RAND Journal of Economics
Vol. 44, No. 3, Fall 2013
pp. 569–583

The effect of Internet distribution on brick-and-mortar sales

Andrea Pozzi*

I examine the introduction of an online shopping service by a large supermarket chain also operating a network of brick-and-mortar stores. The establishment of the Internet channel led to a 13 percent increase in overall revenues, with limited cannibalization of traditional sales. I study the mechanisms underlying this result, focusing on two areas. First, I demonstrate the importance of the reduction of customers' travel costs in the attraction of new business. Second, I provide some evidence that revenues increase more in markets where the chain faces more competitors, suggesting that the online channel can help divert business from rival supermarkets.

1. Introduction

■ The early age of electronic commerce is associated to the names of retailers that sold only or mainly online, like Amazon, Dell, and Webvan. However, after some dithering, many brick-and-mortar retail chains, especially large, big box ones, have heavily invested in the technology, creating online divisions alongside their network of physical outlets. Today, "hybrid" retailers—those selling both online and through traditional stores—hold significant shares of the online market in several sectors, from books to electronics to apparel. The strength of this trend suggests that leading chain stores see upsides from selling online, but there is so far little empirical evidence of the nature of the gains achieved, their magnitude, or how they materialize.

This study focuses on a particular motive driving a traditional retailer's choice of entering the online market: market share expansion. Big box retailers tend to locate in suburban and less densely populated areas which makes it harder for them to compete for customers with higher transportation costs like urban (Gautier and Zenou, 2010) or lower income ones (Chiou, 2009). Selling online can prove a successful strategy to overcome the location disadvantage and foster

*Einaudi Institute for Economics and Finance; andrea.pozzi@eief.it.
This article previously circulated under the title, "Who Is Hurt By E-Commerce? Crowding Out and Business Stealing in Online Grocery." I am especially grateful to Liran Einav for invaluable discussions at various stages of this project. I received useful comments and suggestions from David Autor, Effi Benmelech, Tim Bresnahan, Luigi Guiso, Jakub Kastl, Fabiano Schivardi, Alessandra Voena, the editor, and two anonymous referees as well as participants in presentations at Alicante, Cagliari, EIEF, IFN-Stockholm, Paris School of Economics, Collegio Carlo Alberto, the 2010 IO FOOD conference (Toulouse), the Workshop on the Economics of ICT 2011 (Evora), the 2011 ZEW Conference on the economics of ICT (Mannheim), EARIE 2011 (Stockholm), the 2012 Management and Economics of ICT Conference (Munich), the 2012 Telecom ParisTech conference on the economics of ICTs (Paris), and the AEA meeting 2013 (San Diego). I thank Stan Ernst and Neal H. Hooker for sharing their data on the number of US-based e-grocers. Financial support from SIEPR in the form of the B.F. Haley and E.S. Shaw dissertation fellowship is gratefully acknowledged. All errors are my own.

sales. In fact, the Internet breaks the link between shoppers' distance from a store and their convenience of buying there, thereby allowing big retailers to leverage their better prices and wider product availability on a group of customers they were previously less likely to attract.

However, although the new channel can capture extra revenues, it also exposes the firm to the risk of crowding out its own brick-and-mortar sales. For instance, Gentzkow (2007) documents that the introduction of a digital version of the *Washington Post* reduced the number of readers of the print edition. In fact, opening an Internet distribution channel involves a similar trade-off to that faced by a multiproduct firm considering whether to introduce a new product (Shaked and Sutton, 1990) or by a chain opening a new store (Holmes, 2011; Nishida, 2012).

I contribute new evidence on the effect of online distribution on a retailer's traditional sales by describing the case of a large supermarket chain (henceforth, the Retailer) which added an e-commerce service to its network of brick-and-mortar stores. Although the analysis relies on data from a single firm, there is no reason to believe that the mechanisms driving the result are idiosyncratic to this particular application. The effects I document are likely to be experienced by other companies when expanding their retail offer to the online channel.

The supermarket industry is an ideal setting for the study. First, revenue expansion is likely to be the chief reason leading a supermarket chain to sell online. The perishability of the goods and the time-sensitive nature of the delivery do not allow the chain to centralize operations over large geographical areas. This suggests that e-commerce cannot deliver huge efficiency gains on the cost side in this industry. Second, as grocery shopping is a frequent activity, transportation costs are particularly salient: most customers are unwilling to travel far to buy their groceries.[1] Therefore, selling online can significantly enhance the appeal of a grocer to households who do not live near to its stores.

I provide two complementary pieces of evidence on the effect of the implementation of online distribution on revenues. I start by examining household behavior and investigate whether the introduction of Internet shopping leads customers to spend more at the Retailer. Next, I use aggregated data on store sales to look directly at how this reflects on revenues of the chain.

The first exercise exploits scanner data on grocery purchases for a large panel of households who shopped both online and in-store at the chain. The household data are unique in that they separately report expenditure on both shopping channels for each customer. Unlike most studies comparing online and traditional shopping,[2] I have direct information on involvement in e-commerce at the individual level. Moreover, because online and in-store purchases occur at the same company, differences in behavior across channel cannot be due to heterogeneity in quality or reputation between online and traditional retailers. Even prices do not represent a confounding factor in this setting as the Retailer commits to offering the same prices and promotions online and in traditional stores.

I use these data to quantify the fraction of a household's online shopping that represents additional revenue for the chain, as opposed to simple substitution for purchases in brick-and-mortar stores. The detailed information available allows for a simple approach: I regress a household total (online and in-store) monthly expenditure in grocery at the Retailer on its monthly expenditure in online grocery at the same chain. If the two are uncorrelated, this suggests that purchases made online are offset by transactions that are no longer taking place in stores. Conversely, if online and total expenditure in grocery co-vary perfectly, the online service is only bringing in additional sales.

I find that the chain is mostly accruing new sales from the Internet channel. For each dollar spent online, only 45 cents represent crowded out in-store expenditure. Moreover, I observe that

[1] Ellickson and Grieco (2013) find that the catchment area of a Wal-Mart supercenter for grocery goods has only a two-mile radius. Orhun (2013) shows that the density of population living more than two miles away from the location of a store has no impact on its profits.

[2] Traditional measures of engagement in online activity have been based on proxies (Brown and Goolsbee, 2002) or surveys (Goolsbee, 2000; Gentzkow, 2007) and contained no information on the amount spent online. Ellison and Ellison (2009) uses data on actual online purchases but no information on transactions at traditional outlets.

the crowding out is lower for households living further away from stores of the chain, consistent with the idea that the online service is enhancing the appeal of the Retailer to customers who would otherwise be unlikely to shop there because of the high travel costs. The share of new business is also higher for customers located closer to competing supermarket stores, suggesting that the online service allows the Retailer to break into markets where customers were before captive to rival chains.

