-1}\psi$, and the composite unobservable $\epsilon_i = \tilde{\eta}_i'\psi + v_i$. This equation shows that controlling for the observable q_i is similar to directly controlling for the unobservable ξ_i as long as the measurement error η_i is independent of all else.[15]

Our method to control for ξ_i is reminiscent of the proxy or "replacement functions" approach (Heckman and Vytlacil 2007), which has widely been used for estimating production functions (Olley and Pakes 1996, Levinsohn and Petrin 2003, Ackerberg et al. 2015, and Gandhi et al. 2020). Studies in this literature proxy unobserved productivity with an observable, such as investment, and then invert the proxy for productivity to derive an estimating equation. Note that our approach allows for multidimensional unobservables, whereas the proxy approach to production function estimation typically focuses on a one-dimensional unobservable, with a few exceptions.[16]

[15] We now consider whether including proxy variables could introduce new econometric problems. One problem is the interpretation of the coefficients γ as structural effects of the proxy variables q_i on the outcome y_i. Instead, they are reduced form parameters that capture (i) the effect of the unobserved characteristics ξ_i on y_i and (ii) the correlation structure of $(\xi_i', q_i')'$. Proxy variables are used in the literature on child development production function estimation, e.g., Cunha et al. [2010]. In that literature, it is common to express proxy variables as linear transformations of latent variables that enter production function equations either as inputs or outputs (e.g., a test score may be a proxy for cognitive skills). Proxy variables are therefore mismeasured versions of variables of interest. A large focus of the child development production function literature, therefore, is addressing the econometric problems introduced by measurement error in input and output variables. Since we are not directly interested in the the unobserved characteristics ξ_i, and measurement error does not affect the consistency of our estimator of the parameter of interest α, we do not attempt to deconvolute our proxies into unobservables and measurement errors.

[16] Cunha et al. [2010] consider multidimensional latent factors, including the cognitive and non-cognitive skills of children and parents as well as parental investments in these skills of their children. Demirer [2020] studies production functions with Hicks-neutral productivity and factor-augmenting productivity. Dhyne et al. [2020] consider production function estimation in a multi-product setting by inverting the vector of proxies for the vector of product-specific productivity unobservables.

4.2 Sources of conditional variation in store counts

We address the endogeneity of the offline retail environment (i.e., n_i) by controlling for a rich set of consumer and region characteristics. Conditional on these controls, variation in brick-and-mortar store counts reflects search-and-matching frictions in real estate markets and plausibly exogenous tastes for neighbourhoods. First, a retailer seeking to open a store may choose the store's location based on the variety of suitable properties contemporaneously listed for sale. Similarly, a consumer may decide where to live based on the selection of properties that happen to be currently listed. Additionally, the outcome of a location may depend on the ordering in the search effort at which a prospective buyer finds properties—which is influenced by orderings in real estate catalogues and realtor websites—and on both the prospective buyer's and the seller's bargaining strategies. We assume that the timing at which properties are listed, their ordering in listings publications, and search-and-matching frictions in real estate markets as independent of consumer tastes for online shopping. If this is indeed the case, a reliance on these sources of variation to estimate the effects of offline stores on online shopping does not introduce an endogeneity problem.

Plausibly exogenous tastes for neighbourhoods also induce variation in proximity to stores conditional on our controls. Consider, for illustration, two consumers with the same unobservable tastes for online shopping, but who desire to live in different neighbourhoods of their city. These consumers may prefer different neighbourhoods because of a difference in where they grew up, because they differentially value local non-retail amenities such as parks and schools, or because their friends and family members are differentially distributed across the city. These factors provide variation in tastes for neighbourhoods even conditional on the rich set of controls that we use to capture unobserved tastes for online shopping. Suppose that the first consumer's preferred neigbourhood is nearby area zoned for commercial activities with many stores. Additionally suppose that the second consumer's preferred neighbourhood is far from any brick-and-mortar stores. In this case, the difference in the consumers' tastes for neighbourhoods provides variation in their exposure to offline stores.

4.3 Scale-free measures of rival effects and cross-channel complementarities

Comparing estimated α coefficients across regressions for different time periods and retailing categories is complicated by the fact that spending levels and store counts vary markedly in scale across product categories. This section develops scale-free measures of offline stores' effects on online spending that facilitate comparisons across regressions.

Our discussion of scale-free measures requires additional notation: let $\mathcal{J}^{\text{off}}$ denote the set of offline retailers, and let $\mathcal{J}^{\text{on}}$ denote the set of online retailers.[17] Additionally, let $J = |\mathcal{J}^{\text{off}}|$ denote the number of offline retailers. We do not identify time period or category in our notation to avoid notational clutter, although we run our analysis separately for each period/category pair (e.g., books in 2007–2008 and office supplies stores in 2017–2018).

We begin by constructing a measure of the effect of a particular offline store j on the spending at a particular online store s. Note that consumer i's expected spending at online retailer s conditional on the offline retail environment n_i, consumer characteristics z_i and ξ_i, and region-level taste shocks $\rho_{r(i),s} = \hat{\rho}_{r(i),s} + g_s(w_{r(i)})$ is

$$\mathbb{E}\left[y_{is} \mid n_i, z_i, \xi_i, \rho_{r(i),s}\right] = h(n_i)'\alpha_s + z_i'\tilde{\beta}_s + q_i'\gamma_s + \rho_{r(i),s} + \tilde{\eta}_i'\psi. \tag{7}$$

This equation follows from (6), although we have added store s subscripts to signify that we estimate (6) separately for spending y_{is} at various online stores s. Recall that $\tilde{\eta}_i$ reflects mean-zero measurement error in our proxies q_i for ξ_i; this term will not effect our measures because these measures involve evaluating (7) at variables' means. In (7), $n_i = (n_{i1}, \dots, n_{iJ})$ is a vector including counts n_{ij} of offline stores of retailer j within 20km of consumer i.

