IEF

EIEF Working Paper 14/11
August 2011

Who Is Hurt by E-Commerce?
Crowding out and Business Stealing
in Online Grocery

by

Andrea Pozzi

(EIEF)

Who Is Hurt by E-commerce?
Crowding out and Business Stealing in Online Grocery*

Andrea Pozzi[†]

September 9, 2011

Abstract

I study the impact of e-commerce on competition in retail markets. Using scanner data from a large chain that markets grocery online and through traditional stores, I illustrate that selling online reduces the barrier of geographic differentiation and allows stealing business from competitors. Between 60% and 70% of the sales made online by the chain are stolen from other grocers, the rest coming from self cannibalization. I show that small stores are suffering the largest losses from this reallocation of market shares, as they were more heavily relying on geographic differentiation to survive the competitive pressure of big-box stores.

KEYWORDS: Business stealing, Crowding out, E-commerce, Retail
JEL classification: D22, L21, L81

1 Introduction

This paper studies the interplay between two phenomena that have changed the structure of the retail industry in the recent past. The first is the rise of large, multistore chains at the expenses of smaller, independent stores (Jia, 2008). The second is the diffusion of electronic commerce that has reduced shoppers' search and travel costs (Brown and Goolsbee, 2002; Keeney, 1999).

Large chains can exploit economies of scale and density to offer lower prices and wider product selection that their small size competitors struggle to match. However, big-box stores are usually located in suburban areas. This leaves smaller retailers a chance of shielding themselves from competition by locating more conveniently. Diminishing the relevance of travel cost, the diffusion of e-commerce reduces the advantage from geographic differentiation and threatens to reduce the main competitive advantage of "main street" shops. I examine to what extent large retailers are able to exploit the Internet distribution channel to increase their market share by stealing business from competitors and whether small retailers are at the losing end of this reallocation process.

This conclusion is established in steps. First, I assess whether the Internet can serve as an effective tool for a retail firm to steal business from its competitors. Next, I show that small stores are the ones suffering the most from Internet business stealing to the point that when this new type of competition enters the picture it causes a rise in the exit rate of corner food stores.

By establishing that e-commerce can be used to threaten competitors, I shed light on the determinants of a key strategic decision for retail firms, that of adding an online distribution channel to the traditional one (Lieber and Syverson, 2011). Selling online allows a retailer to compete for customers living far away who are unlikely to visit its outlets. The Internet service has also some features (active 24/7, home delivery, no lines at the cashier) that make it attractive in itself. As a result, the Internet can help a retailer attract new business. On the other hand, there is a risk that the retailer's Internet operations will simply crowd out its own traditional sales providing no additional revenue in front of the cost of operating the new channel.

My analysis focuses on the grocery industry which provides an especially relevant case study. Unlike in other retail sectors, corner food stores have been remarkably resilient to the growth of large supermarket chains (Ellickson and Grieco, 2011). The high frequency of food shopping makes travel costs more relevant and geographically limits the pull of big box stores. Therefore, a shock to the importance of location advantage is likely to have a more visible impact in this context. Moreover, after lagging behind for some time, online grocery has been steadily growing in the recent past. Today many major supermarket chains offer some form of Internet shopping and Amazon.com started stocking food items since July 2007. Mass diffusion of this shopping technology could be one of the key variables shaping the evolution of the industry in the next few years.

I exploit scanner data on grocery purchases by a large sample of households at a national supermarket chain. The retailer sells through a wide network of brick-and-mortar stores and online, via the company website. This data are unique in that they report households' expenditure on both shopping channels. Therefore, I can measure the impact of Internet shopping on traditional purchases without needing to resort to proxies for engagement in e-commerce.

To appraise the role of cannibalization and business stealing, 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 appear to be uncorrelated, this suggests that purchases made online are offset by transactions that are not taking place in 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. In this case, the retailer is likely stealing business from other grocers. Controlling for time effect and household heterogeneity, I find the correlation to be high. Results imply that out of each dollar a household spends online at the chain, sixty to seventy cents represent stolen business from competitors and only forty cents are self-cannibalization.

This empirical strategy is only valid under the assumption that there are no 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 possibility that the correlation between total and online expenditure is biased 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. This variation is induced by the distribution of discount coupons. I provide evidence that neither timing of entry in the online segment nor couponing for discounted delivery fee were correlated with expected demand. The IV estimates are close to the OLS ones, confirming that most of the Internet sales are due to business stealing, rather than self-cannibalization.

Exploring heterogeneity in these effects is informative on how the retailer leverages the online channel to steal business from competitors. I find that the selling on the web helps attracting more incremental sales from lower income families. This group should be more sensitive to the low price offered by big retailers and naturally responds more once e-commerce allows for easier access to it. Households living further away from brick-and-mortar outlets of the supermarket chain are also keener on substituting purchases at other grocers with orders on the retailer website. This result suggests that, thanks to e-commerce, the retailer is recouping potential sales lost due to location disadvantage. When looking at grocery expenditure by product category, it emerges that business stealing is larger for staple products rather than niche ones. Therefore, the threat from online competition is stronger in the categories that represent the core of traditional mom-and-pop's offer but could spare small stores focusing on more specific niches.

Finally, I look directly at the effect of the e-commerce on market structure by documenting how the number of grocers active in a zipcode changes after the supermarket chain introduces online grocery in a market. I show that this event hurts the profits of small food stores, causing exit. Instead, big grocers are unaffected by Internet competition. As a placebo test, I look at the effect on small greengrocers and butchers. Meat and produce are high-touch products and Internet shopping is a poor substitute for traditional stores in those categories. I find no evidence that these stores are affected by the rollout of the web service.

Overall e-commerce emerges as yet another investment that can help large chains squeeze

the competition from small independent stores. Adoption of optical scanners and improved logistics allowed large distributors to take full advantage of their scale and density (Holmes, 2001). Entering the online segment provides them with a way to reduce their location disadvantage and to leverage their better prices and selection on customers that they had a hard time attracting before. Since many corner grocery shops have survived exploiting customers' heterogeneity in travel costs, further diffusion of Internet purchase in the grocery sector may spell doom for mom-and-pop's and independent stores. My results imply that, if Internet shopping will become a common way of purchasing grocery, the sector is poised to experience the same increase in concentration of market share in the hands of big players already registered in other retail markets.

This paper brings together two large but so far separate streams of literature. On the one hand it ties in to contributions estimating the crowding out of Internet markets on traditional ones (Goolsbee, 2001; Gentzkow, 2007; Kroft and Pope, 2011; Seamans and Zhu, 2011). On the other one, it adds to the literature on the determinants of equilibrium market structure in retail (Smith, 2004; Ellickson, 2007; Schivardi and Viviano, 2011). In particular, it introduces a new angle to the debate over competition between large chains and independent stores (Jia, 2008; Basker and Noel, 2008; Grieco, 2010; Matsa, 2011) by stressing the role of e-commerce in reducing the scope for differentiation in location. The impact of e-commerce on spatial differentiation has been previously studied by a series of contributions on Internet and geography (Sinai and Waldfogel, 2004; Chiou, 2009; Forman, Ghose, and Goldfarb, 2009). Goldmanis, Hortacsu, Syverson, and Emre (2010) also explores the role of e-commerce in shaping market structure in retail industries. Whereas they stress the role of the reduction in search cost in fostering selection and reallocation of market share, it is the demise of geographic differentiation that plays a major part in my application.

The rest of the paper as organized follows. In Section 2, I provide background on the Internet grocery business and present the data sources used for the study. I then proceed to discuss how I identify the business stealing potential of the Internet channel and present a quantification in Section 3. Section 4 endogenizes customers' choice of the shopping channel and presents new estimates of business stealing based on instrumental variables. Section 5 provides insights on how the Internet channel can accrue new business for a retailer. The

implication for competition are discussed in Section 6. Section 7 concludes.

2   Environment and Data Description

2.1   Institutional environment

Retail grocery is a convenient market to study the business stealing effect of Internet distribution. The online segment is growing: 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, the market 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.