These findings can be questioned as the choice of the shopping channel may not be exogenous to food consumption. For example, because all online orders are home delivered, Internet shopping is particularly attractive when customers need to make large grocery purchases, which they would find inconvenient to carry around themselves. However, the ordinary least squares (OLS) estimates are confirmed even after I address the endogeneity of online expenditure with an instrumental variables approach. I employ two instruments: first, I exploit the fact that the Retailer introduced the online service at different times in different markets, therefore generating variation in the availability of online shopping. In addition, I take advantage of variation in the fee charged for accessing the e-commerce service generated by the distribution of discount coupons.

The evidence emerging from the analysis of customer behavior implies that the online channel delivers monetary gains for the chain. To confirm this conclusion, I use sales data aggregated at the store level and look directly at the effect of introducing the online service on the revenues of the grocer. Online orders are fulfilled using inventories from local stores; therefore, Internet sales appear as revenues of the store that provided the merchandise. I compare sales of a store before and after online grocery was introduced in the zip code where it is located. Consistently with what emerged from the household-level data, I observe that the revenues of the average store experience a 13 percent increase after introduction of the Internet service.

Finally, I explore how this result changes with market structure by interacting the indicator for e-commerce availability with a set of dummies for the number of competitors in the store's market. The results of this exercise are not conclusive but tend to suggest that the increase in revenues is larger in markets where the Retailer faces more competitors. This is what we would expect if Internet sales came, at least in part, from poaching customers from other supermarkets. I find evidence that the benefit from offering e-commerce also varies depending on whether other competing grocers offer an Internet service in the same area. In markets where rival firms are also operating some form of online distribution, the jump in sales from the introduction of the service is half that experienced in areas where the Retailer has the monopoly in the Internet segment.

This article contributes to a rich literature assessing how the provision of goods and services by traditional firms is affected by the development of Internet-based alternatives (Goolsbee, 2001; Prince, 2007; Seamans and Zhu, 2011; Lieber and Syverson, 2012; Kroft and Pope, forthcoming). However, only a limited number of contributions (Deleersnyder et al., 2002; Gentzkow, 2007) present empirical evidence on the impact of the decision to add Internet commerce to traditional distribution from a firm's perspective. In pointing to the role of online shopping in lifting the constraints of geographical location as one force behind the results, I link this study to an established literature on the impact of e-commerce on spatial differentiation (Sinai and Waldfogel, 2004; Chiou, 2009; Forman, Ghose, and Goldfarb, 2009).

The rest of the article is organized as follows. In Section 2, I provide background on the Internet grocery business and present the data. In Section 3, I use information on household purchases to estimate the amount of new business and crowding out generated by the online channel. Section 4 presents the effect of the introduction of online shopping on store revenues. Section 5 concludes.

2. Environment and data description

■ The Retailer operates over 1500 brick-and-mortar stores across the US and sells online through the company's website. The Internet service is organized according to the "in-store picking" model.[3] Therefore, variety available and other measures of quality (e.g., stockout probability) are comparable across shopping channels. Furthermore, the chain commits to offering the same prices and promotions in-store and online,[4] which ensures that differences in revenues over the two channels are not due to different pricing policies.

The online service is offered in selected zip codes and expanded gradually after starting in 2001. Since then, every month has seen the addition of at least one new zip code to the list of those reached by the service. The Retailer tends to enter the online market in several zip codes at once with large new deployments in spring (March and April) and late summer (August and September). At the end of the first quarter of 2007, online grocery shopping was available in over 1600 zip codes; in roughly 70 percent of them the Retailer is the only grocer offering Internet shopping. As the chain sells online in a subset of the markets it entered with brick-and-mortar stores, the Internet business necessarily represents a small fraction of overall revenues. However, the size of the online segment is not negligible in markets where the web service is available: 9 percent of the trips and 25 percent of the revenues in my sample are generated online.

To shop online, customers must register, providing an address, a phone number, and their loyalty card number.[5] The loyalty card number identifies the household in the data and allows for matching its online and in-store purchases. Upon registration the customer can immediately start shopping, browsing a website structured like a virtual supermarket with goods nested in links directing to different aisles (e.g., cold cereal, canned fruit, etc.). Online orders must be worth at least $50 to be processed and payment occurs at checkout by credit or debit card. Home delivery is available every day of the week and the customer can choose the delivery time. The delivery fee is set at $9.95, but the Retailer frequently issues coupons offering discounts. The fee is also waived or reduced for large orders.

The Retailer provided scanner data relative to all the shopping trips, online and in-store, made at the chain between June 2004 and June 2006 by a sample of almost 10,000 households. Households are in the sample if they shopped at least once in a supermarket store and at least once on the Internet in the period. The data report date, shopping channel, and store of choice (for brick-and-mortar trips) for each of the household's trips as well as the list of goods purchased, as defined by their Universal Product Classification Code (UPC), quantity purchased, price paid, and promotional discounts. Over the two years, I observe 1,492,166 trips, including over 100,000 online orders. The average monthly expenditure at the chain of the average household in the sample is $426.15. If we assume the yearly expenditure in grocery of an average family of four to be ten thousand dollars,[6] I can conjecture that the Retailer accounts for more than half of the grocery need of the typical household in the data, given that the average household size in my sample is 2.5.

The average household in the sample visits a brick-and-mortar store of the chain twice per week and only shops online every six weeks (Table 1).[7] However, online trips are on average much larger than in-store ones. The existence of the $50 minimum order requirement for online orders explains this difference. If I condition on large trips (e.g., worth more than $100) where such requirement is less likely to bind, the average trip online and in-store are worth roughly the same. The existence of a delivery fee also contributes to explain the large size (both

[3] In-store picking requires that online demand in a given area is fulfilled exploiting inventory of local brick-and-mortar stores, rather than stocks in dedicated warehouses.

[4] Stores are grouped into price areas by geographic proximity. Online prices match those of the store which supplies the goods to fulfill the order.

[5] Customer who do not have a loyalty card can apply for one while registering for the online service.

[6] $10,692 is the figure mentioned in "Guerrilla grocery shopping," *Consumer Reports*, January 2010. Last retrieved on January 23, 2011 at www.consumerreports.org/cro/shopping/2010/january/guerrilla-grocery-shopping/overview/guerrilla-grocery-shopping-ov.htm.

[7] Summary statistics in Table 1 understate the importance of online shopping. Although all the households in the sample eventually become e-shoppers, not all of them have adopted the technology at the very beginning of the period. The service is not even available in all the zip codes at that time.