We measure the effect of n_{ij} on spending by the percentage change in expected spending when the number n_{ij} of stores of offline retailer j is exogenously increased from $\bar{n}_j$ to $\bar{n}_j + 1$, holding all other explanatory variables fixed at their mean values. Letting $\bar{x}$ denote the population mean of a random variable x_i, we define the relative effect of j on s as

$$\begin{aligned} \theta_{js} &= \frac{\mathbb{E}\left[y_{is} \mid \bar{n}_j + 1, \bar{n}_{-j}, \bar{z}, \bar{\xi}, \bar{\rho}_s\right] - \mathbb{E}\left[y_{is} \mid \bar{n}_j, \bar{n}_{-j}, \bar{z}, \bar{\xi}, \bar{\rho}_s\right]}{\mathbb{E}\left[y_{is} \mid \bar{n}_j, \bar{n}_{-j}, \bar{z}, \bar{\xi}, \bar{\rho}_s\right]} \\ &= \frac{\left(h(\bar{n}_j + 1, \bar{n}_{-j}) - h(\bar{n}_j, \bar{n}_{-j})\right)'\alpha_s}{h(\bar{n})'\alpha_s + \bar{z}'\tilde{\beta}_s + \bar{q}'\gamma_s + \bar{\rho}_s}. \end{aligned} \tag{8}$$

Although θ_{js} depends on unknown parameters of the online spending model and on population means, we can consistently estimate θ_{js} by substituting consistent estimators of these parameters and sample means into (8).

Using the store-pair-specific effects defined by (8), we now define the measures of rival

[17]In our empirical analysis, $\mathcal{J}^{\text{off}}$ includes the major offline retailers listed in Table 1 as well as a composite "other" store which combines offline stores not included in the table. The set $\mathcal{J}^{\text{on}}$ includes the major online retailers which appear in Tables 2 and 3.

and own effects. First, define the *store-specific average rival effect* as

$$\theta_s^{\text{rival}} = \sum_{j \in \mathcal{J}^{\text{off}} \setminus \{s\}} w_{js}^{\text{off}} \theta_{js} \tag{9}$$

where $\{w_{js}^{\text{off}}\}_j$ are weights for offline retailers j satisfying $\sum_{j \in \mathcal{J}^{\text{off}}} w_{js}^{\text{off}} = 1$ and $w_{js}^{\text{off}} = 0$ for $j = s$. We set w_{js}^{off} proportional to retailer j's number of stores in the time period under analysis.

The rival effect defined in (9) includes all effects of rival offline stores on online spending. Therefore, a positive value for this measure would indicate that a showrooming effect outweighs the competitive effect as long as the former effect is weakly positive and the latter is weakly negative. The converse of this statement is true in the case in which the rival effect is negative.

We also interpret averages of our store-specific average rival and own effects for each retailing category in each time period. These averages are defined as

$$\bar{\theta}^{\text{rival}} = \sum_{s \in \mathcal{J}^{\text{on}} \cap \mathcal{J}^{\text{off}}} w_s^{\text{on}} \theta_s^{\text{rival}} \tag{10}$$

$$\bar{\theta}^{\text{own}} = \sum_{s \in \mathcal{J}^{\text{on}} \cap \mathcal{J}^{\text{off}}} w_s^{\text{on}} \theta_{ss} \tag{11}$$

for weights $\{w_s^{\text{on}}\}_s$ on stores s that sum to one across s. In practice, we set w_s^{on} proportional to the mean of y_{is}. Note that each of the average measures defined above is taken over multichannel retailers $s \in \mathcal{J}^{\text{on}} \cap \mathcal{J}^{\text{off}}$. We also compute $\bar{\theta}^{\text{rival}}$ as an average over all online retailers, including Amazon:

$$\bar{\theta}_{\text{incl}}^{\text{rival}} = \sum_{s \in \mathcal{J}^{\text{on}}} w_s^{\text{on}} \theta_s^{\text{rival}}.$$

5 Estimation details

We now turn to several specific details of our estimation procedure. Our primary estimation equation is a linear specification of (6):

$$y_i = \alpha' h(n_i) + z_i' \beta + q_i' \gamma + \hat{\rho}_{R(i)} + w_{r(i)}' \phi + \varepsilon_i, \tag{12}$$

where $\hat{\rho}_{R(i)}$ is a fixed effect for i's census region $R(i)$ and $h(n_i)$ is a vector of counts of retailers' locations within 20km of consumer i transformed by $x \mapsto \log(x+1)$. The spending outcomes y_i that we consider are the panelist's annual online spending at various retailers and in various product categories. The z_i variables are the characteristics of panelist i, which include: an indicator for household income exceeding \$75,000; indicators for white and black racial backgrounds; indicators for the consumer's head-of-household being less than 40 years of age and falling between 40 and 54 years of age; the consumer's household size; an indicator for the presence of children in the consumer's household; an indicator for the consumer being Hispanic; an indicator for the consumer having broadband internet access; an indicator for the panelist being a college graduate; and the log of the population within 20km of the panelist's ZIP code of residence. We include this population control because tastes for online shopping may vary with population density across regions. In the regressions with spending outcomes particular to specialized retailing categories, we also include the counts of Walmart, Target, and Costco stores within 20km of the consumer transformed by the $x \mapsto \log(x+1)$ as controls. These counts partially control for the overall level of retail activity in a consumer's vicinity that is not specific to the category in question. Additionally, the q_i vector includes the internet usage variables described by Table 4, which are intended to proxy for unobserved taste characteristics. The $w_{r(i)}$ vector, meanwhile, includes averages of demographic variables in the consumer's region $r(i)$ as described by Section 2. As noted by Section 4, ϵ_i is a composite unobservable that captures idiosyncratic spending shocks that relate neither to the offline retail environment as summarized by n_i nor the consumer's persistent tastes for online shopping, mismeasurement of y_i, and noise in our measurement of unobserved taste characteristics ξ_i by the proxies q_i. Finally, we control for year fixed effects in each regression.

