The Effect of Internet Distribution on Brick-and-mortar Sales*

Andrea Pozzi†

October 10, 2012

Abstract

I study the introduction of an online shopping service for a large supermarket chain that also operates a wide network of brick-and-mortar stores. Entering the new market presents the usual trade-off between market share expansion and cannibalization. Adding online distribution allows the retailer to compete for customers located far away from it stores and with higher transportation costs. At the same time, the online channel can crowd out business from its own traditional stores. Using scanner data on household purchases, I show that online sales represent incremental business and crowding out is lower than 40 cent for each dollar spent on the Internet. Overall store revenues raise by 13 percent once the service is introduced in the zipcode where the store is located.

KEYWORDS: Market structure, Business stealing, E-commerce, Retail
JEL classification: D22, L21, L81

*I am especially grateful to Liran Einav for useful discussions at various stages of this project. I also thank David Autor, Effi Benmelech, Tim Bresnahan, Luigi Guiso, Jakub Kastl, Fabiano Schivardi, Alessandra Voena as well as participants in presentations at Alicante, Cagliari, EIEF, IFN-Stockholm, the 6*th* IO FOOD conference (Toulouse), the 2*nd* Workshop on the Economics of ICT (Evora), the 9*th* ZEW Conference on the economics of ICT (Mannheim), and the 38*th* EARIE Conference (Stockholm) for comments and suggestions. 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.

†Einaudi Institute for Economics and Finance. E-mail: andrea.pozzi@eief.it

1 Introduction

The introduction and diffusion of electronic commerce has transformed retail markets over the past decade. Although startup firms were the pioneers of the new channel; after some dithering many leading traditional retail chains have embraced the technology creating online divisions alongside their network of brick-and-mortar outlets. The sheer magnitude of this trend, which touches all major sectors from books to electronics to apparel, suggests that big retailers see potential benefits from this strategic choice. However, direct empirical evidence is lacking on the source of such advantages and on their size.

The online channel has some feature that can help a traditional retailer to appeal to more customers (available 24/7, no lines at the cashier, etc.). Furthermore, due to home delivery, the Internet promised to break the link between a shopper's distance from a store and its convenience in buying there. This is especially important for big box retailers that dominate thanks to better prices and wider product availability but tend to locate their outlets in suburban and less densely populated locations. This puts them at a disadvantage when competing for customers with higher transportation costs like urban (Gautier and Zenou, 2010) or lower income ones (Chiou, 2009). Selling online could represent an effective way for them to overcome such disadvantage and leverage their competitive strength on a set of customers they were previously struggling to attract. On the other hand, customers have been shown to consider online and traditional shopping as substitute (Gentzkow, 2007). Online sales could simply displace brick-and-mortar ones; leaving retailers operating both channels to shoulder the fixed cost to setup their Internet operations to accrue little or no additional business.

I study the case of a large grocery retailer that added an online shopping service to its brick-and-mortar business and document the effect of this choice on the chain's revenue. I show that the introduction of the online service has a strong positive impact on the chain 's sales. The bulk of the transactions occurring online represent new business for the chain; whereas the crowding out of the brick-and-mortar business is limited. The incremental effect is stronger for customers living further away from stores of the chain and for those located closer to competitors, suggesting that the reduction of the disadvantages of suburban location is one of the forces driving the result. Furthermore, the Retailer benefits more from the introduction of e-commerce in markets where it faces more competition, consistent with the idea that at least part of the extra revenues derive from business stolen from competitors.

The supermarket industry provides a good environment to study the impact of Internet distribution to on sales. Online grocery is a growing market. The emergence of the "in-store picking"[1] business model has improved the quality of the service and reduced the fixed cost for offering it. As a result, demand is expanding and more retailers are offering the service. The sector is currently dominated by brick-and-mortar firms who expanded into the online business after a first wave of Internet players had already entered with poor results (e.g. Webvan). The perishability of the items and time sensitivity of the delivery does not allow for centralizing operations over large geographical areas, making cost reductions and synergies harder to achieve than in other e-commerce sectors. The limited opportunities for improvement on the cost side point to market share enhancement as a major motive for starting online operations in the grocery sector. Moreover, the high frequency of food shopping makes travel costs particularly salient. Therefore the drag of less convenient location stings more in this industry than in other retail ones and the potential upside from selling online is larger.

