IEF

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 08/13
March 2013

E-commerce as a Stockpiling Technology: Implications for Consumer Savings

by

Andrea Pozzi

(EIEF)

E-commerce as a Stockpiling Technology: Implications for Consumer Savings *

Andrea Pozzi†

March 29, 2013

Abstract

Shopping on the Internet spares customers the discomfort of carrying around heavy and bulky baskets of goods, since the service usually includes home delivery. This makes e-commerce a technology well suited to helping consumers to buy in bulk or to stockpile items on discount. I use grocery scanner data provided by a supermarket chain selling both online and through traditional stores to show that the introduction of e-commerce leads to an increase in bulk purchase and stockpiling behavior by customers. Since bulk and discounted items are sold at a lower price per unit, my findings highlight a new dimension in which online shopping can be beneficial to consumers. According to my calculations, the reduction in the cost of stockpiling triggered by the introduction of electronic commerce generates significant savings.

Keywords: E-commerce, Retail, Stockpiling, Consumer savings

JEL classification: D22, L21, L81

*This paper is based on the third chapter of my Stanford PhD dissertation. I am thankful to Liran Einav, Tim Bresnahan, and Jakub Kastl for their advice and encouragement. I also thank the Editors, two anonymous referees, Tomás Rodriguez Barraquer and participants in the Stanford IO Workshop for useful comments and suggestions. I gratefully acknowledge financial support from SIEPR, in the form of the B.F. Haley and E.S. Shaw dissertation fellowship. All errors are mine.

†Einaudi Institute for Economics and Finance, andrea.pozzi@eief.it

1 Introduction

The introduction and spread of e-commerce has transformed retail markets, changing the way consumers shop for goods and services. For instance, the Internet facilitates price comparison between sellers (Brynjolfsson and Smith, 2000; Brown and Goolsbee, 2002) and offers a larger variety of products than traditional businesses (Brynjolfsson, Yu, and Smith, 2003). Another fundamental difference between online and regular shopping is that Internet customers shop without traveling to a store and receive their purchases delivered at home, instead of picking and carrying them themselves. Delivery to door is not a new service, nor is it exclusively associated with electronic commerce. Items ordered through catalogues are typically delivered at home and some brick-and-mortar stores offer delivery services. Nonetheless, Internet shopping has greatly increased the number of people who have access to this service and the range of goods for which it is supplied.

Home delivery is of particular significance for regularly purchased goods. In this context, customers have the incentive to stockpile for future consumption to take advantage of nonlinear prices in quantity or of temporary price discounts. They face a trade-off between the benefit of stockpiling and the cost of doing so. Though cost is typically represented as a storage cost, it includes other dimensions that Chintagunta, Chu, and Cebollada (2012) label as *physical costs*. These refer to the discomfort associated with carrying the grocery basket around the store and transporting it home. Consumers sustain physical costs every time they shop but we may think they are higher the larger the number of items bought in the trip or the bigger their size and weight. Providing mass access to home delivery, electronic commerce eliminates the physical costs of shopping and removes a substantial hurdle to households' stockpiling decisions.

This paper documents the effect of the introduction of e-commerce on households' propensity to buy in bulk and to stockpile on promotional items. Consistent with what a simple theory model would predict, the elimination of physical costs makes stockpiling and buying in bulk more popular. If the sign of the effect was anticipated, its magnitude is surprisingly large. A back-of-the-envelope calculation of the savings consumers can achieve once these hurdles to stockpiling are removed suggests that the savings represented by the increase in bulk purchases for soda and laundry detergent are worth around one percent of the total expenditure in those product categories. The increased efficiency in stockpiling on discounted items allows for even more significant expenditure reductions, close to 4 percent of the yearly household groceries budget. Though home delivery has received far less attention than other much heralded advantages offered by shopping online, it plays an important role in making e-commerce a superior stockpiling technology that delivers non-negligible benefits to users.