We estimate (12) by ordinary least squares. To reduce the dependence of our results on outlier observations, we trim our data before running regressions in which the y_i outcome is a spending level. In particular, we eliminate observations in which the outcome variable exceeds its 98th percentile conditional on the variable taking on a positive value. In our regressions for cross-channel retailers, we include store counts for Walmart, Costco, and Target as regressors. In the specialized retailing categories (i.e., books, office supplies, and electronics), we include offline store counts for the large retailers listed in Table 1 in addition to counts of other stores in the category that are within 20km of the consumer. In the books category, for example, this includes independent bookstores as well as smaller chains that are not among those on which we focus.

We transform our store count variables taking logs (after adding one to ensure positivity) based on our hypothesis that offline stores' effects on online spending are likely to be diminishing in the number of offline stores. One reason why these effects are likely to be diminishing is that offline stores may affect online spending by creating awareness of their associated retailing chain, and that consumers in a market are likely to be aware of a retailer once it has a few offline stores, leaving little scope for additional offline stores to further increase awareness.[18]

6 Results

6.1 Preliminaries

This section presents our results from the regressions described by Section 5 as well as the scale-free measures of the rival and own effects as discussed in Section 4.3. Our regression tables include only the estimates of the parameters of interest, which are the effects of offline stores counts. Throughout our discussion of the results, we emphasize estimates for the 2007–2008 time period on account of the fact that—as documented by Tables 2 and 3—we observe more spending at multichannel retailers in this time period. This allows us to estimate effects more precisely for 2007–2008 than for 2017–2018. As discussed in Section 2, we attribute the decrease in observed sales at multichannel retailers' online stores in part to the Comscore panel's decrease in coverage of online transactions between 2008 and 2017. Recall that we analyze store counts for the large retailers listed in Table 1, and online sales for these retailers' online stores and for Amazon.

6.2 Hypothesized differences between categories

We hypothesize that, on account of showrooming effects, rival effects are weakest in the books category among the specialized retailing categories that we study. This hypothesis owes to both reasoning about the products sold in these categories and the heterogeneity between online sales in these categories as documented by Hortaçsu and Syverson [2015]. Hortaçsu and Syverson [2015] note that the e-commerce share of books and magazines in 2013 was 44.2%, much higher than that of electronics and appliances (23.1%) and office equipment and supplies (17.3%). There are several reasons to expect that product categories in which e-commerce is a prominent channel are also those for which the showrooming effect is strongest. These reasons relate to (i) heterogeneity in the ease of learning about online products from offline experiences, (ii) the role of product variety, and (iii) urgency/shipping costs. First, we posit that e-commerce will have a higher market share in categories in which it is relatively easy for consumers to resolve their uncertainty about their tastes for products sold online. Offline stores that offer opportunities to meaningfully learn about products

[18] Wang and Goldfarb [2017] find, for instance, that awareness largely explains the positive effect of a retailer opening an offline store on that retailer's online sales in areas in which the retailer does not already have a strong presence.

sold online—which give rise to showrooming effects—provide one way to resolve this uncertainty.[19] Product variety also stands to generate a correlation between online retail's prominence in a category and the effect of offline stores on online sales in that category. In a product category with a great deal of variety, offline stores may help consumers discover new varieties and thus bolster consumer interest in the category. Also, consumers may especially like to buy from online retailers in retail categories with a great deal of product variety because they cannot be sure that offline stores with limited inventories will have the exact variety that they seek. Thus, we expect categories with more product variety to have higher online sales near offline stores and also to have greater shares of online sales.[20] Last, urgency also contributes to a cross-category correlation between online retail's prominence and the effect of offline stores on online sales. To illustrate, suppose that consumers urgently demand health products, but not books. Then, consumers will be less likely to buy health products online because ordering from the online channel involves waiting for a delivery. But consumers who cannot easily travel to an offline store to obtain their urgently desired good may buy online because their travel cost exceeds the cost of waiting, which induces a positive relationship between distance from an offline store and probability of buying online. We can re-interpret "urgency" as shipping costs more generally in the argument above.

The relationship between online retail's prominence in a category and the relevance of showrooming effects in the category suggests that the ordering of specialized retail categories by strength of rival effects (i.e., competitive effects offset by showrooming effects) will be the reverse of their ordering in terms of e-commerce's share of sales in the category. Thus, as noted above, we expect books to have the weakest rival effects. This hypothesis is supported by a consideration of the products sold in these categories as well. We expect the showrooming effect to be strong in the books category because browsing an offline bookstore provides a consumer with information about the same books that are sold at rival online bookstores.

[19] There are other reasons why a consumer could easily learn about products sold online; a consumer can learn about music products sold online, for example, by listening to music on the radio or the internet. The showrooming effect could also induce online entry and thus improve the online retail market from the consumer's perspective by making it more competitive.

[20] The substitutability of varieties is important here: a consumer may seek particular book titles or generic clothing items (e.g., a t-shirt of some kind, as opposed to an exact item). Given this consideration, the effective extent of variety seems lower in apparel and health and beauty than in books, music, and videos. Managing an inventory of diverse products may be less costly in a centralized warehouse than in a large network of stores; this could drive online entry in categories with many distinct products.

6.3 Estimates

Summary of results. Tables 9 and 10 summarize our results. Whereas Table 9 reports estimates from category-specific regressions of total spending on overall store counts, Table 10 provides estimates of the average effects $\bar{\theta}^{\text{rival}}$ and $\bar{\theta}^{\text{own}}$ as defined in Section 4.3. We discuss these tables' estimates throughout this section.

Table 9: Overall spending regressions

(a) 2007–2008

	Cross-category retailers (1)	Bookstores (2)	Electronics (3)	Office supplies (4)
N. Stores: Total	-11.754^{***}	0.789^{***}	2.517^{**}	0.468
	(2.476)	(0.190)	(1.199)	(0.771)
Mean dep. var.	187.35	9.14	47.37	12.91
Observations	145,345	146,506	146,404	146,765

(b) 2017–2018

	Cross-category retailers (1)	Bookstores (2)	Electronics (3)	Office supplies (4)
N. Stores: Total	-6.894^{***}	0.373^{***}	0.315	0.077
	(1.766)	(0.127)	(0.531)	(0.118)
Mean dep. var.	101.83	4.51	22.28	2.46
Observations	170,169	171,029	170,818	171,131

Note: These tables present the coefficients from the regressions of the overall expenditures on the number of offline stores in each category. Panel 9a displays the results for 2007–2008 and panel 9b displays those for 2017–2018. The "Mean dep. var" row presents the averages of the dependent variable (expenditures in dollars). Heteroskedasticity-robust standard errors in parentheses.