TABLE 1 Household Shopping Behavior, By Channel of Purchase

	Mean	Standard Deviation	Percentiles 10th	25th	50th	75th	90th
Panel A: All Trips (n = 1,492,166)							
Monthly expenditure	426.15	335.38	79.33	182.99	358.75	589.72	845.24
Trips per month	7.61	6.94	2	3	6	10	15
Expenditure per trip	56.01	68.17	4.46	10.97	29.79	76.9	148.32
Basket size	19.14	24.47	1	3	9	27	53
Total trips	160.05	143.53	32	66	125	212	320
Panel B: In-Store Trips (n = 1,372,180)							
Monthly expenditure	326.73	302.98	25.52	99.95	250.48	472.78	722.69
Trips per month	7	7.02	1	2	5	9	15
Expenditure per trip	46.71	58.39	4.08	9.99	25.82	60.22	120.26
Basket size	15.52	20	1	3	7	21	43
Total trips	147.18	144.4	20	52	110	199	309
Panel C: Online Trips (n = 119,986)							
Monthly expenditure	99.42	200.7	0	0	0	143.13	337.57
Trips per month	.61	1.08	0	0	0	1	2
Expenditure per trip	162.52	80.38	80.47	108.34	149.27	194.19	257.81
Basket size	60.49	31.8	29	40	55	74	97
Total trips	12.87	17.33	1	3	7	16	32
Panel D: Distance to Closest Supermarket Stores (n = 9,323)							
Distance to Retailer's store	1.43	2.01	0.37	0.64	1.08	1.70	2.56
Distance to competitors' store	2.55	5.27	0.53	0.86	1.44	2.47	4.45

Notes: Total and per trip expenditures are expressed in 2006 dollars. Figures for expenditure per trip and basket size are averages of households averages (i.e., the average expenditure per trip of the average household). Basket size is defined as the number of items (UPCs) purchased in a shopping trip. Distance from the closest supermarket store of the Retailer and from the closest competitor is computed in miles using data provided by the Retailer (for the former) and geodesic coordinates from References USA (for the latter). The sample includes the over 9000 households who shopped *at least* once online and *at least* once in-store at the grocery chain between June 2004 and June 2006.

in expenditure and basket size) of online trips: households pay a fixed cost to receive home delivery, with no cost for adding items.

I also have information on the Retailer's revenues thanks to a weekly panel detailing sales by UPC for a sample of 118 stores between January 2004 and December 2006. The stores were drawn to ensure representativeness of the different price areas and the online service is introduced in each of these markets, though at different points in time. For each store UPC-week triplet, the data record the quantity sold and the revenue, both gross and net, of promotional discounts.

3. Household level analysis

■ In this section, I document the change in the households' expenditure pattern triggered by the introduction of e-commerce. The goal is to determine to what extent online shopping displaces brick-and-mortar purchases at the Retailer's stores and in which measure it instead captures expenditure at other retailers or consumption alternative to grocery (e.g., dining at restaurants). The former determines the fraction of a customer's online purchases that are simply crowding out in-store business; the latter singles out the share of online sales which represent new business for the chain.

I regress the total amount (online and in-store) spent on grocery at the chain by a household in a month on its online expenditure, effectively computing the correlation between total and online expenditure at the Retailer. If sales online are new business for the Retailer, months with higher Internet expenditure should be reflected into higher total expenditure at the chain. If instead the crowding out were perfect, each dollar spent online would be offset by a reduction in the

TABLE 2 The Effect of Introducing Internet Shopping on Households' Consumption at the Retailer

	(1) OLS	(2) OLS	(3) IV	(4) OLS	(5) OLS
Online expenditure	0.665***	0.750***	0.553***	0.620***	0.683***
	(0.012)	(0.024)	(0.028)	(0.010)	(0.011)
Online expenditure* distance from competitors		– 0.008*			
		(0.005)			
Online expenditure* distance from Retailer		0.042***			
		(0.011)			
Lagged total expenditure				0.159***	
				(0.009)	
Household f.e.	Yes	No	Yes	Yes	Yes
Observations	196,148	172,113	167,590	180,725	239,167
R-squared	0.29	0.26	0.22	0.25	0.31
Number of hhid	9,323	7,789	7,789	9,194	11,629

Notes: This table reports estimates of the composition of online expenditure for customers of the Retailer. The model estimated is the one in equation (1): the coefficient on *online expenditure* (β in equation (1)) represents incremental revenues and (1 − β) gives an estimate of crowding out. The unit of observation is a household-month; standard errors (in parenthesis) are clustered at the household level. In column 2, I include demographic variables from the Census 2000 matched using the block group of residence of the household. Variables included are: share of males, share of Blacks, share of Hispanics, share of people aged 25-34, 35-44, 45-54, 55-64, and over 65, share of families, share of college graduates, share of employed, median household income, and share of commuters for 60 minutes or longer. I also include the distance in miles between the household residence and the closest store of the chain and the distance in miles between the household residence and closest store of a competitor. The former is computed using data provided by the Retailer, the latter using geodesic coordinates from References USA. These coefficients are not reported for reasons of brevity, but full results are available upon request. The instruments used in column 3 are a dummy for availability of the service in the zip code of residence of the household and an indicator variable signaling the availability of a coupon for a household in a particular month. In column 5, I consider only expenditure in perishable and nonstorable items. All monetary amounts are expressed in 2006 dollars. Significance levels: *10%, **5%, ***1%.

in-store expenditure by the household and the overall amount spent would be flat across months with different intensities of online shopping.

Exploiting cross-sectional identification is undesirable in this context because correlation between online and total expenditure in grocery could be driven by unobserved heterogeneity among households. For example, wealthier households are likely to shop for higher amounts both in-store and online, causing an upward bias to the measured correlation. I therefore include household fixed effects and identify the correlation exclusively based on within-household variation. To account for seasonal patterns and aggregate shocks to demand for grocery, a full set of year-month fixed effects is also included.

I report results from the following regression

$$Total\_Expenditure_{it} = \alpha_i + \tau_t + \beta Online\_Expenditure_{it} + \varepsilon_{it}, \tag{1}$$

where α_i and τ_t are household and time fixed effects. *Total expenditure* and *Online expenditure* are expressed in 2006 dollars and computed net of promotional discounts. *Online expenditure* is also net of the fee paid for home delivery. As sales are expressed in levels, this specification delivers an easy interpretation in terms of cannibalization and incremental business rates. Out of each dollar a household spends on the online channel, β dollars are new business for the chain; whereas (1 − β) dollars represents purchases that the household would have made at the Retailer's brick-and-mortar stores and quantify crowding out.