I provide two complementary pieces of evidence on the effect of the implementation of online distribution. I start exploiting scanner data on household grocery purchases at the chain. The household data are unique in that they separately report expenditure on both shopping channels for each customer. Therefore, I observe directly engagement in online shopping and do not need to resort to proxies to measure engagement in e-commerce. The availability of expenditure by channel for each household allows for a simple approach to quantifying cannibalization and business expansion. I measure the correlation between a household total (online and in-store) monthly expenditure in grocery at the retailer and 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 not taking place in

[1] "In-store picking" requires that online demand in a given area is fulfilled exploiting inventory of local brick-and-mortar stores. It is best suited to retailers selling on both channels at the same time but online only grocers have also adopted it striking deals with traditional retail chains.

store. Therefore; the crowding out is complete. Conversely, if online and total expenditure in grocery co-vary perfectly, there is no cannibalization and the online channel is bringing in additional sales. The results imply that, when shopping on the Internet, customers are bringing new business to the chain. Out of one dollar spent online, only 40 cents represent crowded out in-store expenditure.

The straight comparison of online and offline expenditure for the same agent is made possible by convenient features of the application. The retailer has chosen to offer the same prices and promotion online and in the traditional stores; this makes revenue data more comparable across channel. Furthermore, due to in-store picking, online orders are fulfilled using inventories of brick-and-mortar stores close to the delivery address for the online customer. This implies that assortment and stockout processes are similar in the two environments. Finally, I am comparing online and traditional channel for a same retailer, meaning that brand reputation should not affect differences in sales patterns online and in-store. This empirical approach would, however, be invalid if there were unobserved factors impacting the household decision to shop on the Internet -therefore generating positive expenditure online in a given month- and the overall demand for grocery of the consumer. I address the endogeneity of the choice of the shopping channel by using two different instruments for online expenditure. First, I exploit the fact that the retailer introduced the online service at different times in different markets, therefore generating variation in availability of online shopping. As an alternative, I take advantage of variation in the fee charged for accessing the e-commerce service generated by the distribution of discount coupons. I provide evidence that neither timing of entry in the online segment nor couponing for discounted delivery fee are correlated with expected demand. The IV estimates confirm the OLS results.

I then look directly at the effect of introducing the online service on the revenues of the grocer. Due to the in-store fulfillment model, online sales figure as revenues of the store that provided the merchandize. I regress the (log of) total sales in a store in a given month on an indicator function taking value one if the retailer sells online in the market where the store is located. As argued before, the timing of the introduction of the service can be considered exogenous to sales. The coefficient is positive and large: monthly sales of the average store increase by 13 percent after online commerce is introduced in its market. I further explore how this result changes with market structure by interacting the indicator for e-commerce availability with the number of competitors in a store 's market. I find that the revenue enhancement induced by the online channel is stronger in markets with more competitors. This corroborates the hypothesis that selling online helps the chain stealing business from competitors.

This paper contributes to the literature on e-commerce and Internet markets; in particular, it relates to the debate on the crowding out induced by the Internet availability on traditional distribution (Goolsbee, 2001; Gentzkow, 2007; Kroft and Pope, 2012; Seamans and Zhu, 2011). It also sheds light on two important dimensions of competition in retail markets. First, it provides new evidence that selling online can allow large retailers to overcome location disadvantage, fitting in an established literature on the impact of e-commerce on spatial differentiation (Sinai and Waldfogel, 2004; Chiou, 2009; Forman, Ghose, and Goldfarb, 2009). Second, the addition of a new distribution channel revolves around the same trade-off (extra revenues vs. cannibalization) that characterizes the choice of opening a new store (Holmes, 2011; Nishida, 2012).

The rest of the paper as organized follows. In Section 2, I provide background on the Internet grocery business and present the data. In Section 3 I use household purchase 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 supermarket chain that provided the data (henceforth "the Retailer") operates over 1,500 brick-and-mortar stores across the US and sells online through the company's website. Prices are set weekly and the pricing strategy is "high\low": goods are sold at a relatively high price but there are frequent promotions and discounts. The Retailer's stores are grouped by geographic proximity into price areas: prices are the same for stores within the same price area but may vary across different price areas. The Retailer adopts the in-store picking model and each stores has a dedicated fleet of trucks to deliver to Internet customers. This implies that variety and other measures of quality (e.g. stockout probability) are comparable across shopping channels. The Retailer also commits to offer the same prices and promotions over the two distribution channels.