Cross-category retailers. We begin by studying large retailers that sell products across many categories. Column (1) of Table 9a and Table 9b provides results for our 2007–2008 and 2017–2018 samples, respectively, of the regressions of overall online spending on the overall number of stores operated by Walmart, Target, and Costco within 20km of the consumer. These results suggest that offline stores have a negative effect on online spending, which suggests substitutability between online and offline retail channels. In Online Appendix O.3, we show that the estimated effect of cross-category stores on online spending remains negative under a change in regression specification (Poisson regression instead of ordinary least squares) and a change in dependent variable (positive spending indicator instead of expenditure levels), although these alternative estimates are imprecise.

Table 10: Category-level rival and own effects on expenditures

(a) 2007–2008

	Cross-category retailers	Bookstores	Electronics	Office supplies
	(1)	(2)	(3)	(4)
Rival	-0.038	-0.031	-0.011	-0.030
	(0.011)	(0.013)	(0.017)	(0.016)
Rival	-0.026	0.021	0.001	-0.030
(incl. amazon)	(0.006)	(0.005)	(0.013)	(0.016)
Own	0.194	0.323	0.071	0.263
	(0.019)	(0.049)	(0.054)	(0.027)

(b) 2017–2018

	Cross-category retailers	Bookstores	Electronics	Office supplies
	(1)	(2)	(3)	(4)
Rival	-0.020	-0.011	0.048	-0.018
	(0.012)	(0.036)	(0.039)	(0.045)
Rival	-0.013	0.008	0.007	-0.009
(incl. amazon)	(0.004)	(0.008)	(0.011)	(0.013)
Own	0.086	0.108	0.016	0.172
	(0.021)	(0.069)	(0.059)	(0.074)

Note: Each column presents the category-level average rival effect and own effects. Panel 10a displays the results for 2007–2008 and panel 10b displays those for 2017–2018. The "Rival (incl. amazon)" row shows the average rival effects including Amazon. Standard errors are computed by the delta method.

Our overall effect conceals distinctions between offline stores' effects on spending at their associated online store versus their effects on their competitors' online sales. Tables 11a and 12a display estimates for regressions on the 2007–2008 and 2017–2018 samples, respectively, in which we analyze expenditures at each retailer's online store separately and include each retailer's offline store count as a distinct regressor. The estimated effects of multichannel retailers' own offline store counts on their own online sales are generally positive and statistically significant across stores and time periods (except Walmart in 2017–2018). Conversely, the estimated effects of rival offline stores on a retailer's own online sales are typically negative. The signs of these coefficients reflect both the competitive effect and the showrooming effect; thus, we interpret the negative rival effects as evidence that competitive effect generally outstrips the showrooming effect for large cross-category retailers.

Tables 11b and 12b present the rival and own effects for cross-category retailers. There are three notable patterns which are to be seen in these measures for the other categories and other years. First, for all retailers, the rival effect is negative and the own effect is positive, as is expected from the regression results. This is a type of heterogeneity that is not documented by previous studies that focus on a single retailer. Additionally, the rival and own effects are heterogeneous across retailers even in the same category. Amazon faces substantially weaker rival effects—a sales reduction of 1.6% in response to entry of an offline rival store—than the other multichannel retailers in the same sector; the second weakest effect is the sales reduction of 3.1% percent on Walmart's sales. This may reflect that Amazon successfully realized showrooming effects.

Books. We now consider spending at online stores that specialize in a particular product category. The purpose of this analysis is to understand how the rival and own effects differ across specific categories. For example, if we expect that consumers can learn more information on books sold online through offline interactions at physical bookstores than they can for other types of products, then we expect that the showrooming effect will largely offset the competitive effect and that the rival effect for booksellers will consequently be small. We find some support for this hypothesis below. Note that, within each category, we construct our expenditure measures using only sales of products within the category in question (e.g., we exclude sales of products other than books by bookstores). Additionally, our category-specific overall spending regressions include stores other than those operated by the large retailers listed in Table 1.

Column (2) of Tables 9a and 9b presents results from regressions of overall spending on books on the total number of bookstores within 20km of the consumer for the 2007–2008 and the 2017–2018 samples, respectively. In contrast to the cross-category regressions, the estimated overall effects are positive and statistically significant for both time periods. The positive effect of bookstores on online book sales is suggestive evidence of the strong showrooming effect that we hypothesized for the books category. We also estimate the distinct effects of own and rival offline stores on each retailer's online sales. Tables 13a and 14a display the results for the 2007–2008 and 2017–2018 samples, respectively. The results exhibit a basic pattern seen in the cross-category results: a multichannel retailer's own offline stores increase its online sales. It is less clear, though, that a retailer's online sales suffer from rival offline stores given that we estimate a positive rival effect for Amazon in addition to the aforementioned estimate of a positive effect of bookstores on overall online book spending. This is consistent with the hypothesis that offline bookstore experiences lead consumers to purchase books online due to a showrooming effect.

Electronics. We now turn to the electronics category. The overall spending results appear in the columns labelled (3) in Table 9, and the store-specific results appear in Tables 15 and 16. We see that overall spending is positively related to the total number of offline stores as in the books category, but that the store-specific relationships are mixed. Although Best Buy's and Circuit City's offline stores exhibit positive effects on their respective online sales and negative effects on the other's online sales in 2007, Radio Shack and Apple do not fit this pattern seen in results for other categories. None of the regression estimates for 2017–2018 are statistically significant. The rival and own effects for 2007–2008 are also qualitatively similar to other product categories.