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 quite 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

---

[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.

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 average household visits a brick-and-mortar store twice per week and only shops online every six weeks (Table 1).[3] 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.

---

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

[3]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.

2.2 Data sources

This study exploits several sources of data, which I introduce below. More details on the data and variables construction are provided in the Appendix.

Data on online and in-store purchases. The main dataset consists of scanner data relative to all the shopping trips, online and in-store, made at the supermarket chain providing the data 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. They are identified by a household ID linked to the loyalty card (possibly multiple ones) held by members of the family.

Data documenting individual purchases both in regular shops and on the Internet are rare. [4] That they come from the same company reduces concerns that differences in behavior across channel could be due to sample selection or to differences in reputation of the online and traditional retailer seeing as they are one and the same. 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.[5] Since the average household size in my sample is 2.5, I can speculate that the Retailer accounts for more than half of the grocery need of the typical household in the data.

The Retailer also 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. The

---

[4]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.

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

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).

Nielsen Homescan data. The Homescan database collected by AC Nielsen contains scanner data from households grocery purchases but includes *all* the stores in several markets across the US. Households sign up to enter the sample and are selected to ensure demographic and geographic representativeness. They are provided with a scanning device which they use to record the products (UPCs) they purchased and are asked to manually report the retail store where they shopped.[6] The information content is comparable to that present in the Retailer's data: date of the trip, UPCs purchased, prices and quantities as well as details on the characteristics of the products. Homescan also provides detailed demographic information on the household in the samples.

The Homescan sample used in this study covers food items purchases only (non food product categories, such as liquid detergent, are excluded) for the year 2004, therefore partially overlapping with the span covered by the main data. I only consider the 12 Homescan markets where the Retailer is active for a total of 10,244 customers and 1,212,047 grocery shopping trips. In this subsample, nearly 10% of the trips took place at stores of the Retailer's chain and 5,268 of the households registered at HomeScan shopped there at least once.

Zipcode Business Pattern. The Zipcode Business Patterns (ZBP) are published annually by the US Bureau of Census and contain information on the number of establishments of US located business by industry code, their level of payroll and employment. I use information for the years 2001-2007 on the "Retail trade" industry which is further disaggregated into *supermarkets and other grocery*, *meat markets*, *fish markets*, and *fruit and vegetable markets*. The decision of exploiting zipcode level data, rather than the more commonly used county ones, stems from the fact that the zipcode is the geographic unit for which the Retailer de-

---

[6]See for example Einav, Leibtag, and Nevo (2010) for an in-depth description of the database.

cides on deployment of its web service. This cleanly characterizes the area where the impact of this move is more likely to be felt and provides a handy way to define a market.

3 Measuring business stealing and crowding out

3.1 Identification

I consider a multistop shopping framework in which households potentially split their shopping among several retailers. I observe their 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 assume that these options are substitute[7] and I am interested in measuring the cross elasticity of purchases on the Retailer's website to purchases on its traditional stores and to those made at competitors'. The former measures how much the Retailer cannibalizes its own sales by operating on the Internet channel; the latter quantifies business stealing.

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

$$Total\_Expenditure_{it} = \alpha + \beta Online\_Expenditure_{it} + \varepsilon_{it} \qquad (1)$$

where *Total expenditure* and *Online expenditure* are expressed in 2006 dollars and computed

---

[7]This assumption implies that the introduction of e-commerce does not expand the market and that, at the industry level, online business is a pure reallocation of market shares. Whereas in other retail industries, the Internet "Long Tail" has brought a market expansion allowing customers to purchase goods they would have not found in traditional outlets (Brynjolfsson, Yu, and Smith, 2003), this phenomenon does not fit grocery well. Perishability of the items makes stocking niche goods much riskier. Moreover, the in-store picking business model has kept varieties offered online and in-store close. As a result, in the data the bulk of purchases online and in-store involves the same product categories.

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 business stealing. Out of each dollar a household spends on the online channel, β cents are new business for the chain and, therefore, are due to business stealing. Conversely $(1 - \beta)$ 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.[8]

Relying on longitudinal data also implies that I do not need to worry about changes in the composition of the customer base due to the introduction of the online service. Even if the availability of Internet shopping attracts new customers to the Retailer's chain who are systematically different from historical customers, cross-sectional differences between these households do not contribute to identification.

---

[8]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 4.

3.2 Results

Results from the estimation of equation 1 are reported in Table 2. Column (1) displays a first assessment of business stealing from a pooled regression, controlling only for time effects. The point estimate of β is .74 implying that for every dollar spent online, 74 cents represent purchases the household would have made at competitors and only the residual 26 cents is being displaced from brick-and-mortar sales. This pooled regression without household controls probably overstates the correlation between online and total purchases since certain individual characteristics may be correlated with higher expenditure both online and in-store. Column (2) adds household fixed effects and column (3) replaces them with observable characteristics matched at the block group level from Census 2000. Even after controlling for household heterogeneity, the business stealing effect is strong. The 67 cents per dollar of estimated incremental business imply that households are substituting trips to competitors of the Retailer with orders on its website.

The remaining columns in Table 2 check robustness of this result. In Columns (4) I aggregate grocery expenditure for all the households living in the same zipcode and run a regression analogous to that in column (2) but with zipcode fixed effects. Since households shop online on average less than once a month, a time span longer than that is perhaps needed to identify any displacement that online shopping may induce on traditional sales. Column (5) addresses this concerns by considering quarterly rather monthly household expenditure. Column (6) adds the total number of shopping trips made by the household in the month to the controls. The small change in the estimate of β hints that business stealing does not come through an increased number of trips the household make to the Retailer but it is rather due to large the amount spent when it shops online.

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 stealing from competitors. The final two columns control whether the positive association between online and total sales fades once I take into account the inventory motive (Hendel and Nevo, 2006). Column (7) controls for lagged expenditure in grocery which proxies for household inventory. In that

12

specification, I assume that a household coming out of months with similar level of grocery spending holds a comparable level of inventory. Column (8) takes a different approach at shutting off the potential effect of stockpiling. I reproduce the basic estimates considering only expenditure in perishable grocery products, such as eggs or milk, which cannot be stockpiled.[9]

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

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

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.[10] 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.[11] 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

---

[9]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.

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

[11]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.

over the two years: this represents 18% of the total amount spent by the median household in the sample.

4   Endogenous shopping channel selection

In Section 3, I identified business stealing and crowding out exploiting within-household variation in expenditure in online grocery. As documented before, online orders are infrequent relative to traditional grocery trips, implying that the extensive margin of engagement in e-commerce (i.e. comparison for the same households between months when he spends positive amounts on the web and those when he does not purchase on the Internet) is relevant to identification. Neither the decision of becoming an online shopper nor that of shopping online in a given month are random, however. It follows that the existence of unobserved shocks to demand for grocery correlated with the choice of shopping on the web could compromise a large chunk of my identifying variation. 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 overstate the amount of business stealing. In this section, I check that estimates of business stealing are robust to addressing endogeneity concerns.

A first approach entails eliminating the potentially troublesome variation by dropping from the sample all the months where a household did not shop online. Conditioning on positive expenditure online removes concerns due to the endogeneity in the shopping channel selection; business stealing is identified out of variation in the actual amount spent. This solution is implemented in the first column of Table 3. Estimated business stealing rises to 80 cents for a dollar spent online, even higher than in the baseline results. However, this magnitude does not represent a true estimate of business stealing as much as an assessment of its intensive margin. To better address the endogeneity issue, I exploit features of the data to implement an instrumental variable strategy.

I present results of three instrumental variables estimates. First, I instrument online expenditure with 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, the 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.[12] 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.[13] 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.

The second instrument I use is a dummy taking value one if 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.[14] Identification through coupon

---

[12]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.

[13]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.

[14]The imputation of coupon holding is obviously subject to error. For example, if no household redeems

holding relies on Pozzi (2010) 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 influence 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 analogous to that provided for coupon holding.