The Retailer started offering the option of shopping online in 1999 but the service was significantly reorganized in 2001. The online service was gradually expanded: at the end of the first quarter of 2007 online grocery shopping was available in over 1,600 zipcodes. The online business represents a small fraction of total revenues generated by the chain. Nevertheless, the Retailer is particularly strong in that market where it is often the dominant online retailer and frequently faces no competition. In areas where online shopping is available its size is non negligible. 9% of the trips in the sample are online orders and they account for almost 25% of the sales.

To access the online service, customers are asked to register by providing an address, a phone number, and their loyalty card number[2]. 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 seven days per week and the customer can choose her delivery time conditional on availability. There is a delivery fee set at $9.95 but the Retailer frequently issues coupons offering discounts. The fee is also waived or reduced for large orders.

The first dataset used in the analysis consists of scanner data relative to all the shopping trips, online and in-store, made at the Retailer's chain between June 2004 and June 2006 by a sample of almost 10,000 households. Data documenting individual purchases both in regular shops and on the Internet are rare. [3] Moreover, the fact that data come from the same company reduces concerns that differences in behavior across channel could be due

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

[3]Engagement in online activity has been traditionally inferred with proxies such as penetration of Internet connection (Brown and Goolsbee, 2002) or survey data (Goolsbee, 2000; Gentzkow, 2007). In all these studies participation in online shopping is measured as discrete and there is no information on its intensity (i.e. the amount spent online). Ellison and Ellison (2009) has data on actual online purchases but no information on transactions occurred at traditional outlets.

to sample selection or to differences in quality or reputation between online and traditional retailers.

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. They are identified by a household ID linked to the loyalty card (possibly multiple ones) held by members of the family. The information contained in the data includes date, shopping channel, and store of choice (for brick-and-mortar trips) for each of the household's trips. Furthermore, I have a complete description of the collection of goods purchased as defined by their Universal Product Classification Code (UPC) including quantity purchased, price paid, and promotional discounts.

Over the two years, I observe 1,492,166 trips. The great majority (1,372,180) occurred in stores but I also count over 100,000 online orders. The average monthly expenditure at the chain of the average household in the sample is $426.15. Industry sources set at $10,692 the yearly expenditure in grocery of an average family of four.[4] Since the average household size in my sample is 2.5, I can conjecture that the Retailer accounts for more than half of the grocery need of the typical household in the data.

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).[5] 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 we 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 in expenditure and basket size) of online trips: households pay a fixed cost to receive home delivery whereas there is no cost for adding items.

I also exploit a panel detailing store weekly revenues by UPC for a sample of 118 stores between January 2004 and December 2006. The stores were drawn to ensure representativeness from the different price areas set by the Retailer and the online service is introduced

[4] "Guerrilla grocery shopping", *Consumer Reports*, January 2010. Last retrieved on January 23rd, 2011.

[5] 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 zipcodes at that time. This generates by construction many months where household have no online trips and therefore, zero online expenditure.

in each of their markets. Finally, the Retailer provided a list of all the zipcodes covered by the online service with the date of first availability of the service. That enables me to trace the rollout process through time. The deployment of the online service started in late 2001 and was still ongoing at the beginning of 2007. Every month in that interval has seen the addition of at least one new zipcode to the list of those reached by the service. The Retailer tends to enter the online market in several zipcodes at once with large new deployments in Spring (March and April) and late Summer (August and September).

3 Household level analysis

I start by documenting the change in the households' expenditure pattern triggered by the introduction of e-commerce. I observe customers' grocery purchases when they occur at one of the Retailer's brick-and-mortar stores or on its website, but I do not when they buy at another grocer. I also do not observe expenditure in activities that are substitute to grocery, such as eating out or home production. I am interested in measuring the cross elasticity of purchases in the Retailer's brick-and-mortar stores and of expenditure at other retailers or in "grocery substitute" categories. 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 instead new business for the chain.

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

I report results from the following regression

$$Total\_Expenditure_{it} = \alpha + \beta Online\_Expenditure_{it} + \varepsilon_{it} \tag{1}$$

where *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.