Office supplies. The final category that we consider is office supplies retail. The columns labelled (4) in Table 9 present results for regressions of overall office supplies spending on the total number of office supplies stores within 20km of the consumer. The estimate of the overall effect of stores on spending is statistically insignificant for each time period. Tables 17 and 18 provide the store-specific spending results. These results for 2007–2008 suggest several important effects of particular offline stores on online sales. First, we estimate strong own-store effects; for each of Office Max, Office Depot, and Staples, we find that increasing the number of offline stores increases their own online sales. The rival effects are generally negative between multichannel retailers, with the exception that offline Office Max locations boost Office Depot's online sales. These estimates translate to relatively strong negative rival effects together with relatively strong positive own effects, as Tables 17b and 18b show.

6.4 Comparison across categories

We conclude this section by comparing the rival and own effects across categories. Table 10 shows the category-level rival and own effects defined by equations (10) and (11). A conspicuous pattern that emerges is that excluding Amazon leads to stronger average rival effects. This could be explained by showrooming effects — Amazon did not offer offline informative services, but can freeride on its multichannel competitors' informative services to achieve higher sales among consumers living nearby these competitors' offline stores. The fact that rival effects (including Amazon) are most positive in the books category, which we argue is especially prone to showrooming effects, further suggests the empirical relevance of showrooming. We also note that office supply retailers face relatively strong own effects. This may reflect that office supply stores offer valuable cross-channel services. Staples and Office Max, for example, allow consumers to submit documents online for in-store printing and to otherwise purchase items online for in-store pickup.

Explanations for cross-channel complementarities. We now consider various explanations for the robustly positive and large own effects that we estimate. Our analysis of these explanations is based on archived versions of retailers' websites but not a statistical analysis of the effects of retailers' service offerings on the extent of their cross-channel complementarities. The first explanation is that the option to return items purchased from a retailer online at one of the retailer's offline stores makes online purchasing more attractive. All of the recordings of retailers' websites from 2007 that we checked—including those for Walmart, Target, Costco, Barnes & Noble, Staples, Office Max, Best Buy, and Circuit City—indicate the acceptance of in-store returns for items purchased online. This service makes online ordering more appealing to a consumer who can easily visit an offline store to return items that do not match the consumer's expectations.

Another potential driver of cross-channel complementarities is "buy online, pickup in-store" (BOPIS) services. As noted above, office supplies stores in our sample allow consumers to submit documents for printing at brick-and-mortar stores. More generally, retailers often allow consumers to purchase items online for pick-up at a physical store; oftentimes, retailers offer to ship items that are not carried by an offline store to that store for consumers to pick up without charging consumers shipping fees. Walmart, for instance, launched its "Site to Store" program—which allows consumers to ship items listed online to offline stores for pick-up without paying shipping fees—in March 2007. BOPIS programs were common in 2007 and have remained common across the retailers we study. They allow consumers who live near a BOPIS-practicing retailer's offline stores to resolve uncertainty about an offline store's inventories of desired items and to avoid shipping costs.

Retailer loyalty programs that apply to retailers' online and offline stores could also give rise to cross-channel complementarities. Target, for example, offered a loyalty program called "Red Card" that allowed members to claim savings on products purchased both offline and online. Staples similarly advertised a "Staples Rewards" program applying to its offline and online stores on its website in 2007. When a retailer has a greater presence in a consumer's vicinity, the consumer has more opportunities to shop at the retailer. This increases the consumer's potential savings from joining the retailer's loyalty program. Joining the loyalty program in turn makes shopping at the retailer's online store more appealing.

Some retailers limit their online services to consumers living near offline stores. Staples, for example, restricted furniture deliveries placed online to consumers living within 20 miles of an offline Staples location in 2007. Additionally, Office Max restricted free deliveries for orders over $50 to consumers within 20 miles of an Office Max store in 2007. cross-channel complementarities could owe in part to a greater availability of a retailer's online services to consumers living near one of the retailer's offline stores. Retailer's linkages between online deliveries and offline store locations also suggest an important role of offline stores in retailers' logistical processes for shipping products to consumers.

Other potential sources of cross-channel complementarities include offline stores' function as advertisements for their associated retailers, and their function in providing consumers with information about the fit-and-feel characteristics of products sold exclusively by these retailers. It is worth noting that some stores in our sample, including Costco and Target, offer online-exclusive items on their websites. Costco even claimed in 2007 that "most items available on our web site are unique to costco.com" rather than available in offline Costco stores. This disjointedness of online and offline product lines would seem to weaken the connection between a retailer's online and offline stores and therefore reduce the scope for cross-channel complementarities.

7 Conclusion

This paper developed a framework for the estimation of the effects of offline stores on online shopping, and applied this framework to a novel combination of data on offline store locations and online spending. Our framework accounts for rich unobserved preference heterogeneity using a combination of consumer-level variables that proxy for tastes and region-level variables that control for systematic differences in tastes for online shopping across geography. We found that cross-channel complementarities are more empirically relevant than cannibalization: a retailer's own offline stores generally increase its online sales. The offline stores of a retailer's rival generally decrease this retailer's online sales, although these negative rival effects are much smaller than positive own effects noted above. This conclusion applies to the retail categories of bookstores, electronics stores, and office supplies stores. We also find that Amazon often experiences sales increases when its multichannel rivals open offline stores, which could be explained by Amazon successfully taking advantage of showrooming effects. Our estimates suggest that offline bookstores raise online sales of books, a category that we hypothesize is especially prone to showrooming effects. By contrast, the number of large cross-category stores reduces total online spending. These results additionally suggest a role for offline stores in supporting their associated retailers' online sales, and possibly their competitors' sales.

Although we find that brick-and-mortar stores bolster retailers' online sales—that is, multichannel retailing benefits from synergies between channels—the past decade has witnessed the increased importance of Amazon, a primarily online retailer, in e-commerce. This raises the questions about the merits and relevance of multichannel retailing strategies. We argue that Amazon has benefited from economies of scale and scope that have permitted its success despite not realizing cross-channel complementarities to the extent of its rivals; see, e.g., Houde et al. [2022] for analysis of Amazon's logistical network and associated economies of scale. In addition, Amazon's expansion into brick-and-mortar retail through its acquisition of Whole Foods and its introduction of offline stores (under the brands Amazon Go, Amazon Fresh, and Amazon Style) suggests that it understands the gains from realizing cross-channel complementarities. As studied by Bell et al. [2018], the eyewear retailer Warby Parker similarly expanded from a primarily online model to a multichannel model and benefited from cross-channel complementarities as a result.