The setting for this study is the grocery industry. I use two years of scanner data on grocery purchases made by a large panel of households at a major US supermarket chain that sells groceries in brick-and-mortar stores and online. Although e-commerce in grocery items is still a niche market in the US, it is steadily growing and has already achieved a wide reach in other countries.[1] Grocery is a classic environment in which to study stockpiling behavior.[2]: many grocery goods are storable for periods ranging from a few weeks (e.g. yogurt) to several months (e.g. soda) or even years (e.g. canned food and detergents). In addition, supermarkets often price packaged goods nonlinearly and subject them to frequent temporary price reductions. Finally, quantifying the savings that consumers can achieve through stockpiling in grocery it is important because they account for a sizeable share of consumption for American households. According to the 2010 Consumer Expenditure Survey, food at home, housekeeping and personal care items account for over 10 percent of consumers' total expenditure.

The estimation strategy used in this paper exploits the panel structure of the

[1] For instance, Tesco online delivery service covers the U.K. in almost its entirety.

[2] See Hendel and Nevo (2004) for a survey of studies providing evidence of stockpiling behavior using supermarket scanner data.

data. I observe the same household shopping on some occasions in supermarket stores, on others ordering online. Within-household variation on the shopping channel provides a source of identification cleaner than the cross-sectional variation used in most studies investigating the effects of e-commerce. In fact, it reduces the concerns about self-selection into online shopping, which may be driving the results if we were to compare the behavior of a sample of online shoppers to that of another sample of traditional customers. Other features of the setting lend strength to the empirical strategy. First, the retailer is committed to offering the same prices and promotions online and in brick-and-mortar stores. Heterogeneous pricing policies across channels would have been a confounding factor for the analysis, as expectations over prices are a major driver of the decision to stockpile. Second, the information comes from a single supermarket chain, ensuring that the seller reputation, the brand name and the assortment are the same across environments.

Even though I am able to compare one same household when shopping in different channels, the possibility that households sort their trips still provides an identification challenge. For instance, if customers systematically shop on the Internet when they have independently decided to build up stocks, a positive association between stockpiling and online shopping would not necessarily imply that e-commerce increases the propensity to stockpile or buy in bulk. I deal with this issue by exploiting the staggered rollout of the online service by the supermarket chain. In fact, the option of shopping on the web became available at different times to households living in different zipcodes. Internet shopping is eventually introduced in all the zipcodes included in the data and the timing of the rollout of the service is mainly dictated by supply side considerations; therefore it is orthogonal to households' shopping decisions. This ensures that comparing customers' purchasing patterns before and after the introduction of the service delivers an estimate of the causal effect of e-commerce on the tendency to stockpile.

I find that when the chain introduces online grocery shopping in a market, the households living there increase their share of expenditure in shopping instances involving large bulks of goods. The share-of-wallet of purchases involving at least 300 ounces of liquid laundry detergent goes up by 15 percent and that of trips including 24 or more cans of soda rises by 90 percent. The average quantity purchased of an item conditional on being on sale is also positively affected by the availability of e-commerce. The share of expenditure on discounted items rises by a figure between 10 and 20 percent. This latter result can be linked to an effect on stockpiling behavior only if it is driven by an increase in the quantity bought conditional on buying on discount. To assess whether this is the case, I contrast the effect of e-commerce on the extensive margin of promotional purchases (i.e. the probability of buying a good when it is in promotion) and its intensive margin (i.e. the quantity purchased, conditional on buying on sale). Consistently with my claims, I find that online shopping does not increase the likelihood that customers buy items when they are on promotion but has a positive impact on the amount they buy when they do so.

That e-commerce can foster bulk purchases and stockpiling of items on sales has important implications for consumers, since Griffith, Leibtag, Leicester, and Nevo (2009) show that those two are major sources of savings for households purchasing groceries. My estimates provide a chance to quantify the dollar impact of lowering the cost of stockpiling for households. As mentioned, the extra savings account for a small but far from insignificant fraction of the overall grocery consumption of the average household.