The baseline estimates in column 1 of Table 2 indicate that crowding out is modest. For every dollar spent online, 67 cents represent fresh business for the chain and only the residual 33 cents are displaced from its brick-and-mortar sales.[8] This finding is robust both to moving

[8] Not surprisingly, this figure is larger than the self-cannibalization induced by new stores opening whose estimates range between 13 percent (Nishida, 2012) and 25 percent (Schiraldi, Seiler, and Smith, 2011).

the unit of observation from a household to all the households living in a same zip code and to looking at longer time horizons.

This result prompts two questions. The first one relates the mechanism that associates the addition of an Internet distribution channel with the gain of new business; the second one concerns the sources of the incremental sales. On the first issue, I have emphasized the role of e-commerce in reducing transportation costs for shoppers. After the introduction of the website, customers located far away from the Retailer's stores face a lower cost of shopping there and can do so more often than they would have otherwise. The household level data provide a direct way to test whether this mechanism plays a role by looking at how the share of new business captured on the Internet varies depending on the location of the shopper. In column 2, I interact online expenditure with the distance in miles between the customer and the closest store of the chain. Because the distance from the closest shop does not vary in time for a household, I can no longer include household fixed effects in the regression. Therefore, I control for cross-sectional heterogeneity, including a set of demographic characteristics matched from Census 2000. It emerges that households living further away from stores of the chain indeed generate more additional business (and less cannibalization). Moving a household one extra mile away from the closest store of the chain (a shift that, according to the information in panel D of Table 1, corresponds to the one needed to move it from the 25th to the 75th percentile in the distribution), increases its share of online expenditure representing fresh sales by 6 percent.

As for the source of the new revenues accrued by the Retailer, they could be originated by two nonmutually exclusive channels. On the one hand, the Retailer may be gaining shopping trips from customers substituting for the outside good. For example, once buying grocery is easier, people may decide to buy food and cook rather than dining out. At the same time, the website is increasing the appeal of the chain relative to that of competitors. Households living close to a competing store must have found it convenient to shop there rather than visit one of the Retailer's. The introduction of the online service makes such customers contestable as the transportation cost from shopping at the Retailer becomes negligible. I find that shoppers who live close to competitors generate a higher share of incremental purchases when shopping on the Retailer's website. This is indirect evidence that part of the extra sales generated online represent business diverted from rival grocers. However, without additional data or strong assumptions, I cannot separately identify the contribution to the result of market expansion and business stealing.

The identification approach described above may be compromised by the existence of unobserved individual shocks to demand for grocery correlated with the choice of shopping on the web. For instance, if people systematically ordered online to exploit home delivery when they happen to be in need of large amounts of grocery (e.g., when throwing a party), the estimate of β in equation (1) would be biased upward. As a consequence, I would be underestimating the displacement of brick-and-mortar sales induced by online shopping. I address this issue in column 3 of Table 2, where I present instrumental variables estimates that control for the potential endogeneity of online expenditure.

I use two distinct instruments. The first is an indicator variable denoting availability of online shopping in the zip code of residence of the household and takes advantage of the fact that the Retailer was expanding the number of zip codes where it allowed customers to order online throughout the sample period. In practice, this instrument compares average grocery expenditure at the chain for a household before and after it had the chance to purchase grocery online. One could question the validity of the instrument because the Retailer's decision to introduce online distribution in a market is obviously based on the expected demand. However, by sample construction, all the zip codes in the data are eventually reached by the online service. Hence, as long as the timing of rollout is uncorrelated with demand considerations, the instrument is valid. Anecdotal evidence emerging from conversations with managers of the chain provides support to this assumption. Ease of deployment, knowledge of the area, and logistics are mentioned as key factors in deciding which areas to reach first rather than expected demand.[9] Furthermore, there are benefits in rolling out the service in geographically closed markets similar to those identified by Holmes (2011) for Wal-Mart stores opening and by Toivanen and Waterson (2011) for McDonald's expansion.[10] This stresses the relevance of logistic considerations over demand motives in deciding when to enter a market. The Appendix provides more formal evidence that causality runs from rollout to demand, rather than the other way around.

The distribution of coupons entitling customers to a discount fee for the Internet service in a given month can also be used as an instrument. In fact, Pozzi (2012) shows that the availability of coupons for free or discounted delivery has a strong impact on the decision to shop online. The Retailer follows a "blanket" approach and mails coupons with discounts to all registered customers living in a given zip code. Therefore, coupons availability is by construction orthogonal to individual shocks to demand for grocery.[11] Even if coupon issuing is influenced by seasonality, with more coupons being mailed closer to sweeps season, this does not compromise the validity of the instrument as aggregate trends are picked up by time dummies.

The IV estimates reported use indicator variables for e-commerce availability and coupon holding as instruments.[12] The first stage (not reported) shows that both instruments are positively and significantly correlated with online expenditure. This is expected as they all increase the probability of doing any online shopping at all. Estimates of business stealing are again positive, precisely estimated and economically substantial. More importantly, though lower than the original OLS estimate of crowding out, they are quite close to it.

One lingering concern relates to the possibility that I am not capturing the intertemporal cannibalization of online shopping on brick-and-mortar sales. In fact, as online orders are delivered at home, e-commerce is well suited for large stock-up purchases which fulfill grocery demand for current and future periods. In columns 4 and 5, I check whether the positive association between online and total sales fades once I take into account the inventory motive (Hendel and Nevo, 2006). Column 4 controls for lagged expenditure in grocery which proxies for household inventory. In that specification, I assume that a household coming out of months with similar level of grocery spending holds a comparable level of inventory. Column 5 takes a different approach to shut down the effect of stockpiling. I estimate equation (1) considering only expenditure in perishable grocery products, such as eggs or milk, which cannot be stockpiled.[13] The resulting changes in the estimated share of new business gained online are small and do not alter the economic bottom line.

The share of incremental business brought in for each household by the Internet channel has obvious implications for the Retailer's revenues. The estimates just presented can be used to compute the dollar value of the online distribution channel to the grocer as follows:

$$Incremental\ sales = (Fitted\ sales|_{\beta=\hat{\beta}} - Fitted\ sales|_{\beta=0}). \tag{2}$$

The estimated value of the channel ranges between 11.5 and 14 millions of dollars over the two years. This represents a tiny fraction of the Retailer's overall yearly revenues.[14] However, the

[9] The observed rollout sequence is consistent with these statements. The online shopping option was first offered in zip codes located around the Retailer's headquarters. Later on, the chain did not jump straight to the obvious big markets: Portland and San Jose were reached before San Francisco, Los Angeles, Philadelphia, and Washington, D.C.