Following the intuition given above, an estimate of β close to one would support the idea that the online business is incremental for the chain. Instead a β close to zero would suggest almost complete crowding out of online sales on store sales. Since sales are expressed in levels, the results of the regression have an easy interpretation in terms of cannibalization and incremental business. Out of each dollar a household spends on the online channel, β cents are new business for the chain; whereas $(1-\beta)$ cents represents purchases that the household would have made at the Retailer's brick-and-mortar stores and quantify crowding out.

Exploiting cross-sectional identification is undesirable in this context since association between online and total purchases could be driven by unobserved heterogeneity among households. For example, wealthier households are likely to shop for higher amounts both in-store and online, therefore generating a spurious positive correlation. For this reason, my main approach relies on the panel dimension of the data. I include household fixed effects and identify the correlation exclusively based on within-household variation. In effect, the coefficient β is identified by comparing total grocery purchases of the same household in months with different level of online grocery shopping. To account for seasonal patterns and aggregate shocks to demand for grocery, a full set of year-month fixed effects is also included.[6]

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 26 cents is being displaced from its brick-and-mortar sales.[7] In column (2) I examine whether the fact that the Internet allows the Retailer to overcome its location disadvantage contributes to this result by interacting online expenditure with the distance in miles from the

[6] The assumption that total grocery needs for a household do not vary once online shopping is introduced is key to be able to interpret β as a measure of business stealing. Note, however, that this does not imply that household demand for grocery should be fixed through time. I allow it to vary with seasonality, controlling for it with time effects. Even idiosyncratic changes in demand for grocery do not compromise identification unless they are correlated with the decision of shopping online. This concern is addressed in Section **??**.

[7] Preliminary simulations in Schiraldi, Seiler, and Smith (2011) provide a term of comparison for this figure. They find that, if Tesco were to open a new superstore in Oxford, this would induce a 25% cannibalization on the chain's other stores.

closer store of the chain and with the distance from the closest competing supermarket store. Households living further away from stores of the chain display less crowding out than those living close to it. ONE STD DEV... This finding is consistent with the Internet allowing the Retailer to reach more easily customers who were before reluctant to shop there, due to the high transportation costs they would have faced. Conversely, the closer the customer is located to competitors of the chain, the lower the self-cannibalization. Households living close to a competing store must have found convenient to shop there rather than visit one of the Retailer's. The introduction of the online service makes such customers contestable since the transportation cost from shopping at the Retailer is also negligible.

The existence of unobserved shocks to demand for grocery correlated with the choice of shopping on the web would make online expenditure endogenous and compromise the identification approach I follow. For instance, if people systematically ordered online to exploit home delivery when they happen to be in need of larger amount of grocery (e.g. when throwing a party), the fixed effect approach I used so far would underestimate the crowding out. In column (3), 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 zipcode of residence of the household. I take advantage of the fact that the Retailer was expanding the number of zipcodes where it allowed to order online throughout the sample period. The analysis relies on the subsample of 352 zipcodes where the e-commerce service was rolled out between June 2004 and June 2006 and the instrument is a dummy that takes value one once the service has been deployed in the zipcode. In practice, this instrument compares grocery expenditure at the chain for a household before and after he had the chance to purchase grocery online.

One could question orthogonality of the instrument to demand for grocery since the Retailer obviously targets markets for online entry based on their expected profitability. However, by sample construction, all the zipcodes in the data are eventually reached by the online service. Hence as long as conditional on online entry 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.[8] 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.[9] This stresses the relevance of logistic considerations over demand motives in deciding when to enter a market. Appendix B provides more formal evidence that causality runs from rollout to demand, rather than the other way around.

I also use as instrument an indicator function signalling that a household holds a coupon entitling to a discount fee for the Internet service in a given month. The Retailer follows a "blanket" approach and mails coupons with discounts to all registered customers living in a given zipcode. Because of this feature, it is enough for me to observe one household redeeming a discounted delivery fee in a given month to infer that all the households living in the same zipcode must have had one too, whether they used it or not.[10] Identification through coupon holding relies on Pozzi (2012) that shows how availability of coupons for free or discounted delivery has a strong impact on the probability of shopping online. Since coupons are mailed to all households living in a given zipcode, their availability is by construction orthogonal to individual shocks to demand for grocery, fulfilling the exclusion restriction. Even if coupon issuing is influenced by seasonality, with more coupon 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 final instrument employed is a slight variation on this last one. I exploit variation in the size of the discount which ranges between one dollar to full waiver of the $9.95 cost of internet order. The justification for the validity of this instrument is

[8]Simple observation of the sequence of rollout is consistent with these statements. The first group of zipcodes where the online shopping option was offered was clustered around the location of the Retailer's headquarter. The city counts a population of around 60,000 and is at the intersection of two major interstate roads. Even 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 DC.