Our work suggests several fruitful directions for future research. The first is the decomposition of cross-channel complementarities. We suggest several explanations for the positive cross-channel complementarities that we estimate across categories, time periods, and model specifications; these include in-store returns; "buy online, pickup in-store" services; reward programs applying to retailers' online and offline stores; offline stores' effects on retailer awareness; and the value of offline stores in providing consumers with information about the fit-and-feel characteristics of a retailer's exclusive products. We base these explanations on retailers' observed policies, but we do not attempt to quantitatively disentangle their contributions to cross-channel complementarities.

Although we focus on the effects of the offline retail environment on online sales, we note that the effects of the online retail environment on offline sales are also important in shaping retail industries. A retailer that invests in an online store, for example, is not only affected by online sales and the costs associated with developing and maintaining its online store, but also the effect of its online store on its offline sales. This effect reflects both cannibalization and cross-channel complementarities working in the opposite direction as those analyzed in this paper. In general, the market structure of retailing industries will depend on offline-to-online effects and online-to-offline effects; we leave the study of the latter and of these effects' interactions in determining equilibrium market structures to future work.

Table 11: Store-specific cross-category spending in 2007–2008

(a) Coefficients

	Spending			
	amazon	costco.com	target.com	walmart.com
	(1)	(2)	(3)	(4)
N. Stores: Costco	0.466	2.089***	0.212*	0.266
	(0.299)	(0.262)	(0.128)	(0.185)
N. Stores: Target	−0.282	0.122	0.408***	−0.752***
	(0.348)	(0.288)	(0.146)	(0.230)
N. Stores: Walmart	−1.295***	−0.683**	−0.572***	0.640***
	(0.331)	(0.332)	(0.140)	(0.189)
Mean dep. var.	14.10	2.51	3.20	5.71
Observations	146,451	146,857	146,770	146,694
R^2	0.057	0.007	0.017	0.023

(b) Rival effects and own effects

	amazon	costco.com	target.com	walmart.com
	(1)	(2)	(3)	(4)
Rival	-0.016	-0.052	-0.038	-0.031
	(0.006)	(0.030)	(0.018)	(0.013)
Own		0.695	0.063	0.047
		(0.075)	(0.023)	(0.014)

Note: Panel 11a presents the coefficients from the regressions of the expenditures at a given online retailer on the numbers of offline stores of each retailer. The "Mean dep. var" row presents the average expenditures at each online retailer. Robust standard errors in parentheses. Panel 11b displays the scale-free measures of the rival and own effects. Standard errors are computed by the delta method.

Table 12: Store-specific cross-category spending in 2017–2018

(a) Coefficients

	Spending			
	amazon	costco.com	target.com	walmart.com
	(1)	(2)	(3)	(4)
N. Stores: Costco	-0.135	0.851^{***}	0.010	-0.027
	(0.776)	(0.210)	(0.070)	(0.186)
N. Stores: Target	0.809	-0.015	0.159^{*}	-0.111
	(0.984)	(0.281)	(0.085)	(0.242)
N. Stores: Walmart	-3.688^{***}	-0.278	-0.209^{***}	-0.007
	(0.766)	(0.231)	(0.076)	(0.174)
Mean dep. var.	41.38	1.74	0.98	4.77
Observations	170,599	171,207	171,189	171,074
R^2	0.143	0.009	0.019	0.047

(b) Rival and own effects

	amazon	costco.com	target.com	walmart.com
	(1)	(2)	(3)	(4)
Rival	-0.012	-0.034	-0.051	-0.008
	(0.004)	(0.033)	(0.027)	(0.014)
Own		0.331	0.069	-0.000
		(0.079)	(0.037)	(0.013)

Note: See the notes for Table 11.

Table 13: Store-specific books spending in 2007–2008

(a) Coefficients

	Spending		
	amazon	barnesandnoble.com	booksamillion.com
	(1)	(2)	(3)
N. Stores: Barnes	0.161	0.344***	0.015
	(0.170)	(0.052)	(0.016)
N. Stores: Books-a-Million	0.272*	−0.063	0.056**
	(0.163)	(0.051)	(0.022)
N. Stores: Borders	0.468***	−0.203***	−0.021
	(0.154)	(0.059)	(0.014)
N. Stores: Other	0.555***	−0.036	−0.008
	(0.144)	(0.047)	(0.012)
N. Stores: Waldenbooks	0.008	0.113**	−0.015
	(0.138)	(0.044)	(0.011)
Mean dep. var.	5.53	0.86	0.06
Observations	146,629	146,819	146,869
R^2	0.034	0.008	0.002

(b) Rival and own effects

	amazon	barnesandnoble.com	booksamillion.com
	(1)	(2)	(3)
Rival	0.030	-0.030	-0.052
	(0.006)	(0.013)	(0.042)
Own		0.234	1.579
		(0.035)	(0.582)

Note: See the notes for Table 11.

Table 14: Store-specific books spending in 2017–2018

(a) Coefficients

	Spending	
	amazon	barnesandnoble.com
	(1)	(2)
N. Stores: Barnes	0.123	0.025
	(0.135)	(0.016)
N. Stores: Books-a-Million	-0.300^{**}	0.001
	(0.129)	(0.019)
N. Stores: Other	0.236^{*}	-0.006
	(0.125)	(0.014)
Mean dep. var.	3.80	0.13
Observations	171,056	171,215
R^2	0.052	0.003

(b) Rival and own effects

	amazon	barnesandnoble.com
	(1)	(2)
Rival	0.009	-0.011
	(0.008)	(0.036)
Own		0.108
		(0.070)

Note: See the notes for Table 11.