[10] In my application, such benefits are mainly linked to reductions in the cost of delivery. Adjacent zip codes can be served by the truck fleet of a same brick-and-mortar store. Jumping to zip codes further away would instead require the fixed cost investment of equipping another local store with its own fleet.

[11] This also allows me to recover coupon holding for households who do not redeem them. More details on the construction of this instrument are provided in the Appendix.

[12] Alternatively, I have experimented using the size of the discount on the delivery fee instead of the indicator for coupon holding, obtaining similar results.

[13] For the purpose of this exercise, products that are technically storable but with a high cost of inventory are also considered as "nonstorable." This includes ice cream and frozen dinners, which can be stockpiled only by households with large freezer units.

[14] The Retailer is selling online only in selected areas. Therefore, the bulk of revenues must necessarily come from the brick-and-mortar division.

figure is significant in two respects. First, it suggests that the extra revenues gained thanks to the online division could be big enough to cover the fixed costs of setting it up, given that variable costs can be covered by the delivery fee.[15] Moreover, the incremental sales per customer are not negligible in size. The point estimate from the preferred specification in column 3 implies that the online channel brought in additional \$1362 per customer over the two years: this represents 18 percent of the total amount spent by the median household in the sample.

4. Store-level analysis

■ I extend the analysis based on household transaction data using a distinct data set that contains weekly revenues by UPC for a sample of stores of the chain. Whereas individual data are available only for households using the loyalty card, store revenues also include transactions by customers who do not hold one. Given that online orders are fulfilled using the inventory of brick-and-mortar outlets, Internet purchases are included as revenues for the store that provided the goods. However, the data do not distinguish between brick-and-mortar and online sales.

The store-level analysis complements the results obtained with household-level data in two main ways. First, it allows for a more direct approach to quantifying the impact of e-commerce on revenues, which could only be assessed through a back-of-the-envelope calculation when using household data. Second, it allows to address one potential vulnerability of the household-level regressions. The evidence of the previous section relied, in fact, solely on households shopping at the chain before and after the online service was introduced. Regular customers who use the loyalty card only when shopping online and new customers who started shopping at the chain after the service was introduced did not contribute to identification. Failing to consider the first group could lead to overestimation of the incremental business drawn in by the Internet, whereas omission of the latter is likely to bias it downward. Store data include purchases of both these groups, allowing me to circumvent the problem.

I aggregate sales at the store-month level and use only the 118 stores located in zip codes where the service was introduced between June 2004 and June 2006. In Table 3, I report results from regressions of the following form

$$\ln(Total\_sales_{szt}) = \alpha_s + \tau_t + \beta Online\_Available_{zt} + \varepsilon_{szt}, \tag{3}$$

where s indexes a particular store set in zip code z and t indicates a month. The variable *Online available* signals that the e-commerce service was provided in the market where the store is located. As argued before, the timing of the introduction of the service in a particular market is not driven by demand considerations. Therefore, I consider this regressor as exogenous to store revenues. Store fixed effects take care of time-invariant unobserved differences across locations and time dummies account for seasonal patterns.

In column 1, I define the market of a store as the zip code where it is located and find that store revenues go up by 13 percent after online shopping becomes available in the zip code, confirming that the Internet channel does not simply displace the Retailer traditional sales but generates new business. In column 2, I broaden the definition of a store market to include all the zip codes whose centroid is closer to it than to any other outlet of the chain.[16] I then regress monthly store revenues on the share of the zip codes in its market in which e-commerce is available, weighting each zip code by its population. Increases in the penetration of the web service in the market of a store have a positive and sizeable effect on its revenues. One standard deviation increase in the penetration of e-commerce in the store's market is associated with a 3.2 percent gain in revenues. I also experimented with a specification where log revenues are regressed directly on the number

[15] The estimated value of the online channel over the two years covers about 50 percent of the alleged initial investment in the online operations as reported in a news article. The source cannot be reported as it would identify the Retailer.

[16] As it is not always the case that online demand for a zip code is served by the store closest to it, this variable will be constructed with some error. This introduces a classical measurement error bias in the coefficient.

TABLE 3 The Effect of Introducing Internet Shopping on Store Revenues

	(1)	(2)	(3)	(4)	(5)	(6)
Access in the store's zip code	0.13** (0.012)			0.28** (0.115)	0.25** (.118)	0.25** (.096)
% zip codes with access in the store's market		0.06** (0.032)				
% zip codes with coupons in the store's market			0.08* (0.046)			
Access in the store's zip code * monopoly				– 0.01 (0.125)	– 0.17 (.157)	
Access in the store's zip code * duopoly				– 0.31* (0.176)	– 0.08 (0.163)	
Access in the store's zip code * two competitors				0.01 (0.140)		
Access in the store's zip code * three competitors				– 0.49*** (0.187)		
Access in the store's zip code * online competitors						– 0.13* (0.081)
Store f.e.	Yes	Yes	Yes	No	No	No
Observations	3,041	2,926	2,926	2,963	2,963	2,807
R-squared	0.22	0.02	0.03	0.20	0.15	0.14

Notes: The dependent variable is the logarithm of total monthly store revenues. A store's market is defined as the zip code where it is located in columns 1, 4, 5, and 6; whereas it includes all the zip codes to whose centroid the store is closer than any other store of the chain in columns 2 and 3. The shares of zip codes with access to the online service and of those with coupons for delivery fee are computed using population weights based on information from Census 2000. The number of competitors in a store's zip code is computed using information on store location from Reference USA. Column 4 considers all supermarket stores competing with the Retailer 's chain, whereas column 5 only includes stores of major supermarket chains. Specifications in columns 4 and 5 include market level controls from Census 2000: share of Blacks, share of Hispanics, share of people aged 25-34, 35-44, 45-54, 55-64, and over 65, share of families, share of college graduates, median household income. The point estimates of the market structure dummies in colums 4-6 are not reported for reasons of brevity. All specifications include month-year fixed effects. Standard errors (in parenthesis) are clustered at the store level. Significance levels: *10%, **5%, ***1%.

of individuals who have access to it. The result implies that introducing Internet grocery in a zip code with a population of 100,000 units leads to half a percentage point increase in revenues.

In column 3, I exploit a different source of variation to identify the impact of online shopping on total store revenues; I look at the distribution of coupons for free or discounted delivery of online orders. The chain does not handpick customers to which the discounts are mailed, which makes coupon availability exogenous to household grocery demand. I include time dummies to control for seasonal effects that can influence both the revenue pattern and the coupon strategy. I keep the same market definition as in column 2 and regress log revenues on the fraction of zip codes in the store's market that have been targeted for coupon distribution. Once again, zip codes are weighted according to their population. Store revenues go up in months when coupons stimulate access to Internet commerce in its area. One standard deviation increase in the share of zip codes targeted for coupon distribution raises sales by 3.9 percent.