This paper provides two novel insights. First, I extend the literature on inventory models delivering empirical evidence on the impact of the cost of stockpiling on shopping behavior. Past contributions examined the effect of price changes (Neslin, Henderson, and Quelch, 1985; Boizot, Robin, and Visser, 2001; Hendel and Nevo, 2006b) and inventory held (Neslin and Schneider Stone, 1996; Hendel and Nevo, 2006b) on the decision to stockpile. Efforts to quantitatively explore the impact of variation in the cost of stockpiling have been plagued by the lack of direct data. I expand the definition of stockpiling cost beyond the storage cost considered by traditional inventory models (Erdem, Imai, and Keane, 2003; Hendel and Nevo, 2006a) to include the physical discomfort from carrying the goods purchased. In my application, the introduction of e-commerce provides exogenous variation in the physical cost of stockpiling that does not impact on storage cost, therefore allowing to shed light on the importance of this determinant.

Second, I further the understanding of the effect of information technology on consumers' behavior. Prior literature has studied the impact of online shopping on price sensitivity (Chu, Chintagunta, and Cebollada, 2008; Ellison and Ellison, 2009), brand loyalty (Danaher, Wilson, and Davis, 2003; Pozzi, 2012), and product choice (Brynjolfsson, Hu, and Simester, 2011; Goldfarb, McDevitt, Samila, and Silverman, 2012; Zentner, Smith, and Kaya, forthcoming). I study how the availability of this technology affects inventory decisions. Characterizing the Internet as a superior technology for stockpiling and quantifying the monetary gains for shoppers, I highlight another benefit that online commerce can deliver to consumers in addition to the well known gains derived from lowered search costs and increased provision of variety. To the best of my knowledge, this is the first paper to focus on the home delivery aspect of Internet commerce and to study its implications for the incentive to stockpile.

The rest of the paper is structured as follows. Section 2 describes the data used in the analysis; Section 3 presents the identification strategy and empirical specification of the econometric model. Section 4 comments on the results of the empirical analysis and Section 5 computes the consumers' savings implied by the estimates. Section 6 concludes.

2 Data description

The analysis is based on scanner data on purchases at a large US supermarket chain that sells both online and in regular stores. The sample is a panel of 11,646 households who shopped for grocery *both* at stores of the chain *and* using its web service between

June 2004 and June 2006. Purchases by all members of a household are linked using an identification number that the retailer associates to all the loyalty cards belonging to members of the same family.[3] For each shopping trip, data record the household identification number, the date, the complete list of items bought as identified by the Universal Product Classification (UPC) code or barcode, the quantity purchased, the price paid for each good and the depth of promotional price cuts. Importantly, it also records whether the purchase took place at a brick-and-mortar supermarket or on the Internet. For a random subset of the household I have information on the Census block group where the shoppers live so that I can integrate the data with demographic information from the Census 2000.

Online customers shop through virtual aisles where products are listed with their price; goods on sale are marked with an icon similar to that showing discounted items on the shelf in regular stores. A picture of the item and nutrition factors are also available, when applicable. The retailer commits to offering the same prices online and in-store and even promotions are not channel specific. Online orders are met using goods in stocks in brick-and-mortar stores rather than from dedicated warehouses. As a consequence, the variety of goods offered and the probability of stockout for any given UPC are similar for online and in-store purchases.

All Internet orders are delivered home in a day and time slot chosen by the shopper. The grocer only accepts online purchases worth $50 or more and charges a delivery fee on top of the cost of grocery. The full delivery fee in the period analyzed was $9.95 but promotional coupons were frequently issued for discounted or free delivery; customers ordering for more than $150 and $250 also qualify for, respectively, discounted or free delivery.