Columns (2)-(4) report IV estimates using e-commerce availability, coupon holding and coupon size as instruments, respectively. Each of the specifications still includes household and year-month fixed effects. For each specification, I also report the coefficients from the first stage. All the three 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. Though business stealing is higher when instrumenting with rollout (62 cents per dollar) and lower when using coupon holding (56 cents per dollar), the estimates are quite close across specifications. More important, they are close to the original OLS estimates. In fact, the fixed effects specification in column (2) of Table 2 placed business stealing at 67 cents per dollar spent online.

If we believe in the validity of the instruments, this set of exercises is reassuring. The endogeneity bias is not large enough to reverse the conclusions drawn from OLS estimates: on the Internet the Retailer attracts for the most part trips that households would have otherwise made at the competition. For the rest of the analysis, I will present estimates based on instrumental variables with coupon holding as the instrument. Instrumenting with coupons allows to use the full sample, whereas when using rollout I have to restrict the sample to zipcodes where the service was deployed during the sample span. Moreover, coupon

---

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.

holdings vary considerably within household through time. Therefore, this instrument is better suited to match variation in shopping channel decision than service availability that only changes once for the same household in the time series.

5   Heterogeneous effects

The evidence presented so far points to a large role of e-commerce in reallocating market shares. Once the Retailer starts offering grocery online, some household will prefer shopping there rather than visiting the competition. Which types of customers are more likely to be attracted by the online service? Which product categories can the Retailer leverage on the Internet? In this section I explore cross-sectional heterogeneity in the size of the business stealing effect.

5.1   Heterogeneity across customers

I start by looking at whether business stealing varies with customers' income. I classify households in high-income and low-income based on the median household income of their block group as reported in Census 2000. High income households are those belonging to the top decile (over $107,421) in the income per capita distribution; low income households are those in the bottom decile (below $37,250).[15] I proceed to run the same specification as in Column (2) of Table 2 for the two subsamples separately.

The results appear in columns (1) and (2) of Table 4. For both types of customers, the business stealing is significant. However, it is substantially higher for low income household; whereas for more affluent customers business stealing is below the estimated average effect. The Retailer does not especially try to cater to low income customers. Still its size results in bargaining power with suppliers, which in turns allows to offer better prices and promotions than most of the competition. Because poorer households are more sensitive to good deals, it is natural that they respond more once access to the Retailer is made more convenient by

---

[15]Top and bottom decile in wealth distribution are larger than those of the overall population. This is not surprising given that I am looking at a sample of online shoppers. Wealthier people are more likely to have Internet access (Goldfarb and Prince, 2008) and e-commerce is more likely to be available in more affluent markets.

the online channel. Wealthier households likely decide on the store where they want to shop based on other motives.

I apply the same scheme to assess the effect of distance from stores of the chain. A different response between household living closer to a store of the chain and the ones living further away is to be expected. The latter group was more constrained in its access to the Retailer and may have chosen to shop elsewhere because of convenience. Once the Internet removes such disadvantage, these households may be willing to substitute some of the shopping they used to do at stores closer to their location with online orders on the Retailer's website. For households living close to a supermarket stores of the Retailer instead, the introduction of e-commerce does not alter the picture as much although Internet orders still provide benefit that can induce some substitution (e.g. home delivery). Consistently, online purchases of households in the top decile for distance from the closest store of the Retailer's chain (more than 2.6 miles) are in large part substituting trips that they would have otherwise made at the competition (column (3)). For households living closer to a store (less than .4 miles) cannibalization is 40 cents per dollar: higher than for the average customer (column (4)).

Finally, I examine how the impact of online grocery varies for customers with different levels of loyalty to the Retailer. Because I do not observe grocery purchases at other retailers, I do not know what share of its grocery consumption each household fulfills at the Retailer's stores. I take the total amount spent at the chain for the two years included in the sample as a measure of the household loyalty to the chain. Consumers spending more at the chain are more likely to be purchasing a large fraction of their grocery there. This measure has obvious limits. For instance, high expenditure at the Retailer may simply signal high income or large family size. To mitigate this problem, I only look at households who are homogeneous under both these dimensions and assume that their total grocery consumption should be comparable.

I focus on a subsample of 912 households whose median income is between $25,000 and $35,000 and whose average family size is between 2.3 and 2.7 members. These numbers have been chosen because they describe a symmetric interval around median values. In columns (5) and (6) I present estimates of business stealing for the subsample of households in the

top decile for overall expenditure at the chain in the two year between Jun 2004 and June 2006 (over $17,348) and those in the bottom decile ($1,803 or less). The differences in the estimated effects are stark. When shopping online, less loyal customers are increasing their share of purchase at the Retailer. For them the business stealing is estimated at 74 cent per dollar. For the most loyal customers, I cannot reject the hypothesis that the online channel does not trigger any business stealing at all. An intuitive explanation of this finding is that for customers who are already purchasing most of their grocery at the Retailer, there is not much scope for stealing any business. Opportunities for poaching trips are instead maximal with occasional customers. Therefore; the business stealing potential of the Internet channel seems related to the ability of making occasional customers shop more often at the chain.

One lingering concern from the approach followed above is that total expenditure could capture difference in tastes rather than loyalty. Two identical customers who are equally loyal to the Retailer would not appear so if one of them likes to purchase more expensive goods and therefore spends higher amounts. As a robustness check, I repeat the exercise using the number of trips over the two years as a measure of loyalty. This metric is independent from the cost of the basket of goods purchased. I report the results in columns (7) and (8) for the most loyal (320 trips or more) and least loyal (less than 32 trips) respectively finding similar results.

5.2   New customers

Besides inducing current customers of the chain to increase their share of purchase at the Retailer, deployment of the online business operation may convince new customers to start shopping there. New customers are an extreme case of business stealing as for them there is no risk of cannibalization. Here I examine whether a flow of new customers joins the Retailer's chain in coincidence with the rollout of the service in a given zipcode. I treat as "new customers" all the households who start shopping at the grocery chain only after the first six months of my data span have elapsed.

Column (1) of Table 5 presents estimates of the effect of offering e-grocery on the probability that the chain will attract at least a new customer in a given zipcode-month. I

find that the Retailer is more likely to catch new clients in zipcodes where it has already rolled out online shopping. The result is robust to the inclusion of zipcode fixed effects (column (2)). The gap in the probability of attracting new customers before and after the Internet service is introduced in a zipcode is 11 percentage points. This means that introducing e-commerce in the zipcode doubles the probability of acquiring a new customer.

I reach a similar conclusion when looking at the intensive margin (column (4)). The estimate from a Poisson count model implies that the number of new customers gained by the chain in a zipcode-month where e-commerce is available is twice as large as that which they used to gain in the same zipcode when the online division was not delivering in the area. The increase in the appeal to new customers is not immediate. Column (5) shows the result of a linear probability model where the dummy for having attracted any new customer is regressed on an indicator that take value one only in the month of rollout. The probability of catching new clients in the very month of introduction of the online service is not significantly different from that of any other month. The most likely explanation of this finding is that it takes some time for the news of the rollout to reach customers who are not currently shopping at the chain and are possibly harder to reach for the Retailer's informative advertising.

I have implicitly assumed that customers joining the chain for the first time in zipcode reached by the e-commerce service do so *because* of it. I test this assumption by comparing shopping behavior for new customers who joined the chain in a zipcode where shopping online was possible and new customers who joined in a zipcode when that option was not available yet. If the former group is driven by desire to exploit online shopping while the latter is not, they should display differences in their demand for online grocery. The top panel of Figure 1 shows a kernel density estimate of the ratio of online to total expenditure in grocery for households in the two groups. The bottom panel does the same for the distribution of the ratio of online to total shopping trips. To make the two groups comparable, the figures for customers who joined before e-commerce was introduced are computed keeping into account only trips occurred after they first shopped online. The distribution of the share of online to total expenditure for customers first purchasing at the chain in areas where e-commerce was available shows more mass in the right tail; the same holds for the ratio of online trips

to total trips. People joining in zipcodes offering Internet shopping make more intensive use of the online option in their subsequent shopping history. This is consistent with the idea that the availability of this particular service drew them to the chain in the first place.