Table 15: Store-specific electronics spending in 2007–2008

(a) Coefficients

	Spending				
	amazon	apple.com	bestbuy.com	circuitcity.com	radioshack.com
	(1)	(2)	(3)	(4)	(5)
N. Stores: Apple	0.277	−0.162	−0.359*	−0.252	0.047**
	(0.205)	(0.283)	(0.203)	(0.216)	(0.022)
N. Stores: Best Buy	0.031	−0.298	0.589**	−0.530*	0.018
	(0.260)	(0.406)	(0.256)	(0.316)	(0.026)
N. Stores: Circuit City	−0.008	0.530	−0.562**	0.554*	−0.003
	(0.245)	(0.352)	(0.265)	(0.298)	(0.032)
N. Stores: Radio Shack	0.322	0.482	0.196	0.158	−0.059*
	(0.288)	(0.415)	(0.274)	(0.267)	(0.035)
Mean dep. var.	3.22	2.39	2.31	2.13	0.08
Observations	146,819	146,853	146,847	146,850	146,869
R^2	0.011	0.002	0.004	0.004	0.001

(b) Rival and own effects

	amazon	apple.com	bestbuy.com	circuitcity.com	radioshack.com
	(1)	(2)	(3)	(4)	(5)
Rival	0.025	0.048	-0.046	-0.044	0.159
	(0.014)	(0.032)	(0.027)	(0.031)	(0.098)
Own		-0.070	0.146	0.160	-0.254
		(0.123)	(0.063)	(0.086)	(0.140)

Note: See the notes for Table 11.

Table 16: Store-specific electronics spending in 2017–2018

(a) Coefficients

	Spending		
	amazon	apple.com	bestbuy.com
	(1)	(2)	(3)
N. Stores: Apple	0.462	−0.048	0.141
	(0.338)	(0.193)	(0.215)
N. Stores: Best Buy	−0.379	0.315	0.181
	(0.423)	(0.209)	(0.247)
N. Stores: Radio Shack	−0.300	0.259	−0.179
	(0.276)	(0.160)	(0.191)
Mean dep. var.	11.69	1.09	2.26
Observations	170,953	171,220	171,202
R^2	0.064	0.002	0.006

(b) Rival effects and own effects

	amazon	apple.com	bestbuy.com
	(1)	(2)	(3)
Rival	-0.005	0.163	-0.007
	(0.009)	(0.082)	(0.043)
Own		-0.035	0.041
		(0.143)	(0.056)

Note: See the notes for Table 11.

Table 17: Store-specific office supplies spending in 2007–2008

(a) Coefficients

	Spending			
	amazon	officedepot.com	officemax.com	staples.com
	(1)	(2)	(3)	(4)
N. Stores: Office Depot	0.028	2.190***	−0.209**	−0.278
	(0.019)	(0.370)	(0.104)	(0.438)
N. Stores: Office Max	0.007	0.790**	0.172***	−0.903**
	(0.013)	(0.366)	(0.060)	(0.380)
N. Stores: Other	−0.019	−0.349	−0.045	−0.534
	(0.018)	(0.324)	(0.085)	(0.382)
N. Stores: Staples	0.010	0.005	−0.116	1.875***
	(0.013)	(0.348)	(0.083)	(0.280)
Mean dep. var.	0.07	3.59	0.33	4.54
Observations	146,870	146,856	146,869	146,848
R^2	0.001	0.003	0.001	0.005

(b) Rival and own effects

	amazon	officedepot.com	officemax.com	staples.com
	(1)	(2)	(3)	(4)
Rival	0.041	0.016	-0.162	-0.057
	(0.049)	(0.025)	(0.064)	(0.021)
Own		0.322	0.319	0.212
		(0.050)	(0.100)	(0.029)

Note: See the notes for Table 11.

Bibliography

Daniel Ackerberg, C. Lanier Benkard, Steven Berry, and Ariel Pakes. Chapter 63 econometric tools for analyzing market outcomes. volume 6 of *Handbook of Econometrics*, pages 4171–4276. Elsevier, 2007.

Daniel Ackerberg, Gregory S. Crawford, and Jinyong Hahn. Orthogonal instruments:

Table 18: Store-specific office supplies spending in 2017–2018

(a) Coefficients

	Spending		
	amazon	officedepot.com	staples.com
	(1)	(2)	(3)
N. Stores: Office Depot	−0.064	0.062	−0.052
	(0.078)	(0.048)	(0.044)
N. Stores: Office Max	−0.022	0.058	−0.009
	(0.074)	(0.052)	(0.044)
N. Stores: Other	0.019	−0.006	−0.025
	(0.076)	(0.043)	(0.037)
N. Stores: Staples	−0.027	−0.001	0.075**
	(0.077)	(0.057)	(0.035)
Mean dep. var.	1.69	0.21	0.22
Observations	171,162	171,222	171,221
R^2	0.029	0.002	0.002

(b) Rival and own effects

	amazon	officedepot.com	staples.com
	(1)	(2)	(3)
Rival	-0.007	0.033	-0.066
	(0.012)	(0.072)	(0.056)
Own		0.177	0.168
		(0.133)	(0.075)

Note: See the notes for Table 11.

estimating price elasticities in the presence of endogenous product characteristics. 2011.

Daniel A. Ackerberg, Kevin Caves, and Garth Frazer. Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451, 2015.

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

David R. Bell, Santiago Gallino, and Antonio Moreno. Offline showrooms in omnichannel retail: Demand and operational benefits. *Management Science*, 64(4):1629–1651, 2018.

Steven T. Berry and Philip A. Haile. Identification in differentiated product markets. *Annual Review of Economics*, 8:27–52, 2016.

Erik Brynjolfsson, Yu (Jeffrey) Hu, and Mohammad S. Rahman. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science*, 55(11):1755–1765, 2009.

Dennis W. Carlton and Judith A. Chevalier. Free riding and sales strategies for the internet. *The Journal of Industrial Economics*, 49(4):441–461, 2001.

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

Ba M. Chu, David T. Jacho-Chávez, and Oliver B. Linton. Standard errors for non-parametric regression. *Econometric Reviews*, 39(7):674–690, 2020.