In total I observe 1,492,166 shopping trips; 9 percent of them are Internet purchases. Over the years attrition is minimal: 90 percent of the household in the sample

[3]Purchases made without using the loyalty card are not part of the data. However, the penetration rate of the card is high and customers use it regularly as it is the only way to take advantage of promotions.

are seen still shopping in the last month of data and 95 percent of them shop for the last time in one of the last two months in the sample. Customers shops for groceries on average twice per week. Online orders are less frequent than visits to stores: the average household purchases on the web once every five weeks. However, Internet orders are larger than brick-and-mortar ones both in value ($163 versus $47) and in the number of items in the shopping basket (60 items online versus 16 items for regular trips). The absence of the cost represented by carrying heavy baskets for online orders and the fixed cost of delivery explain why Internet orders tend to serve as stock-up trips. level

3 E-commerce as a stockpiling technology

The availability of e-commerce provides customers with a way to reduce the cost of stockpiling. Specifically, home delivery reduces the physical cost of shopping, which is particularly salient for baskets that include many items or bulky and heavy goods. As a result, we would expect the introduction of this technology to have a non negative impact on a household's choice to stockpile.[4] In this section, I present my empirical approach to measuring the impact of the introduction of online shopping on a household's propensity to buy in bulk and to stockpile, and I discuss the main challenges to identification. level

3.1 Identification

Identifying the effect of e-commerce on consumer behavior poses several challenges. First, one has to deal with the problem of self-selection. If we compare an online shopper with a traditional customer, their behavior will most likely be different. However, this may be due to unobserved heterogeneity, potentially correlated with

[4]This mechanism is quite easy to formalize. It can be shown that introducing a physical cost of carrying around the basket as a component of the stockpiling cost in a simple inventory setup as in Perrone (2009) delivers the prediction that the optimal quantity purchased is nondecreasing in the physical cost component.

the decision to be an online shopper. In this case, the comparison would yield a biased estimate of the causal effect of online shopping. I address this concern by exploiting a unique feature of the dataset: I observe each household in the sample shopping both on the web and in the stores. It is therefore possible to estimate the impact of online commerce by comparing the behavior of any given consumer when shopping in the two different environments.

The other major threat to identification is reverse causality. A positive association between stockpiling and the decision of shopping online could be driven by the fact that customers choose to go online when they want to stock up. In such a case, a simple regression of quantity purchased on the shopping channel chosen would overestimate the influence of e-commerce on stockpiling. In order to tackle this problem, I take advantage of the staggered rollout of the web service across zipcodes and use the availability of online shopping as a source of variation in the choice of the channel which is orthogonal to consumers' stockpiling decisions.

The retailer added the option to shop online in 2002. The service was not made available at once in all the markets where the chain was present but introduced progressively. Figure 1 shows the cumulative number of zipcodes where the retailer delivered grocery online. The service was expanding throughout the time window covered by my sample. New zipcodes were included each month, with big expansions in May and September. Internet shopping is eventually offered in all of the zipcodes included in the data. However, around 40 percent of them receive the rollout within the time window covered by my sample.

I treat the introduction of online shopping in a zipcode as a source of variation in access to the online channel orthogonal to households' decision about stockpiling and use it to elicit the causal effect of e-commerce on inventory behavior. The assumption of exogeneity of the rollout seems plausible in this context. Although the retailer may choose which market to enter with the online service based on expectations of market characteristics which are also correlated with shopping behavior (wealth, age, population density), such conduct would not affect my estimates. In fact, the online service is eventually introduced in *all* the zipcodes included in the data. Therefore, strategic behavior on the part of the retailer in the rollout decision is of concern for identification purposes only if the timing of the rollout in different markets is correlated with such characteristics.

I address this potential problem in two ways. First, whenever possible I include household fixed effects in the empirical specification. This ensures that identification of the effect of e-commerce is free of any correlation with invariant characteristics of the population of consumers. Second, I argue that the timing of the rollout was mostly influenced by logistical factors, rather than by expectations of consumers' reactions. In fact, the web service is organized so that deliveries for online orders leave from each of the brick-and-mortar stores, rather than from a centralized warehouse. This makes the service easier to implement in some areas (e.g. because of proximity to main roads) and creates the incentive to expand the service in neighboring zipcodes, which can be served by the same fleet of trucks. Anecdotal evidence is provided by the order of rollout: the chain introduced the service in a number of zipcodes close to its headquarters long before jumping to obviously more attractive markets such as Washington D.C. and Los Angeles.