5.3   Heterogeneity across product categories

Grocery retailers sell a wide and heterogeneous set of goods. So far, I have been abstracting from this aspect by looking at the aggregate amount spent in grocery. Understanding which categories a retailer is more able to leverage online is of obvious interest for managers. Furthermore, it sheds light on the nature of the business stealing induced by the Internet channel. I look at two forms of heterogeneity across items sold. First, I assess whether there is any difference in business stealing between popular and niche products. Next, I analyze how the effect varies according to the strength of the Retailer in a particular product category.

To perform the analysis, I combine the data on household purchases at the Retailer's with the Nielsen Homescan panel that contains information on grocery purchases at *every* retailer visited by a *different* sample of households. The Homescan panel serves two purposes. Exploiting the fact that it covers a representative sample of consumers, I use sales reported there to compute the share of total expenditure in grocery (wallet share) accounted for by each category. This statistic is used to separate the products into "head" products, accounting for a large share of grocery purchases, and "tail" products, bought by a small fraction of consumers. Because Homescan data contain shopping trips at each grocery shop visited by households in the sample, they can also be used to calculate the market share held by the Retailer in each product categories. This way, I can distinguish between categories where the Retailer is strong and those where it is a fringe player. Since I have Homescan data available only for 2004, the exercise will assume that both wallet share by product category and the Retailer's market share in each of them do not swing to largely in 2005 and 2006.

I estimate equation 1 separately for each product category, obtaining an estimate of business stealing for each of them. I focus on the product categories with the fifty largest wallet share to ensure a large enough sample size for each of the regressions. I also discard the three product categories for which the estimated coefficient was not significant at the

usual levels. In Figure 2 I plot the estimated business stealing for each category against its wallet share. I find a positive and significant relation between the two suggesting that selling online allows to steal more sales in staple categories. The result contrasts with findings for other industries where the Internet is mostly used to purchased "tail" goods that would be hard to find at brick-and-mortar retailers.

It is important to notice that popular items (bread, milk, soft drinks, cereals, etc.) are generally at the core of the offer of traditional mom-and-pop food stores which rarely carry niche items. Therefore, they are primed to suffer a serious treat from online grocery providers. Small stores focused on ethnic or special items should stand a better chance to fend off the challenge brought by the new channel.

The Retailer is not equally successful in each of the product categories carried: some of them could be more distant from the core business, in others the competition could be stronger. As a results, it holds heterogeneous market shares for different product lines. This suggests testing whether crowding out is higher in product categories where the retailer has a stronger hold. As in the similar exercise looking at consumers' loyalty, I expect that there should be less scope for the chain to steal sales in segments where it is already dominant. Therefore business stealing should be lower in those categories.

Figure 3 displays on the x-axis the market share held by the Retailer in a given product category and on the y-axis the estimated business stealing in that category. Most of the variation in the retailer's market shares for different product is confined in the range between 8% and 40%, although for some categories the chain has a market share above 50%.[16] Business stealing also differs greatly for different goods being limited to 20 cents per dollars in some cases but getting as high as 90 cents per dollar. The plot suggests a negative relationship between market share held and business stealing: crowding out is more severe for product categories in which the retailer is already strong. However, I cannot reject the hypothesis that the two are uncorrelated.

---

[16]Results are robust to excluding product categories that are outliers.

6   Who is Hurt by E-commerce?

The evidence presented so far shows that the introduction of Internet grocery allows for substantial business stealing. Who is losing the sales stolen by online retailers? In this section I try to address such question by looking at the impact of the introduction of e-commerce on the equilibrium market structure in the grocery retail sector.

Since I do not observe revenues nor profits at grocery stores other than the Retailer's, I follow an indirect approach to measure the effect of the deployment of online grocery on competitors profits. I assume that every firm active in the market should be making nonnegative profits and will decide to exit otherwise. When the Retailer introduces online shopping, it should steal business its competitors hurting their profits. As a result, we should observe some exit following the rollout of e-commerce.

I use data provided by the Retailer reporting the date in which it became possible to order online in each of the 1,203 zipcodes reached by the service between January 2001 and July 2007. I rely on the Zipcode Business Patterns for information on the number of *supermarkets and other grocery* (NAICS 445110) operating in a given zipcode in the same time interval. The choice of defining a zipcode as the relevant market seems the best suited to capture the effect of e-commerce. In fact, the introduction of the service in a zipcode implies that all and only the households living in that zipcode can now purchase online. It seems natural that the impact should be felt most by stores sited in that same location. I regress the number of supermarket on a constant and a dummy indicating whether the Retailer was offering online shopping in the zipcode. Since I am including zipcode fixed effects, the estimated coefficient will document whether introduction of e-commerce in a given zipcode has resulted in a variation in the number of grocery stores active in the area. Time fixed effects are also included in all specifications to account for macroeconomic trends.

Data on e-grocery availability come from a single retail chain, I do not observe whether and when other grocers may have introduced online services similar to that offered by the Retailer. Such missing information may lead to misjudge the impact of the rollout of the Retailer's web service but does not affect the big picture conclusions on the effect of online shopping on competition. If the Retailer is the first to enter the online segment in a market,

then I will correctly measure the impact of its rollout which also coincides with the effect of Internet commerce availability. If some chain is already selling online when the Retailer decides to enter the market, this will likely mean that e-commerce is already having an effect on the number of stores active in equilibrium. For instance, some stores may have already shut down business. Therefore, the reaction to the Retailer's rollout will be attenuated and my estimates are biased against finding an effect of online entry. It follows that the results presented below can be interpreted as lower bounds on the effect of Internet commerce on market structure.

Table 6 summarizes the results. The first column shows that the overall number of supermarkets and food stores active in a zipcode does not change after the Retailer introduces the online service. The estimated coefficient is negative but small and not statistically significant. The picture, however, changes when I look at differential impact for small versus large stores. I follow Jia (2008) in defining small stores as those employing less than 20 people. Large stores are instead defined as those counting 100 employees or more. This definition is likely to be correlated with the ownership structure of the store (information not provided in the Zipcode Business Patterns). Independent stores tend to be of small size; whereas outlet of big chains are bigger and employ more people. In the following analysis I will refer to the small stores also as "independent" or "mom-and-pops" and I will assume that the large stores are likely part of a supermarket chain. Columns (2) and (3) present estimated of the effect of Internet rollout on the number of small and large stores respectively. Introducing online grocery in a market the Retailer displaces about .18 small grocery stores, a 5.6% reduction. This is about one-tenth of the effect of the entry of a Wal-Mart store documented in Jia (2008) IV estimates. The effect on the number of large supermarket stores is instead positive and not significant: big chains outlet are not affected by the introduction of e-commerce. The result is robust to the definition of "small store". In column (4) I follow Haltiwanger, Jarmin, and Krizan (2010) and define a small store as one employing less than 10 people without affecting the results.

Why is Internet commerce disruptive for small outlets but not for large chains? This finding supports that online grocery allows big-box stores to overcome their location disadvantage with respect to small, urban ones. The effect of e-commerce is, therefore, asymmetric.

Mom-and-pop's now face competition from low-price and high-variety supermarket chains even for the set of customers with high disutility of travel, who were before captive. The large grocers face little change. They competed with the Retailer on the set of customers willing to travel to purchase groceries. The store choice of these households should not be swayed by the opportunity to have home delivery.

Columns (5)-(7) allow for the effect on exit of -respectively- all, small, and large supermarket stores to be different according to the time elapsed since the Retailer offered Internet shopping. Consistently with the findings commented above, it is the small stores who are affected whereas there is no impact on the profits of large chains. Net exit is observed within the first twelve months since the introduction of the Internet service and its rate is constant through time. Given the data at hand, I can only estimate precisely the impact of e-commerce availability up to the two years following its introduction in a market. Longer run effects are pooled in a single coefficient. It is possible that the direst consequences of e-grocery access will play out only in the long term. In fact, take up of the new service may be slow and it may take time before the dent it creates in the profit of independent stores is large enough to trigger exit. Therefore, the full blast of Internet commerce introduction on the competitive outcomes may be even larger than what I estimate.