To compare quantitatively the results of the household and store level estimates, I perform the following back-of-the-envelope calculation. I compute the dollar value of the channel implied by the results in column 1 of Table 3 following a similar procedure to that described in equation (2). In order to perform the exercise, I need to separate aggregate store sales between revenues generate in-store and online; I assume that 10 percent of total store revenues were obtained by fulfilling online orders.[17] I can then back out the share of online sales representing incremental business consistent with the monetary gains implied by the store-level data. The obtained figure is

[17] Chain executives estimate that 3 percent of the overall revenues come from the online channel. This, however, includes many stores who do not offer e-grocery. The figure must also be lower than the 25 percent emerging from the household data, because store data include business from households who never order on the Internet.

0.58 and should be compared with the IV estimate of 0.553 used to compute the monetary value of the channel from the household data. Although the two figures do not coincide, they are in the same ballpark. One possible reason for the discrepancy is that the estimates based on household data do not consider business generated by new customers attracted by the introduction of the Internet shopping channel, which are instead factored in the store-level exercise. If we assume that this is the sole source of the inconsistency between the two numbers, we can decompose overall impact of online shopping on the Retailer's revenues in a part due to the increase expenditure of existed customers (95 percent of the effect) and expenditure of new customers acquired thanks to the introduction of the feature (the residual 5 percent).

Part of the new revenues is represented by business diverted from competing stores. If the number of stores operating in a market is informative about market size (Toivanen and Waterson, 2005), we would expect markets with more competitors to offer greater potential for business stealing and to lead to a stronger revenue enhancing effect. I investigate how the impact of e-commerce on store revenues varies with market structure in the last three columns of Table 3. I identify the number of rivals operating in the same zip code of a Retailer's store using data on location from Reference USA,[18] I consider all supermarket stores (NAICS code = 44511002) including small "mom-and-pop" stores but discard department and convenience stores and warehouse clubs.

I create four separate indicator variables denoting whether the Retailer is the only supermarket store in the zip code (7 percent of the cases) or whether it has one (10 percent of all cases), two (11 percent), or three competitors (8 percent), respectively. The excluded group is the set of markets where the Retailer faces four or more rivals (64 percent of the markets).[19] This approach is more flexible than including the number of competitors as a regressors, which would impose a linear effect. Because I only have a snapshot of market structure at one point in time, store fixed effects are not identified, and I replace them with zip code characteristics (wealth, age, education, etc.) obtained from Census 2000 to control for cross-sectional differences between markets. As usual, I account for time trends by including a full set of time dummies.

The dummies for market structure, not reported for brevity, are all positive: the Retailer enjoys higher revenues in markets where fewer rivals are present. The interaction dummies for the case of one and three competitors are negative (column 4). As the excluded group is "four or more competitors," this implies that the revenue surge induced by the introduction of the service is lower for markets with fewer competitors. Consistently, the effect is also smaller when the Retailer is a monopolist, but the coefficient is imprecisely measured, likely due to the rare occurrence of such cases. The benefit from rolling out online shopping is estimated to be larger when the chain faces two competitors than when there are four or more of them. This is not consistent with our prior; however, the point estimate is not statistically significantly different from zero in this case.

In column 5, I repeat the exercise considering only outlets of "big competitors," that is, multistore chains with number of employees and revenues similar to those of the Retailer. Here I only define three dummies: monopoly markets, duopoly, and markets with two or more competitors as it is rarely the case in the data that more than two or three big supermarket chains have a store in the same zip code. The interaction coefficients have the expected sign: revenues increase less in markets where there is lower potential for business stealing. However, they are not significant. This may be read as an indication that the results for the whole sample were driven by the effect on small chains and individual stores, who suffer the bulk of the business stealing.

The analysis presented so far has been a partial equilibrium one where I focused on the unilateral decision of the Retailer to enter the online market. It is natural to wonder what happens

[18] My data pull from Reference USA dates to May 2012; whereas the window spanned by the Retailer data is 2004-2006. I adopt a conservative approach and drop all stores in Reference USA who have not been in the sample for at least six years as of May 2012.

[19] Qualitative results are not sensitive to using a larger set of dummies, although some market configurations occur in too few cases to measure precisely the associated coefficient.

when rival chains respond to the Retailer's decision to introduce e-commerce by doing the same thing. Some insights can be gained by looking at the effect of competition in the supply of the service on the amount of new business gained on the Internet. To describe competition in the online grocery market I use data gathered by Berning, Ernst, and Hooker (2005) listing the set of zip codes where Internet shopping for grocery was offered as of September 2004 and reporting the identities of the firms providing the service in each of them. The Retailer faces at most one competitor[20] and is the only retailer selling online in 70 percent of the zip codes where it rolls out the service. The interaction between online availability and online competition (column 6) implies that the additional business generated online is split among the grocers providing the service. In particular, the presence of a rival e-grocer halves the revenue growth induced by the Internet channel for the Retailer.

5. Conclusions

■ I presented results on the effect of the introduction of an online shopping service for a large supermarket chain that also operates a wide network of brick-and-mortar stores. I showed that selling online allows the Retailer to considerably expand its sales with only modest self-cannibalization and document two interesting features of this result. First, I described that—as indicated by the heterogeneity of the effect for customers located at different distances from the Retailer and its competitors—the reduction in transportation cost for customers shopping online at the Retailer is one of its driving forces. Second, I relate the magnitude of the revenue enhancement to the strength of the competition faced by the chain. I only find suggestive evidence of stronger effects in areas where the chain faces more brick-and-mortar competitors, which would be consistent with part of the additional sales coming from business stealing. The results on the effect of competition from alternative online outlets are more conclusive and indicate that it reduces the amount of extra revenues generated by the introduction of the Internet channel.

Appendix

In this appendix, I provide more details on the instrumental variables strategy exploited in Section 3.

□ Instrumental variables strategy.

Date of rollout. To address concerns about the endogeneity in the selection of the shopping channel, I instrument online expenditure with availability of e-commerce in the zip code. Information on the rollout date for each of the over 1000 zip codes where the service was introduced was provided directly by the Retailer. Introduction of the service in a market represents a positive shock to demand for online grocery, which is constrained at zero before Internet shopping is made available. Moreover, because the Retailer rolls out the service simultaneously for all customers living in a zip code, availability is uncorrelated with individual shocks to overall demand for grocery.