Table 1 provides some empirical support for my identification assumption, displaying average characteristics of zipcodes belonging to three waves of e-commerce rollout. The three waves have been identified using the two sizeable jumps, in July 2002 and July 2005, in the number of zipcodes reached from the service which are clearly discernible in Figure 1. The first wave includes the zipcodes where the service became available before August 2002; the second wave contains zipcodes reached by online grocery between August 2002 and July 2005. All the zipcodes where the service was implemented after July 2005 are in the third wave.

In general, the average characteristics of the three subsamples are quite similar. Although some of the differences between waves are statistically significant, it is hard to recognize a pattern. For instance, zipcodes in the second wave display a lower share of Hispanic population than the early rollout ones. However, the latest zipcode to be reached by the service are characterized by a heavier Hispanic presence. Similarly, the population reached in the second wave has a larger share of middle aged individuals. This may suggest that the chain decided to rollout first in areas with younger and more tech savvy populations. However, the population of the later rollout wave is younger than that of the areas where the service was introduced first. Finally, I do not detect any statistical or economically meaningful difference in the median family income of the three waves. This is quite telling, since it is hard to imagine a strategy discriminating the timing of introduction of the service that would not factor in wealth.

The data have other convenient features that help in controlling for potential confounding factors. First, I observe purchases for a panel of households which allows me to use fixed effects to control for time-invariant unobserved heterogeneity across households. This is particularly relevant when studying stockpiling behavior. In fact, one of the main determinants of stockpiling decisions is the storage cost, which is generally unobserved to the econometrician and for which it is hard to find reliable proxies. Using only within-household variation, I can ensure that unobserved heterogeneity in storage cost has no role in determining the results, under the assumption that the cost of storing inventory does not change across time for the same family. Furthermore, the retailer decided to offer the same prices and promotions over the two channels. Therefore, any difference in behavior between the online and the traditional shopping environment cannot be driven by differences in pricing policies. Finally, since both the website and the stores are operated by the same chain under the same name, we can assume that customers do not perceive any change in the brand image of the retailer in the two contexts.

3.2 Econometric specification

I look at the effect of online shopping on two outcomes: the decisions to buy in bulk and stockpile during sales.

The identification of bulk purchases requires defining a threshold above which we can say that the household is buying a "large" number of units. I call bulk purchases those involving quantities equal to or larger than that contained in the largest stock-keeping unit (SKU) within a given product category. SKU sizes were manually coded using the short product description provided by the chain for each UPC in stock. I focus on two product categories, carbonated soft drinks sold in cans and liquid laundry detergent, and restrict attention to the two most sold brands in the data in each of the two categories. For the analysis of stockpiling behavior, coding SKU size is not necessary and I do not have to focus on particular categories or brands. I identify sales using the pricing information provided by the retailer. The data include a gross and a net price for each item sold. When the latter is lower than the former, the item was sold at a promotional price. I use these data to estimate specifications of the following type

$$share\_Bulk_{ht} = a_0 + a_1 Online\_Available_{ht} + a_2 X_{ht} + a_3 W_h + T_t + e_{1,ht} \tag{1}$$

$$share\_Sale_{ht} = b_0 + b_1 Online\_Available_{ht} + b_2 X_{ht} + b_3 W_h + T_t + e_{2,ht} \tag{2}$$

with h indexing the household and t the month. $share\_Bulk_{ht}$ is the share of expenditure in bulks of laundry detergent (or soda) out of the total expenditure in the category in a month. Equation 1 is estimated separately for laundry detergent and soda. $share\_Sale_{ht}$ is the fraction of expenditure in items on sale out of the total expenditure in grocery in a month.