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

[10]The imputation of coupon holding is obviously subject to error. For example, if no household redeems the discount, I would mistakenly infer that no coupon had been mailed. More details on the construction of this instrument are provided in the Appendix.

analogous to that provided for coupon holding.

The IV estimates reported use e-commerce availability and coupon holding as instruments.[11] 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 important, though lower than the original OLS estimates; they are quite close.

Home delivery makes online shopping well suited for large stock-up purchases, suggesting that the positive correlation so far detected could be due to intertemporal cannibalization for the Retailer rather than to contemporaneous business expansion. Columns (4) and (5) 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 potential effect of stockpiling. I estimates equation 1 considering only expenditure in perishable grocery products, such as eggs or milk, which cannot be stockpiled.[12]

Although there is some variation in the estimate of business stealing across specifications, these changes are small and do not change the economic bottom line. The magnitudes range from .61 to .72 implying that 60% to 70% of the sales the Retailer makes online are due to business stealing from competing grocers. That is a substantial figure and points to the online channel as an effective tool to hurt rivals. In order to better understand the economic relevance of the results, I use the estimate to assess implications on the grocer's revenues. The share of incremental sales derived from the online distribution channel to the grocer can be computed as follows

$$Incremental \quad sales = (Fitted \quad sales|_{\beta_1=\hat{\beta}_1} - Fitted \quad sales|_{\beta_1=0}) \tag{2}$$

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

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

The results of such calculation, as implied by the substitution effect estimated in each specification, is reported at the bottom of Table 2 expressed in millions of dollars.

The estimated value of the channel ranges between 11.5 and 14 millions over the two years. This represents a tiny fraction of the Retailer's overall yearly revenues.[13] However, the 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.[14] Moreover, the incremental sales per customer are not negligible in size. The point estimate from the preferred specification in column (2) implies that the online channel brought in additional $ 1,362 per customer over the two years: this represents 18% 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 dataset reporting weekly revenues by UPC for a sample of stores of the chain. The store level data provide complementary evidence on the effect of introducing online distribution on the Retailer 's business. First, in the case of household level regressions the bulk of identification came from households shopping at the chain before and after the online service was introduced. Customers who use the loyalty card when shopping online and those who started shopping at the chain after the service was introduced did not contribute to identification. Failing to consider the first group would lead to overestimate the incremental business drawn in by the Internet; whereas omission of the latter is likely to underestimate it. Looking at store revenues, I can take into account the effect on both of these groups. Moreover, although the household level estimates allowed for a back-of-envelope calculation of the value of the Internet channel for the Retailer, the store level analysis provides a more direct assessment of the bottom line of this effect on revenues.

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

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

I aggregate store revenues by product to obtain total weekly revenues. Since online orders are fulfilled using merchandize in brick-and-mortar outlets, Internet purchases will be included as revenues for the store that provided the goods. However, the data do not distinguish between brick-and-mortar and online sales. I base the regressions on a subset of 118 stores located in markets where the service was introduced between June 2004 and June 2006. I regress the logarithm of total monthly revenues in a store on an indicator variable for availability of the service in the zipcode where the store is located. Note that since the decision of introducing the timing of the introduction of the service in a particular market is not driven by demand considerations, the main regressor of interest can be considered exogenous to store revenues. As usual, zipcode fixed effects take care of unobservable differences across locations and time dummies account for seasonal patterns.

I find (column 1 of Table **??**) that a store revenues are 13 percent higher after online shopping becomes available in the zipcode. This is consistent with the findings of the household level analysis and confirms that the Internet channel does not simply displace the Retailer traditional sales but generates new business. In particular, the result implies that an average store of the chain would gross an extra eight millions per year after online grocery is made available in its market. In column 2, I consider as part of the potential market for a store all the zipcodes for which it is the closest store of the chain. I then regress monthly store revenues on the share of the zipcodes belonging to the store market in which e-commerce is available. Increases in the penetration of the web service in the market of a store have a positive and sizeable effect on its revenues. The coefficient is smaller than the one obtained in the main specification, likely because the relevant market for a store is constructed with some error. In fact, it is not always the case that online demand for a zipcode is served by the store closest to it.