The decision to introduce online shopping in a zip code is clearly influenced by expectations regarding demand. Most likely, the Retailer will roll out the service in zip codes where demand for online grocery is expected to be stronger. These zip codes may be the same ones where overall demand is higher. However, this argument does not compromise identification because: (i) all the zip codes included in my sample are eventually reached by the service; (ii) I include fixed effects in the specification, therefore relying on within-zip code variation.

The main threat to the validity of the instrument comes from the possible correlation between demand and the timing of rollout. Namely, the Retailer could introduce e-grocery when it expects a demand expansion in a market for reasons unobserved by the econometrician. To establish the direction of the causality between demand growth and e-commerce introduction, I use an event study approach. I focus on the zip codes where the service was introduced during the sample span and estimate the impact of current and future availability of e-commerce on demand for grocery. I aggregate grocery consumption for all the households in the sample living in the same zip code and regress this quantity on an indicator variable for availability of online shopping as well as lead indicators as far as five months before to the introduction of the service. If introduction of online grocery is decided as a response to increased demand, current expenditure for grocery in a market could be correlated with future availability of the service. Otherwise, the leads should not be significant. The results are reported in Table A1. The lead variables are generally not significant and the jump in sales is only observed when the Internet channel is actually made available.

[20] The exceptions are two zip codes where online grocery is provided by the Retailers and two other grocers

TABLE A1 Impact of Future E-Commerce Availability on Zip Code Level Sales of the Chain

	(1)	(2)	(3)
Available	262.4***	358.9***	108***
	(37.4)	(119.2)	(40.3)
Available in t + 1		82.9	−89.2
		(110.1)	(60.3)
Available in t + 2		72.4	−77.8
		(86.8)	(64.3)
Available in t + 3		104.1	−75.2
		(89.4)	(55.2)
Available in t + 4		58.2	−121.3*
		(85.2)	(73.1)
Available in t + 5		74.6	−55.2
		(70.9)	(49.5)
N	8,319	8,319	8,319
Zip code f.e.	Yes	No	Yes

Notes: This table assesses the impact of future and current availability of online grocery on the total sales of the chain to the households included in the sample, aggregated at the zip code level. *Available* is a dummy variable that takes value one in each month where the Retailer offers online grocery in the zip code. The set of indicator variables *Available in* $t+s$ denote that the Retailer will start offering online grocery in the zip code in s months. Standard errors (in parenthesis) are clustered at the zip code level. Year-month fixed effects are included in all specifications. The sample includes only the zip codes where the Retailer introduced online grocery between June 2004 and June 2006. Significance levels: *10%, **5%, ***1%.

A final concern relates to the possibility that entry into the online segment may affect the pricing policy of the Retailer. If that were the case and, for instance, the Retailer raised its prices after making e-grocery available, the raise in sales would not automatically imply any business stealing. It is worth stressing that the Retailer is committed to offer the same prices online and in-store. Therefore, a price-induced bump in expenditure would show even in months where the household does not shop online. In other words, a change in pricing policy alone should not be able to generate a positive and significant correlation between online and total grocery consumption. Furthermore, in Figure A1, I document that pricing policy does not seem to change after rollout.

The Retailer provided data on weekly prices for each UPC sold in a subset of stores representative of their pricing areas.[21] Using such data, I constructed an index for the prices posted by the chain in a particular zip code averaging the weekly prices of the 50 most sold UPCs, weighted by revenue generated. The index can be further aggregated to take into account prices in multiple store/zip codes. In Figure A1, I plot the average price index for two subsets of stores operating in zip codes that were involved with the largest rollout events in the sample in February and August 2005. In both cases, I cannot detect a structural break in the time series of the price index after the rollout, which indicates that entry in the online segment did not have impact on the pricing policy.

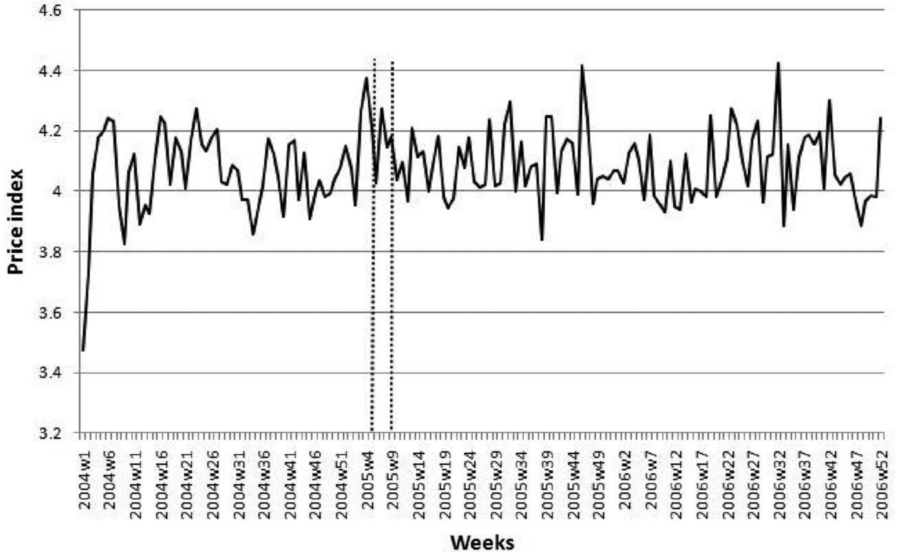
Delivery fee coupons: construction of the instrument. The Retailer data associate a set of UPCs with the fee paid for Internet delivery. So, whenever the customer is ordering online, I observe directly in the data the cost and any discount received for this service. The choice of redeeming a coupon on delivery is potentially endogenous, though. I exploit the Retailer policy in distributing delivery coupons to impute coupon holding for all households even when they decided not to redeem it.

During the sample period, coupons entitling customers to free or discounted home delivery were mailed to all registered households living in a certain area (roughly, a zip code). I proceed as follows in constructing the indicator for coupon availability. I know that all households redeeming a coupon were holding one. Therefore, I count as coupon holders all households billed a delivery fee below the regular amount unless they had shopped for more than \$150 and received a five-dollar discount or they had shopped for more than \$300 and obtained a free delivery. Crossing these thresholds, in fact, would automatically generate a fee reduction, independently of coupon ownership. Once I identify all the households redeeming a coupon in a given month, I assume that all the other ones living in the same zip code must have held one at the same time and for the same amount and I impute coupon ownership based on the zip code of residence. The size of the discount is calculated as the difference from the paid fee and the full \$9.95 one.