The explanatory variable of interest is $Online\_Available_{ht}$, an indicator taking value 1 if the web service is available in the zipcode of residence of household h in month t. Since the rollout of the service was gradual, there are several zipcodes where shopping online becomes possible during the time window covered in the data. The coefficient a_1 (b_1) picks up the difference in the average share of grocery expenditure in bulky items (promotional items) before and after the option of shopping online was introduced. Equations 1 and 2 can be thought of as a "reduced form" of an instrumental variable model where the dependent variables are regressed over the channel in which the trip took place, instrumented with the availability of the online commerce service.

X_{ht} includes household characteristics varying through time (e.g. the number of trips in a particular month) and W_h is a matrix of time-invariant characteristics. In alternative specifications, time-invariant unobserved heterogeneity across households is picked up using household fixed effects. T_t represents a full set of time dummies.

4 Results

4.1 Buying in bulk

I study the effect of the introduction of e-commerce on the propensity to buy in bulk, looking at two particular categories: soda and liquid laundry detergent. In both categories, I focus on the two main brands based on market share. Coke and Pepsi capture 57% of the soda market in my data; whereas Tide and All make up 47% of sales of laundry detergent. These categories are staples of household consumption and can be stored for long periods. Furthermore, both soft drinks and detergent can be bought in a range of different sizes. For soda, I restrict attention to cans which can be bought as single item or in packs of six, eight, twelve, eighteen or twenty-four units. In the data, liquid laundry detergent is sold in four main bottle sizes: 50 ounces, 100 ounces, 200 ounces, and 300 ounces. I say that a household is purchasing laundry detergent in bulk when it is buying 300 ounces or more. For soda, a bulk purchase implies picking at least 24 cans. I include in the definition of large trips both instances where the household picks the largest size available and where it buys that amount in combinations of any size denomination (e.g. two twelve-cans packs).[5]

Figure 2 displays market shares for each package size in the two categories, by channel of purchase. The main fact emerging from the picture is that customers tend to buy larger packs when shopping online. The share of 300 ounce and 24 can purchases is larger for Internet orders than for store transactions. At the same time, single cans and small detergent bottles are more popular choices when customers visit brick-and-mortar stores. Figure 3 shows how the distribution of purchases across sizes adjusts after the introduction of online shopping. Once again, we witness a shift towards larger packages in markets where e-commerce is available. Internet purchases contribute little to sales of small sizes and overwhelmingly so to those of medium and large SKUs. The picture also shows that customers are sorting their trips: not only do they buy more in bulk online but they also shift some of their bulk purchases from stores to the web.

The evidence in the figures is consistent with online shopping fostering bulk purchases. However, it pools together purchases from different households without controlling for differences in their characteristics and it does not deal with the endogeneity in the choice of the shopping channel. I can address both issues by estimating equations 1 and 2.

Table 2 reports, separately for soda and laundry detergent, the estimates of equation 1. The dependent variable is the share of monthly expenditure in *bulky trips* for liquid laundry detergent and soda, where a *bulky trip* implies the purchase of 300 ounces or more (for laundry detergent) and 24 or more cans (for soda).[6] In columns

[5] This is a conservative choice. Although buying a 24-can pack or two 12-can packs amounts to transporting the same bulk, this does not necessarily imply the same level of discomfort from carrying them around. For example, in the former case one has to lift and carry the whole weight of the bulk at once; whereas in the latter case the household can break the transport into two trips, carrying a more maneagable 12-can pack each time. If that is the case, including sums of smaller size denominations to count as bulky trips would lead to measure the dependent variable with error. This would not bias the coefficient but would inflate the standard errors, making it hard to detect an effect statistically significant at the conventional level.

[6] Results are robust to lowering the threshold for a bulky trip to purchases of 18 cans of soda. Counting purchases of 200 oz. of soda as bulky trips leads to a loss of significance in some of the specifications but does not affect the qualitative results.

1 and 5 of Panel A, I control for the total number of trips made by the household and some household characteristics matched at the block group level from Census 2000: ethnicity, a dummy indicating whether the household is a family, age, income, income squared and distance from the closest store of the chain. I also include month fixed effects in the specification. For both laundry detergent and soda availability of e-commerce is associated with an increase in the weight of bulky trips in the shopping basket of the consumers and the effect is economically significant. When it becomes possible to order online, the share of expenditure in bulky items goes up 30 percent for laundry detergent and almost 80 percent for soda.