In an alternative specification, I exploit the distribution of coupons for free or discounted delivery of online orders to understand the impact of increased online activity on store revenues. Recall that the chain mails coupons to all customers living in a same zipcode, allowing me to use household data to reconstruct which zipcodes have been targeted in a given month. I consider as a store 's market the collection of zipcodes for which it is the closest outlet of the chain and I exploit variation in the share of zipcodes in a store 's market that have been targeted for coupon distribution. It emerges that store revenues go up in months when more coupons were handed out in its market, likely boosting Internet demand. One standard deviation increase in the share of zipcodes targeted for coupon distribution raises sales by just over 3 percent.

Lastly, I analyze how the effect of the opening of the online channel on revenues varies with market structure. The increase in revenues brought in by the Internet service have two possible, non mutually exclusive sources. On the one hand, the Internet can allow for an expansion of the market leading household to buy more groceries substituting for, for instance, dining out. On the other hand, the online channel allows the Retailer to compete for new customers or for a larger share of expenditure of its existing ones; this results in business stealing from competitors. One would expect the business stealing component to depend on the number of competitors operating in a store 's market. If the store is a monopolist, there is no scope for business stealing and extra revenues can only comes through market expansion. In market where the Retailer faces many competitors, instead, there is a greater chance to poach some of their customers away.

As a proxy of competition, I use the number of stores operating in the same zipcode. Information on the location of the chain 's store comes directly from the Retailer; I obtained information on competing stores from ReferenceUSA.[15] I consider only supermarket stores (NAICS=44511002), discarding department and convenience stores and warehouse clubs. I create dummy variables for the number of competitors in a Retailer store 's zipcode and I interact them with the indicator for availability of the online service in the store 's zipcode. This approach is more flexible than including the number of competitors as a regressors, which would implicitly assume a linear relationship. In around 7 percent of the cases, a store of the chain is the only supermarket in its zipcode. The market is a duopoly, i.e. the Retailer faces one competitor, in 10 percent of the cases; in 11 percent of the markets the rival stores are two and in another 8 percent there are three competing stores. In the residual 60 percent of the cases, the Retailer has four or more competitors.[16] Because market structure

[15] My data pull from ReferenceUSA 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 ReferenceUSA who have not been in the sample for at least six years as of May 2012.

[16] Qualitative results are not sensitive to using a larger set of dummies, although some market configurations

does not vary with time in the data, I replace store fixed effects with zipcode characteristics (wealth, age, education, etc.) to control for cross-sectional differences between markets. As usual, I account for time trend by including a full set of time dummies.

Column (4) reports the results. The interaction dummies for duopoly and three competitors are negative. Since the excluded group is "four or more competitors", this means that the revenue surge induced by the introduction of the service is lower for market with fewer competitors. For market where the Retailer is monopolist, the coefficient is negative but imprecisely measured, likely due to the relative rare occurrence of such case. The fact that the interaction for the case where the chain faces two competitor is positive is not consistent with my expectations but in this case too the coefficient is not statistically different from zero. In column (5) I only consider outlets of "big competitors", that is multistore chains with number of employees and revenues comparable to that 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 that more than two or three big supermarket chains have a store in the same zipcode. The coefficients have the expected sign: revenues increase more in markets where there is more potential for business stealing. However, they are non significant. This may be read as an indication that the results on the whole sample were driven by the effect on small chains and individual stores who suffer the bulk of the business stealing.

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 documented that selling online allows the Retailer to considerably expand its sales; whereas self-cannibalization is modest. As indicated by the heterogeneity of the effect for customers located at different distances from the Retail and its competitors, the reduction in transportation cost for customers shopping online at the Retailer is one of the driving forces of the result. The fact that revenue enhancement appears to be stronger in areas where the chain faces more competitors suggests that part of the extra sales may be coming from business

occur in too few cases for me to be able to measure precisely the associated coefficient.

stealing.

References

BROWN, J. R., AND A. GOOLSBEE (2002): "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *The Journal of Political Economy*, 110(3), 481–507.

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

ELLISON, G., AND S. ELLISON (2009): "Tax Sensitivity and Home State Preference in Internet Purchasing," *American Economic Journal: Economic Policy*, 1(2), 53–71.