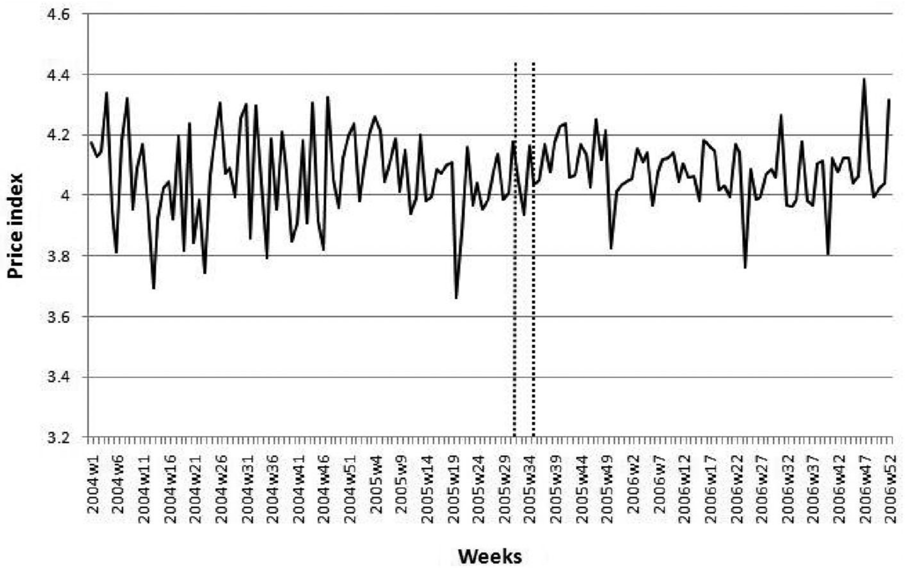
[21] The Retailer declined to disclose the exact composition of each price area.

FIGURE A1

RETAILER PRICING STRATEGY BEFORE AND AFTER INTRODUCING ONLINE GROCERY, SELECTED ZIP CODES



(a) Zip codes with rollout in February 2005



(b) Zip codes with rollout in August 2005

Notes: The figures display the pricing strategy of the Retailer before and after introduction of the Internet grocery service. The series depict movements in a price index constructed as the average of weekly prices for the 50 UPCs most sold at the Retailer chain, weighted for the revenues generated. Panel (a) refers to zip codes where the service was introduced in February 2005; panel (b) portrays information for zip codes that experienced rollout in August 2005. The dotted vertical lines indicate the month of rollout.

References

BERNING, C., ERNST, S., AND HOOKER, N.H. "Are E-Grocers Serving the Right Markets?" *Choices*, Vol. 20 (2005), pp. 77–79.

BROWN, J. AND GOOLSBEE, A. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *The Journal of Political Economy*, Vol. 110 (2002), pp. 481–507.

CHIOU, L. "Empirical Analysis of Competition between Wal-Mart and Other Retail Channels." *Journal of Economics and Management Strategy*, Vol. 18 (2009), pp. 285–322.

DELEERSNYDER, B., GEYSKENS, I., GIELENS, K., AND DEKIMPE, M.G. "How Cannibalistic Is the Internet Channel? A Study of the Newspaper Industry in the United Kingdom and the Netherlands." *International Journal of Research in Marketing*, Vol. 19 (2002), pp. 337–348.

ELLICKSON, P.B. AND GRIECO, P.L. "Wal-Mart and the Geography of Grocery Retailing." *Journal of Urban Economics*, Vol. 75 (2013), pp. 1–14.

ELLISON, G. AND ELLISON, S. "Tax Sensitivity and Home State Preference in Internet Purchasing." *American Economic Journal: Economic Policy*, Vol. 1 (2009), pp. 53–71.

FORMAN, C., GHOSE, A., AND GOLDFARB, A. "Competition between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live." *Management Science*, Vol. 55 (2009), pp. 47–57.

GAUTIER, P. AND ZENOU, Y. "Car Ownership and the Labor Market of Ethnic Minorities." *Journal of Urban Economics*, Vol. 67 (2010), pp. 392–403.

GENTZKOW, M. "Valuing New Goods in A Model with Complementarity: Online Newspapers." *The American Economic Review*, Vol. 97 (2007), pp. 713–744.

GOOLSBEE, A. "In a World Without Borders: The Impact of Taxes on Internet Commerce." *The Quarterly Journal of Economics*, Vol. 115 (2000), pp. 561–576.

———. "Competition in the Computer Industry: Online versus Retail." *The Journal of Industrial Economics*, Vol. 49 (2001), pp. 487–499.

HENDEL, I. AND NEVO, A. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, Vol. 74 (2006), pp. 1637–1673.

HOLMES, T. "The Diffusion of Wal-Mart and Economies of Density." *Econometrica*, Vol. 79 (2011), pp. 253–302.

KROFT, K. AND POPE, D.G. "Does Online Search Crowd-Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist." *Journal of Labor Economics*, (forthcoming).

LIEBER, E. AND SYVERSON, C. "Online versus Offline Competition." In M. Peitz and J. Waldfogel, eds., *Oxford Handbook of the Digital Economy*. Oxford University Press, 2012.

NISHIDA, M. "Estimating a Model of Strategic Network Choice: The Convenience-Store Industry in Okinawa." NET Institute Working Paper no. 08–27, 2012.

ORHUN, Y. "Spatial Differentiation in the Supermarket Industry: The Role of Common Information." *Quantitative Marketing and Economics*, Vol. 11 (2013), pp. 3–37.

POZZI, A. "Shopping Cost and Brand Exploration in Online Grocery." *American Economic Journal: Microeconomics*, Vol. 4 (2012), pp. 96–120.

PRINCE, J.T. "The Beginning of Online/Retail Competition and Its Origins: An Application to Personal Computers." *International Journal of Industrial Organization*, Vol. 25 (2007), pp. 139–156.

SCHIRALDI, P., SEILER, S., AND SMITH, H. "Supermarket Choice with Multi-Store Shopping: Measuring the Effect of Format *Regulation*." Mimeo, London School of Economics, 2011.

SEAMANS, R. AND ZHU, F. "Technology Shocks in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers." NET Institute Working Paper no. 10–11, 2011.

SHAKED, A. AND SUTTON, J. "Multiproduct Firms and Market Structure." *RAND Journal of Economics*, Vol. 21 (1990), pp. 45–62.

SINAI, T. AND WALDFOGEL, J. "Geography and the Internet: Is the Internet A Substitute Or A Complement for Cities?" *Journal of Urban Economics*, Vol. 56 (2004), pp. 1–24.

TOIVANEN, O. AND WATERSON, M. "Market Structure and Entry: Where's the Beef?" *RAND Journal of Economics*, Vol. 36 (2005), pp. 680–699.

———. "Retail Chain Expansion: The Early Years of McDonalds in Great Britain." CEPR Discussion Paper no. 8534, 2011.