To further assert the causality of the results presented so far, I implement a placebo test checking whether e-commerce introduction had any effect on the number of small greengrocers, fish and meat markets in the same time span. Produce, fish and meat are "high touch" goods. Their nature makes them less fit to be shopped online because the Internet does not allow for inspecting the merchandize or choose the cut. Therefore, we would expect that despite deriving a large share of their customer base from their location advantage, local produce, meat and fish market should not be affected by e-grocery as much as independent stores selling packaged goods. Table 7 shows that the impact of Internet grocery introduction on the number of small stores in these three particular categories is not significant. The result survives even once we allow for differential effects in time.

One lingering concern is linked to the fact that the analysis is only based on partial equilibrium. I analyze the effect of the Retailer's introduction of e-grocery but do not consider that its competitors could respond doing the same. If other grocers were allowed

to respond to the Retailer introduction of e-grocery, would the conclusions change? If the engine of the business stealing effect is indeed the demise of the location strategy that allowed small scale stores to compete with the more attractive offer of large chains, this should not be the case. Other supermarket chains selling online would bring further harm to independent stores, giving their captive customers one more alternative to substitute to. All the chains distributing on the Internet would share the gains from business stealing, affecting how much each one profits from the online channel. The bottom line on winners and losers from the introduction of e-commerce in the market would be unaffected, though. As for mom-and-pop's, even if they could enter the online segment[17] they should not be able to steal any business online. They could reach more people but their offer should prove unappealing compared to that of competitors. In fact, that is what had driven them to exploit the location advantage in the first place.

7   Conclusions

The role of the diffusion of e-commerce in the demise of location advantage has long been theorized. I provided direct evidence of this phenomenon in the grocery industry by showing that the introduction of Internet commerce benefits large suburban stores and hurts the small grocers, usually more centrally located. This happens because selling on the Internet allows the big-box stores to leverage their better prices and product selection on customers with high travel cost who were before served by local independent shops. The burden of increased competition is large enough to cause exit of mom-and-pop's and further concentration of market shares in the hands of large supermarket chains.

The results suggest that take up of online grocery should be factored in as an important ingredient when trying to understand the evolution of the industry in the near future. A widespread adoption of this technology will trigger further concentration in a business where the relevance of travel costs had so far ensured high fragmentation relative to other retail sectors. That early deployment of online grocery shopping is already showing an impact on

---

[17]This is an unlikely occurrence. Given the nature of the business the cost of supporting Internet orders with home delivery to a wide enough area are prohibitive for small scale, single-store grocers.

the competitive environment is all the more impressive when we consider that even Wal-Mart entry into grocery seems to have failed to make waves.

An interesting avenue for future research is to examine the reaction of small neighborhood grocery stores to the changed environment. Evidence from this study hints that even after losing the shield of location advantage, they could survive competition of large chains by pursuing other forms of differentiation. For example, they could focus on niches such as ethnic or gourmet food. Therefore, if detailed data on the format of small grocery stores were available, we should be able to detect a shift towards such specialized offers. This is reminiscent of the repositioning of local papers' news content observed by George and Waldfogel (2006) after the New York Times started circulating nationally.

References

BASKER, E., AND M. D. NOEL (2008): "The Evolving Food Chain: Competitive Effects of Wal-Mart's Entry into the Supermarket Industry," *Journal of Economics and Management Strategy*, 17(2), 345–377.

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.

BRYNJOLFSSON, E., J. H. YU, AND M. D. SMITH (2003): "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science*, 49(11), 1580–1596.

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

EINAV, L., E. LEIBTAG, AND A. NEVO (2010): "Recording Discrepancies in Nielsen Homescan Data: Are They Present and Do They Matter?," *Quantitative Marketing and Economics*, 8(2), 207–239.

ELLICKSON, P. B. (2007): "Does Sutton apply to supermarkets?," *Rand Journal of Economics*, 38(1), 43–59.

ELLICKSON, P. B., AND P. L. GRIECO (2011): "Density versus Differentiation: The Impact of Wal-Mart on the Grocery Industry," working paper.

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.

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

GEORGE, L. M., AND J. WALDFOGEL (2006): "The New York Times and the Market for Local Newspapers," *American Economic Review*, 96(1), 435–477.

GOLDFARB, A., AND J. T. PRINCE (2008): "Internet adoption and usage patterns are different: Implications for digital divide," *Information Economics and Policy*, 20(1), 2–15.

GOLDMANIS, M., A. HORTACSU, C. SYVERSON, AND O. EMRE (2010): "E-commerce and Market Structure of Retail Industries," *Economic Journal*, 120(545), 651–682.

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.

GRIECO, P. L. (2010): "Discrete Games with Flexible Information Structures: An Application to Local Grocery Markets," working paper.

HALTIWANGER, J. C., R. JARMIN, AND C. J. KRIZAN (2010): "Mom-and-Pop Meet Big-Box: Complements or Substitutes?," *Journal of Urban Economics*, 67(1), 116–134.

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

HOLMES, T. J. (2001): "Barcodes Lead to Frequent Deliveries and Superstores," *Rand Journal of Economics*, 32, 708–725.

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

JIA, P. (2008): "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 76(6), 1263–1316.

KEENEY, R. L. (1999): "The Value of Internet Commerce to the Customer," *Management Science*, 45, 533–542.

KROFT, K., AND D. G. POPE (2011): "Does Online Search Crowd out Traditional Search and Improve Matching? Evidence from Craigslist," working paper.

LIEBER, E., AND C. SYVERSON (2011): "Online vs. Offline Competition," in *Oxford Handbook of the Digital Economy*, ed. by M. Peitz, and J. Waldfogel. Oxford University Press, forthcoming.

MATSA, D. A. (2011): "Competition and Product Quality in the Supermarket Industry," *Quarterly Journal of Economics*, forthcoming.

POZZI, A. (2010): "Shopping Cost and Brand Exploration in Online Grocery," working paper.

SCHIVARDI, F., AND E. VIVIANO (2011): "Entry Barriers in Retail Trade," *Economic Journal*, 121, 145–170.

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.

SMITH, H. W. (2004): "Supermarket Choice and Supermarket Competition in Market Equilibrium," *The Review of Economic Studies*, 71(1), 235–263.

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: Measures of online crowding out and business stealing

Dependent variable:				_Total expenditure_				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Online expenditure	.74**	.67**	.70**	.72**	.69**	.61**	.67**	.67**
	(.019)	(.012)	(.020)	(.019)	(.019)	(.012)	(.010)	(.013)
Num. of trips						25.6**		
						(.436)		
Total expenditure ($t-1$)							.16**	
							(.008)	
Fixed effects								
Year-month	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Year-quarter	No	No	No	No	Yes	No	No	No
Household	No	Yes	No	No	Yes	Yes	Yes	Yes
Zipcode	No	No	No	Yes	No	No	No	No
Sample restrictions								Non storable
R^2	.22	.22	.26	.22	.28	.5	.4	.27
N households	9,323	9,323	8,019	8,019	9,323	9,323	9,234	9,318
N obs.	196,148	196,148	172,113	171,198	72,736	196,148	186,825	192,237
Implied value of the online channel	14M	12.7M	13.3M	13.7M	13.1M	11.6M	12.7M	12.7M

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 in columns (1)-(3) and (6)-(8) and household-quarter in column (5); standard errors (in parenthesis) are clustered at the household level. In column (4) the unit of observation is zipcode-month and standard errors are clustered at the zipcode level. In column (3) 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. These coefficients are not reported for parsimony but full results are available upon request. In column (8) 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%

32

Table 3: Online crowding out and business stealing: instrumental variables estimates