FORMAN, C., A. GHOSE, AND A. GOLDFARB (2009): "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 55(1), 47–57.

GAUTIER, P., AND Y. ZENOU (2010): "Car ownership and the labor market of ethnic minorities," *Journal of Urban Economics*, 67(3), 392–403.

GENTZKOW, M. (2007): "Valuing New Goods in a Model with Complementarity: Online Newspapers," *The American Economic Review*, 97(3), 713–744.

GOOLSBEE, A. (2000): "In a World without Borders: The Impact of Taxes on Internet Commerce," *The Quarterly Journal of Economics*, 115(2), 561–576.

——— (2001): "Competition in the Computer Industry: Online versus Retail," *The Journal of Industrial Economics*, 49(4), 487–499.

HENDEL, I., AND A. NEVO (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673.

HOLMES, T. J. (2011): "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 79(1), 253–302.

Kroft, K., and D. G. Pope (2012): "Does Online Search Crowd-out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist," *Journal of Labor Economics*, forthcoming.

Nishida, M. (2012): "Estimating a Model of Strategic Network Choice: The Convenience-Store Industry in Okinawa," working paper.

Pozzi, A. (2012): "Shopping Cost and Brand Exploration in Online Grocery," American Economic Journal: Microeconomics, forthcoming.

Schiraldi, P., S. Seiler, and H. Smith (2011): "Supermarket Choice with Multi-Store Shopping: Measuring the Effect of Format Regulation," working paper.

Seamans, R., and F. Zhu (2011): "Technology Shocks in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," working paper.

Sinai, T., and J. Waldfogel (2004): "Geography and the Internet: Is the Internet a substitute or a complement for cities?," *Journal of Urban Economics*, 56(1), 1–24.

Toivanen, O., and M. Waterson (2011): "Retail Chain Expansion: The Early Years of McDonalds in Great Britain," CEPR discussion paper No. 8534.

Tables and Figures

Table 1: Household shopping behavior, by channel of purchase.

	Mean	*Std. dev.*	*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

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. The sample includes the over 9,000 households who shopped *at least* once online and *at least* once in-store at the grocery chain between June 2004 and June 2006.

Table 2: The effect on 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		0.25	0.31
Number of hhid	9,323	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 business stealing and $(1-\beta)$ 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 US 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 US. These coefficients are not reported for parsimony but full results are available upon request. The instrument used in column (3) is 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 non storable items. The implied value of the online channel is expressed in millions of 2006 dollars. All figures are deflated. Significance levels: $* : 5\%$ $** : 1\%$

Table 3: The effect on introducing Internet shopping on store revenues

	(1)	(2)	(3)	(4)	(5)
Access in the store's zipcode	0.131***			0.28**	0.25**
	(0.012)			(0.115)	(.118)
% zipcodes with access in the store's market		0.064**			
		(0.032)			
% zipcodes with coupons in the store's market			0.095*		
			(0.051)		
Access in the store's zipcode * monopoly				-0.098	-0.165
				(0.125)	(.157)
Access in the store's zipcode * duopoly				-0.312*	-0.081
				(0.176)	(0.163)
Access in the store's zipcode * two competitors				0.096	
				(0.140)	
Access in the store's zipcode * three competitors				-0.488***	
				(0.187)	
Store f.e.	Yes	Yes	Yes	No	No
Observations	3,041	2,926	2,926	2,963	2,963
R-squared	0.22	0.09	0.09	0.20	0.15

Notes: The dependent variable is the logarithm of total monthly store revenues. A store's market is defined as the zipcode where it is located in columns (1), (4) and (5); whereas it includes all the zipcodes to whose centroid the store is closer than any other store of the chain in columns (2) and (3). The number of competitors in a store 's zipcode is computed using information on store location from Reference US. 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. All specifications include month-year fixed effects. Standard errors (in parenthesis) are clustered at the store level. Significance levels $:* : 5\%$ $** : 1\%$

A Instrumental variables strategy

A.1 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 zipcode. Information on the rollout date for each of the over 1,000 zipcodes 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, since the Retailer rolls out the service simultaneously for all customers living in a zipcode, availability is uncorrelated with individual shocks to overall demand for grocery.