Gregory S. Crawford. Endogenous product choice: a progress report. *International Journal of Industrial Organization*, 30:315–320, 2012.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Babur De Los Santos, Ali Hortaçsu, and Matthijs R. Wildenbeest. Testing models of consumer search using data on web browsing purchasing behavior. *American Economic Review*, 102(6):2955–2980, 2012.

Mert Demirer. Production function estimation with factor-augmenting technology: An application to markups. *Job Market Paper*, 2020.

Emmanuel Dhyne, Amil Petrin, Valérie Smeets, and Frederic Warzynski. Theory for extending single-product production function estimation to multi-product settings. *Work. Pap., Yale Univ., New Haven, CT Google Scholar Article Location*, 2020.

Paul Dolfen, Liran Einav, Peter J. Klenow, Benjamin Klopack, Jonathan D. Levin, Laurence Levin, and Wayne Best. Assessing the gains from e-commerce. 2019. NBER working paper 25610.

Chris Forman, Anindya Ghose, and Avi Goldfarb. Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Science*, 55(1):47–57, 2009.

Amit Gandhi, Salvador Navarro, and David A. Rivers. On the identification of gross output production functions. *Journal of Political Economy*, 128(8):2973–3016, 2020.

Georg Goetz, Daniel Herold, Phil-Adrian Klotz, and Jan Thomas Schäfer. The substitutability between brick-and-mortar stores and e-commerce-the case of books. Technical report, Joint Discussion Paper Series in Economics, 2020.

Austan Goolsbee. Competition in the computer industry: Online versus retail. *The Journal of Industrial Economics*, 49(4):487–499, 2001.

James J. Heckman and Edward J. Vytlacil. Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. volume 6 of *Handbook of Econometrics*, pages 4875–5143. Elsevier, 2007.

Ali Hortaçsu and Chad Syverson. The ongoing evolution of us retail: A format tug-of-war. *Journal of Economic Perspectives*, 29(4):89–112, November 2015.

Jean-François Houde, Peter Newberry, and Katja Seim. Nexus tax laws and economies of density in e-commerce: A study of amazon's fulfillment center network. *Econometrica*, 2022. Forthcoming.

Bing Jing. Showrooming and webrooming: Information externalities between online and offline sellers. *Marketing Science*, 37(3):469–483, 2018.

Bruno Jullien, Alessandro Pavan, and Marc Rysman. Two-sided markets, pricing, and network effects. In Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, editors, *Handbook of Industrial Organization, Volume 4*, volume 4 of *Handbook of Industrial Organization*, pages 485–592. Elsevier, 2021.

Dmitri Kuksov and Chenxi Liao. When showrooming increases retailer profit. *Journal of Marketing Research*, 55(4):459–473, 2018.

James Levinsohn and Amil Petrin. Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies*, 70(2):317–341, 04 2003.

Alexander MacKay and David Smith. The empirical effects of minimum resale price maintenance. 2014.

Amit Mehra, Subodha Kumar, and Jagmohan S. Raju. Competitive strategies for brick-and-mortar stores to counter "showrooming". *Management Science*, 64(7): 3076–3090, 2018.

G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, 1996.

Andrea Pozzi. The effect of internet distribution on brick-and-mortar sales. *The RAND Journal of Economics*, 44(3):569–583, 2013.

Jeffrey T. Prince. The beginning of online/retail competition and its origins: An application to personal computers. *International Journal of Industrial Organization*, 25(1):139–156, 2007.

Thomas W. Quan and Kevin R. Williams. Product variety, across-market demand heterogeneity, and the value of online retail. *The RAND Journal of Economics*, 49 (4):877–913, 2018.

Scott Shriver and Bryan Bollinger. Demand expansion and cannibalization effects from retail store entry: A structural analysis of multi-channel demand. *Available at SSRN 2600917*, 2020.

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

Joel Waldfogel. Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *The RAND Journal of Economics*, 34(3):557–568, 2003.

Joel Waldfogel. The median voter and the median consumer: Local private goods and population composition. *Journal of Urban Economics*, 63(2):567–582, 2008.

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

A National store counts by category

Table 19: Top book retailers by store count

	2007		2008		2017		2018	
Rank	Retailer	Count	Retailer	Count	Retailer	Count	Retailer	Count
1	Barnes and Noble	929	Barnes and Noble	929	Barnes and Noble	1012	Barnes and Noble	948
2	Borders Books and Music	566	Borders Books and Music	566	Books-a-Million	246	Books-a-Million	231
3	Waldenbooks	383	Waldenbooks	383	Follett Higher Education Group	150	Follett Higher Education Group	147
4	Family Christian Book Store	269	Family Christian Book Store	269	Half Price Books	144	Half Price Books	144
5	Books-a-Million	177	Books-a-Million	177	Scholastic Book Fairs	52	Scholastic Book Fairs	37

Table 20: Top electronics retailers by store count

	2007		2008		2017		2018	
Rank	Retailer	Count	Retailer	Count	Retailer	Count	Retailer	Count
1	Radio Shack	5699	Radio Shack	5699	Best Buy	1171	Best Buy	1092
2	Best Buy	993	Best Buy	993	Radioshack	1133	Radioshack	286
3	Circuit City	753	Circuit City	753	Apple Store	282	Apple Store	278
4	Ritz Camera Ctr	413	Ritz Camera Ctr	413	Bose Corp	115	Eye Level Learning Ctr	167
5	Compusa	235	Magnolia Home Theatre	235	Spectrum	72	Microsoft Corp	110

Table 21: Top office supplies retailers by store count

	2007		2008		2017		2018	
Rank	Retailer	Count	Retailer	Count	Retailer	Count	Retailer	Count
1	Staples	1609	Staples	1609	Staples	1380	Staples	1321
2	Office Depot	1307	Office Depot	1307	Office Depot	968	Office Depot	961
3	Office Max	1068	Office Max	1068	Office Max	622	Office Max	540
4	Cartridge World	77	Corporate Express	77	w b Mason	25	w b Mason	15
5	Corporate Express	55	Indoff Inc	55	Office Shop	14	Office Shop	13