	OLS	IV		IV		IV	
		Indicator for date of rollout		Indicator for coupon		Amount of coupon	
	(1)	1^{st}stage	(2)	1^{st}stage	(3)	1^{st}stage	(4)
Online expenditure	.80**		.62**		.56**		.60**
	(.004)		(.099)		(.029)		(.030)
Rollout indicator		41.7**					
		(2.5)					
Coupon indicator				50.9**			
				(1.1)			
Coupon size						5.3**	
						(.12)	
N obs.	66,624	31,132	31,132	167,590	167,590	167,590	167,590
R^2	.05	.14	.11	.05	.22	.04	.22

Notes: The dependent variable is household total monthly expenditure in grocery at the retailer's chain. Standard errors (in parenthesis) are clustered at the household level and zipcode level respectively. The *Rollout indicator* is a dummy that takes value 1 in each month after the Internet service has been available in the zipcode of residence of the household. The *Coupon indicator* is a dummy that takes value 1 if the household held a coupon for free or discounted Internet delivery. Interpretation is as in Table 2: the coefficient on *online expenditure* (β in equation 1) represents business stealing and $(1 - \beta)$ gives an estimate of crowding out. Column (1) only considers household-months with strictly positive expenditure on the online channel. In column (2) only households living in the 351 zipcodes where the Retailer introduced online grocery between June 2004 and June 2006 are considered. Figures are in 2006 dollars. All specifications include month-year and household fixed effects. Significance levels :* : 5% ** : 1%

Table 4: Heterogeneity in business stealing across households

	Income		Distance from store		Loyalty: expenditure		Loyalty: trips	
	top 10% (1)	bottom 10% (2)	top 10% (3)	bottom 10% (4)	top 10% (5)	bottom 10% (6)	top 10% (7)	bottom 10% (8)
Online expenditure	.55**	.73**	.74**	.60**	.27	.76**	.30	.79**
	(.105)	(.071)	(.090)	(.082)	(.280)	(.194)	(.192)	(.286)
Fixed effects								
Household	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	18,794	13,371	10,662	10,676	2,231	1,224	5,530	1,029
R^2	.23	.45	.26	.21	.27	.53	.24	.61

Notes: This table reports instrumental variables estimates of business stealing for subsamples of the population of customers. The instrument for online expenditure is availability of coupon in a given month. Columns (1) and (2) report results for households in the top and bottom decile of income distribution, where median family income is matched at the block group level from Census 2000. Columns (3) and (4) display estimates for customers at top and bottom deciles in terms of distance from the closest store of the Retailer. The distance is computed "as the crow flies" using geodesic coordinates of the address of residence of the household and that of the location of the supermarket store. Columns (5)-(8) split customers using proxies for their loyalty to the chain: total expenditure in grocery and total number of trips over the two years. Estimates in columns (5)-(8) are based on the 912 households with income per capita between \$25,000 and \$35,000 and with average size of the household between 2.3 and 2.7. Standard errors (in parenthesis) are clustered at the household level. Significance levels :* : 5% ** : 1%

34

Table 5: Impact of selling online on the probability of attracting new customers

Dependent variable	(1) New customers dummy	(2) New customers dummy	(3) New customers dummy	(4) Number of new customers	(5) New customers dummy
Online available	.03**	.11**	.07**	.14**	
	(.003)	(.027)	(.013)	(.032)	
Rollout month					.03
					(.036)
Zipcode fe	No	Yes	Yes	Yes	Yes
N obs.	24,030	2,642	2,642	2,642	2,642
R^2	.02	.03	.13	.03	.03

Notes: This table reports the effect of offering online shopping on the probability that the Retailer attracts new customers. A new customer is defined as a household that has never shopped at the chain in the first six months of data in the sample (June 2004-December 2004). *New customers* takes value 1 if at least one new customer showed up at the chain in a given zipcode-month. *Number of customers* is a count variable for the number of new customers joining the chain in a given zipcode-month. *Online available* takes value 1 in each month after the introduction of online grocery in a zipcode. *Rollout month* takes value 1 only in the exact month when the Internet grocery service was first made available in the zipcode. Specification including zipcode fixed effects (Column (2)-(5)) only include the 352 zipcode where the service was rolled out between june 2004 and june 2006. Columns (1), (2), and (5) report results from a linear probability model. Column (3) refers to a probit specification and column (4) to a Poisson one; for both marginal effects are reported. Year-month dummies are included in all specifications and standard errors (in parenthesis) are clustered at the zipcode level. The mean of the *new customer* dummy is .103 (standard deviation .303); the mean of the *number of new customers* variable is .113 (.350) Significance levels :* : 5% ** : 1%

35

Table 6: Who is hurt by e-commerce? Effect of the introduction of the online grocery on the number of active supermarket and other grocery stores, by size

|	All stores (1)	Small stores (2)	Large stores (3)	Small stores (4)	All stores (5)	Small stores (6)	Large stores (7)
Post rollout	-.01	-.18***	.06	-.16***			
	(.050)	(.043)	(.078)	(.049)			
First year post rollout					.02	-.16***	.04**
					(.048)	(.042)	(.016)
Second year post rollout					-.02	-.19***	.06
					(.058)	(.051)	(.052)
Over two years post rollout					-.01	-.19***	.10
					(.070)	(.061)	(.076)
N obs.	8,307	8,307	8,307	8,307	8,307	8,307	8,307

Notes: This table relates presence of online grocery offered by the Retailer to the number of active supermarket and other grocery stores. The dependent variable is the number of supermarkets active in the zipcode. The dependent variable is the number of establishments of the type specified on top of each column. *Post rollout* is a dummy variable that takes value one in each year where the Retailer offers online grocery in the zipcode. Specifications in columns (5)-(7) allows for heterogeneous effects in the time after the Retailer introduces online grocery in the market (1-12 months, 12-24 months, over 24 months). Column (4) uses a more restrictive definition of small food store. Standard errors (in parenthesis) are clustered at the zipcode level. Year and zipcode fixed effects are included in all specifications. Results are based on all the 1,203 zipcodes where the Retailer introduced online grocery between 2001 and June 2007. Significance levels:∗ : 10% ∗∗ : 5% ∗ ∗ ∗ : 1%

Table 7: Effect of the introduction of the online grocery on the number of active on small meat markets, fish markets and fruit and vegetables markets.

	Produce mkts (1)	Meat mkts (2)	Fish mkts (3)	Produce mkts (4)	Meat mkts (5)	Fish mkts (6)
Post rollout	.01 (.013)	-.03 (.019)	.01 (.009)			
First year post rollout				-.01 (.013)	.01 (.009)	-.02 (.015)
Second year post rollout				-.00 (.014)	.00 (.011)	-.02 (.018)
Over two years post rollout				.04* (.019)	.01 (.011)	-.04 (.022)
N obs.	8,307	8,307	8,307	8,307	8,307	8,307

Notes: This table relates presence of online grocery offered by the Retailer to the number of small fruit and vegetable, meat, and fish markets active in the zipcode. The dependent variable is the number of establishments of the type specified on top of each column. *Post rollout* is a dummy variable that takes value one in each year where the Retailer offers online grocery in the zipcodeSpecifications in columns (4)-(6) allows for heterogeneous effects in the time after the Retailer introduces online grocery in the market (1-12 months, 12-24 months, over 24 months). Standard errors (in parenthesis) are clustered at the zipcode level. Year and zipcode fixed effects are included in all specifications. Results are based on all the 1,203 zipcodes where the Retailer introduced online grocery between 2001 and June 2007. Significance levels:∗ : 10% ∗∗ : 5% ∗ ∗ ∗ : 1%