The decision of introducing online shopping on a zipcode is clearly influenced by expectations over demand. Most likely, the Retailer will roll out the service in zipcodes where demand for online grocery is expected to be stronger. These zipcodes may be the same where overall demand is higher. However, this argument does not compromise identification because: i) all the zipcodes included in my sample are eventually reached by the service; ii) I include fixed effects in the specification, therefore relying on within zipcode 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 to 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 zipcodes 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 zipcode and regress this quantity on an indicator variable for availability of online shopping as well as leads to the introduction of the service in one up to five months. 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 B1. The lead variables are generally not significant and the jump in sales is only observed when the Internet channel is actually made available.

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 B1, I document that pricing policy does not seem to change after rollout.

The Retailer provided data on weekly prices for each UPC's sold in a subset of stores representative of their pricing areas.[17] Using such data, I constructed an index for the prices posted by the chain in a particular zipcode averaging the weekly prices of the 50 most sold UPC's, weighted by revenue generated. The index can be further aggregated to take into account prices in multiple store/zipcodes. In Figure B1 I plot the average price index for two subset of stores operating in zipcodes 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 supporting that entry in the online segment did not have impact on the pricing policy.

A.2 Delivery fee coupons: construction of the instrument

The Retailer data associate a set of UPC's to the fee paid for Internet delivery. Therefore, 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

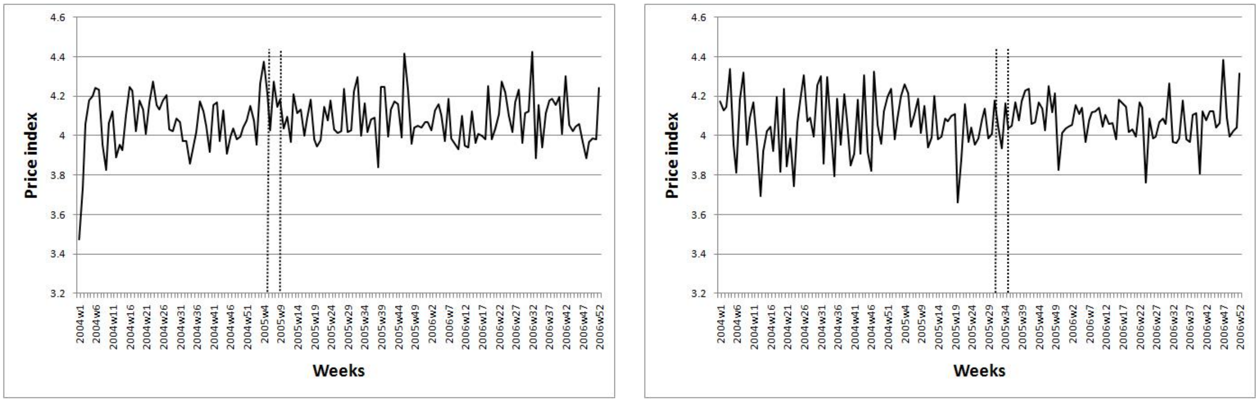
[17] The Retailer declined to disclose the exact composition of each price area.

Table B1: Impact of future e-commerce availability on zipcode 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
Zipcode 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 zipcode level. *Available* is a dummy variable that takes value one in each month where the Retailer offers online grocery in the zipcode. The set of indicator variables *Available in t+s* denote that the Retailer will start offering online grocery in the zipcode in the s months. Standard errors (in parenthesis) are clustered at the zipcode level. Year-month fixed effects are included in all specifications. The sample includes only the zipcodes where the Retailer introduced online grocery between June 2004 and June 2006. Significance levels: $*$: 10% $**$: 5% $***$: 1%

Figure B1: Retailer pricing strategy before and after introducing online grocery, selected zipcodes



(a) Zipcodes with rollout in February 2005 (b) Zipcodes 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) relates zipcodes where the service was introduced in February 2005; panel (b) portrays information for zipcodes that experienced rollout in August 2005. The dotted vertical lines indicate the month of rollout.

coupon holding for all households even when they decided not to redeem it.

During the sample period, coupons entitling to free or discounted home delivery were mailed to all registered households living in a certain area (roughly, a zipcode). I proceed as following 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 dollars discount; or they had shopped for more than \$300 and obtained a free delivery. Crossing these threshold, 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 zipcode must have held one at the same time and for the same amount and I impute coupon ownership based on the zipcode of residence. The size of the discount is calculated as the difference from the paid fee and the full \$9.95 one.