Figure 1: Intensity of usage of online shopping for new customers
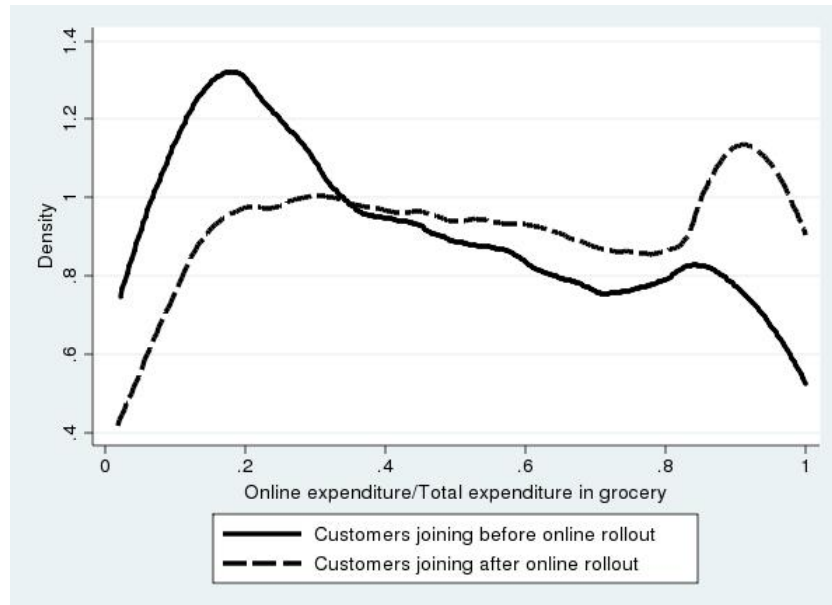
(a) Ratio of online to total grocery expenditure
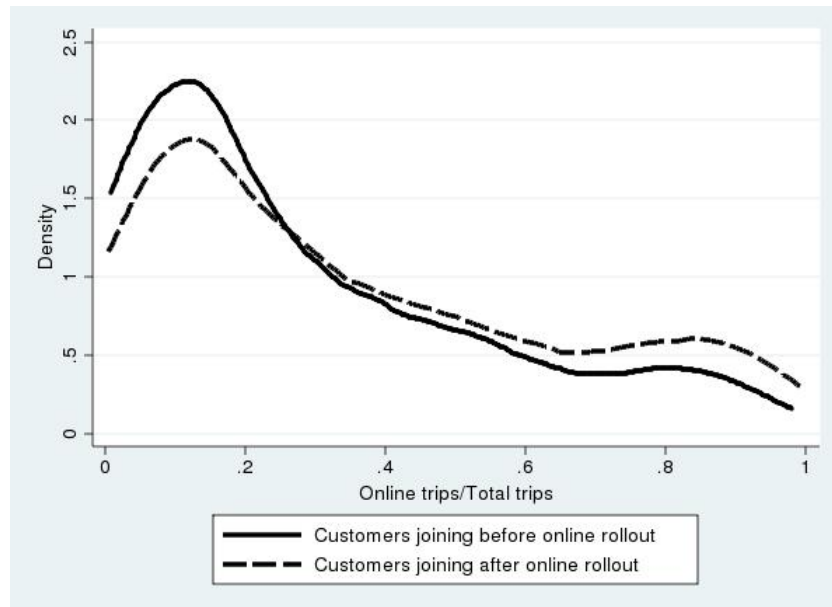

(b) Ratio of online to total grocery trips

Notes: The figures show kernel density estimates of share of online expenditure (panel (a)) and of the share of online trips (panel (b)) for new customers (i.e. households who make their first trip at the chain six months into the data period covered by the sample). Each figures plots two densities: those of new customers who joined in a zipcode where it was already possible to shop online (dashed line) and those who joined in a zipcode where, at the moment of their first trip, online grocery was not offered by the Retailer (although it became available later on). The first group counts 959 households while the second consists of 485 households. A Kolmogorov-Smirnov test rejects equality of the two distributions in both figures.

Figure 2: Heterogeneity in crowding out across product categories: Head vs. tail products
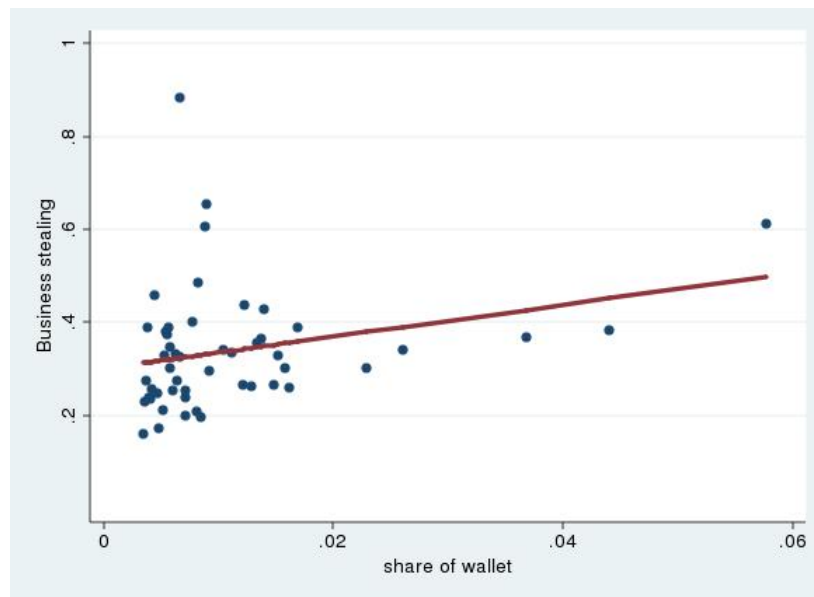
Notes: Each dot in the figure represents a product category. On the horizontal axis is the *wallet share* of the category: the share of total sales represented by the category among all the categories included in the matched Retailer-HomeScan data (see Appendix for detail). On the vertical axis is a category specific estimate *business stealing*. This measure is computed by estimating the model in equation 1 category by category and taking the coefficient β as estimated business stealing. In the regression, *online expenditure* is instrumented using coupon holding. The solid line is the conditional mean of a regression of the estimated business stealing on the wallet share of the category. The slope is 3.41 (with a standard error of 1.72). Only categories with the 50 highest wallet shares are included.

Figure 3: Heterogeneity in crowding out across product categories: Large vs. small market share for the Retailer's chain
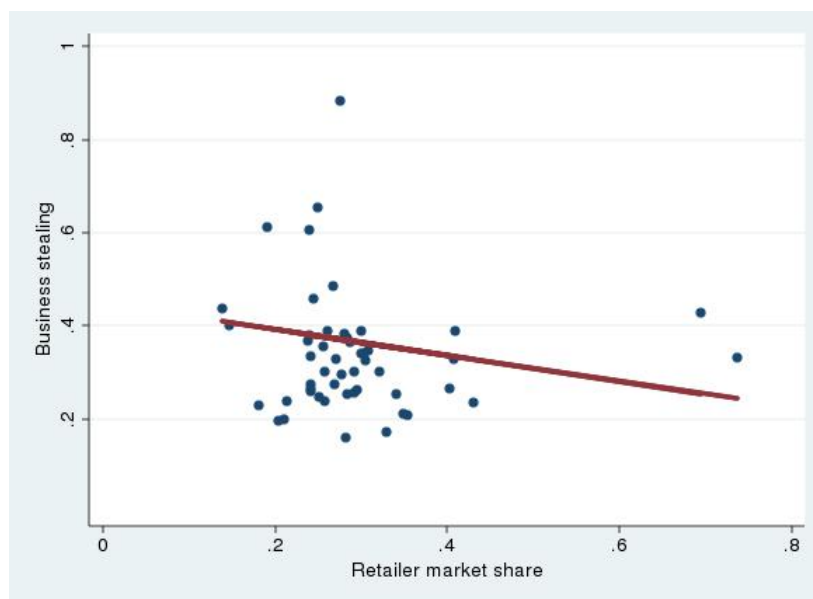
Notes: Each dot in the figure represents a product category. On the horizontal axis is the *market share* held by the Retailer in the category. It is calculated using all the categories included in the matched Retailer-HomeScan data (see Appendix for detail). On the vertical axis is a category specific estimate *business stealing*. This measure is computed by estimating the model in equation 1 category by category and taking the coefficient β as estimated business stealing. In the regression, *online expenditure* is instrumented using coupon holding. The solid line is the conditional mean of a regression of the estimated business stealing on the wallet share of the category. The slope is -.28 (with a standard error of .25). Only categories with the 50 highest wallet shares are included.

A Data appendix

In this section, I provide additional detail on the data sources used for the study and on how they have been combined in the analysis.

A.1 Data on households online and in-store purchases

As noted in the main text, household purchases come from scanner data on individual transactions provided by a national supermarket chain operating traditional stores and selling online. They include shopping trips made by all the households who had been shopping *at least* once in-store and *at least* once online at the chain between June 2004 and June 2006. Purchase data are retained only for those customers shopping with the supermarket loyalty card. Loyalty card usage is high (above 80%) because presenting the card at the cashier is the only way to get access to weekly promotions. The Retailer links all the individual cards belonging to the same household (husband, wife, etc.) to a unique household identification number.

The initial sample includes 11,640 households. To be conservative, I drop 154 households who shop only online or only in-store in the sample period and the 2,165 who place an Internet order worth less than $50. Neither thing was supposed to happen given the description of the sample and the rules to qualify for order online. The results are not sensitive to including these "suspicious" households. The final set consists of 9,323 households distributed over 15 states and 1,344 zipcodes and shopping at 1,423 different stores.

Each trip is associated to a unique identifier which is linked to the identifier of the household performing it, and to the date and the store where it took place. For each trip, all the UPC's purchased in the transaction are listed with the quantity bought and the total amount spent for it, both gross and net of discounts. A difference between the gross and the net amount spent for a UPC implies that the UPC was in promotion when it was bought. Other coupons, such as those giving discount on the delivery fee for online orders, have dedicated UPC's and show as items with negative gross expenditure in the data.

When analyzing individual product categories, I adopt a classification created by the Retailer that aggregates the UPC's into homogeneous groups of goods. Examples of product

categories are "Canned fish", "Laundry detergent", and "Toothpaste". Shoppers in the data purchase products belonging to over 700 different product categories but the largest 200 categories account for nearly 85% of the sales.

Total expenditure for a trip is obtained from the data described above by summing expenditure, net of any discount, in all the UPC's listed for a particular trip ID. For online orders, the fee for home delivery is not included in the total expenditure. Since a date is listed for each trip, it is easy to aggregate the information and obtain the total monthly expenditure in grocery online and in-store for every households. Those are the two main variables used in the analysis.

A.2 Data on households characteristics

Household characteristics come from two sources. The Retailer itself provides information on the distance between each household's residence and the closest store of the chain, as well as the Census block group where the household lives. This information is then used to match Census 2000 demographics aggregated at the block group level with the purchase data.

For 1,301 households the information on distance from the closest store and block group of residence was not available. They are therefore excluded from all exercises which require such information. Households for which the information was not available do not appear to be significantly different from the rest of the sample as far as expenditure in grocery and frequency of purchase are concerned.

A.3 Nielsen Homescan data and construction of product category level variables

Nielsen Homescan database collects similar information to that provided by the Retailer. However, Homescan includes markets all across the US; whereas the Retailer operates only in a subset of them. Furthermore, whereas the retailer data only document purchases at a single chain, Homescan should capture grocery consumption occurring at every store. The Homescan sample used for the study covers the whole year 2004 and only includes purchase

of food items.

I use Homescan transactions only from stores located in zipcodes where at least one of the households in the Retailer data lives. Items in the Retailer and Nielsen data are matched using the UPC's, which should be the same across different stores.[18] Overall, I some 40,000 UPC's are found in both datasets. I group them into product categories using the classification provided by the Retailer rather than the one constructed by Nielsen. Upon inspection, the two categorizations appear broadly consistent with each other.

The wallet share of a category j, used in Figure 2, is constructed as follows. Define $expenditure_j$ as the sum of the expenditure in all UPC's belonging to category j by all the households in all the stores included in the Homescan refined in the way described above. Then

$$Wallet\_share_j = \frac{expenditure_j}{\sum_{j' \in J} expenditure_{j'}} \qquad (3)$$

where J is the set of all the product categories in the data.

To ensure that the sample was large enough, in the analysis I focused on the 150 product categories with the largest wallet share.

The Homescan data also allow to calculate the market share of a particular store or chain s in a given category j, which is used in Figure 3. Define $expenditure_{js}$ as the sum of the expenditure in all UPC's belonging to category j by all the households in store (or store chain) s. Then

$$Market\_share_{js} = \frac{expenditure_{js}}{\sum_{s' \in S} expenditure_{s'}} \qquad (4)$$

where S is the set of all the stores in the data.

A.4 Zipcode business patterns

The number of grocery stores active in a market used in Table 6 come from the Zipcode Business Patterns. They show the count by zipcode of establishments for businesses separately for each sector of activity in the NAICS classification. I use data from 2001 to 2007

---

[18]One exception to that is the food sold by weight, as measured with a scale at the shop. In such case, it is possible that different stores use different codes to identify the item, making the match impossible. This is common for most non packaged goods sold at a supermarket such as produce and fresh bread. However, prepacked produce and bread would be denoted by the same UPC at the Retailer and other grocery stores.

for all the zipcodes where the Retailer decided to introduce the service.

The Zipcode Business Patterns report information on employment at the establishment, which I exploit to identify small and large stores. The former count less than 20 employees; whereas the latter employ over 100 people. Summary statistics on the number of establishments by size are reported in Table A1.

Table A1: Number of establishments in a zipcode, by size.

| | Mean | Std. dev. | Percentiles | | | | |
			10th	25th	50th	75th	90th
All store	6.17	4.91	1	3	5	9	13
Small stores	3.20	3.71	0	1	2	4	8
Large stores	.89	1	0	0	1	1	2

Notes: Small stores list employment below 20 people; large stores employ more than 100 workers.

B Instrumental variables strategy

B.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.[19] 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

---

[19]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:$*: 10\% **: 5\% ***: 1\%$

46

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.

Figure B1: Retailer pricing strategy before and after introducing online grocery, selected zipcodes
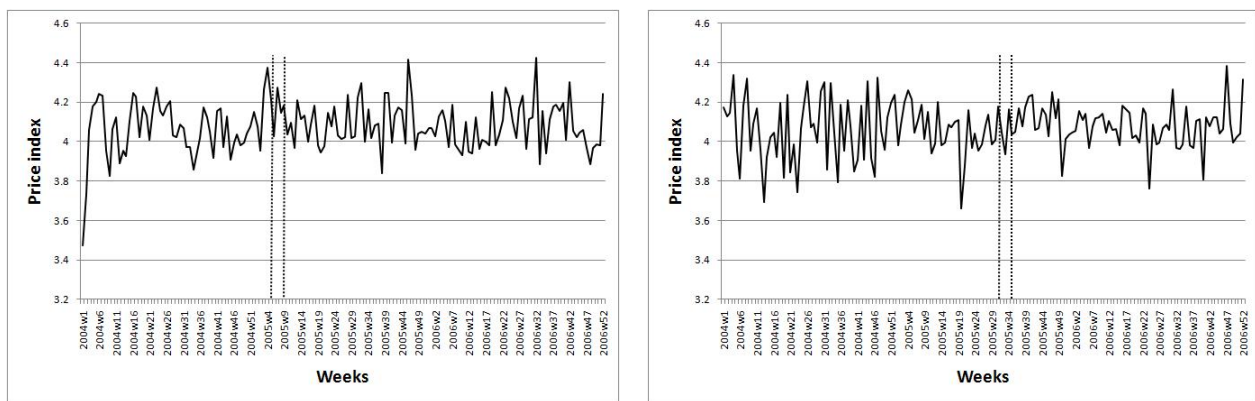
| (a) Zipcodes with rollout in February 2005 | (b) Zipcodes with rollout in August 2005 |

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.

B.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 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.