Including household fixed effects (columns 2 and 6), I identify the effect of e-commerce relying solely on within-household variation. This is important to the interpretation that links the results to the reduction in the cost of carrying heavy items allowed by home delivery. In fact the household fixed effects control, among other things, for heterogeneity in storage cost. Ability to store is an obvious determinant of the decision to buy for deferred consumption and may be correlated with factors that determine propensity to buy online. Though half of the effect for laundry detergent seems to be due to unobserved heterogeneity, introduction of e-commerce is still associated with a sizeable increase in the wallet share of bulk items. The coefficient for the soda category is only marginally changed. As a robustness check, in columns 3 and 7 I also examine the effect of e-commerce availability on the amount spent in bulk purchases of detergent or soda, and I find even bigger effects. Finally, in columns 4 and 8, I use the same set of explanatory variables to estimate a Poisson model where the dependent variable is the number of bulky trips. The effect of online shopping is statistically significant and large.

We may worry that the positive correlation between online commerce and buying in bulk is driven by other factors than the reduction in transportation cost. In particular, online orders must be worth at least $50 of grocery and customers must wait for the delivery to enjoy the goods purchased. This implies that trips undertaken to purchase for immediate consumption can only take place in-store. This type of event is driven by a different motive than the typical grocery: almost by definition the customer enters the store with no intention to stockpile. Furthermore, Internet orders have to be worth more, potentially pushing customers to buy bulky sizes (which have higher total price). Panel B of Table 2 addresses this concern by restricting the analysis to trips worth at least $50 so to make online and traditional trips more comparable. Moreover, the analysis for soda also excludes trips featuring purchase of single-serving soda. This eliminates from the sample what are most likely purchases for immediate consumption; this concern is obviously less present for laundry detergent. Even in this subsample, I still find that the introduction of online commerce fosters expenditure on bulk purchases.

Figure 4 also lends support to the idea that the introduction of e-commerce has a causal impact on shopping behavior. The plot displays the incidence of expenditure in bulky trips as a function of the time to and since the rollout of online shopping for laundry detergent (Figure 4a) and soda (Figure 4b). I focus on a window covering the six months before and after the service is rolled out and normalize to zero the average for the earlier period so that the dots represent the effect compared to that baseline. I detect no trend in the importance of bulk purchases in the period up to the moment of rollout; the incidence of bulk shopping is not significantly different from the baseline until Internet shopping becomes available. After this event, however, it experiences a sustained increase.

4.2 Stockpiling

I now proceed to analyze the effect of online shopping on the decision to stockpile on items that are in promotion. In this case, I look at the universe of goods sold at the supermarket and compute the share of expenditure in UPCs on sale out of the total grocery expenditure of the household in a month. The data provide information on the exact depth of the price cut, allowing me to investigate whether the effect is heterogeneous depending on the size of the discount.[7,8]

Table 3 reports the results of an OLS regression of the share of expenditure on items in promotion over an indicator for availability of online shopping, time dummies and a set of controls. Each panel of the Table reports results for a different definition of the dependent variable, with increasingly higher requirements for a price reduction to be deemed a sale. In Panel A, I include discounts of any size; under this definition the average UPC in the sample has a 43 percent probability of being on sale in a given week. In Panel B and C I only consider as sales price reductions of respectively 50 and 75 percent of the gross price. I experiment controlling with demographic characteristics of the household (column 1) and with household fixed effects (column 2) and find the same qualitative result: the introduction of e-commerce leads to an increased prominence of discounted items in the consumers' basket. Under the less restrictive definition of sales the share of expenditure in promotional items jumps by about 10 percent once the household can start shopping online. One would expect larger effects when focusing only on large discounts (taking 50% or even 75% off the normal price), since deeper price cuts constitute a bigger incentive to stock up. Indeed the estimated increase for 50-percent discounted items is 20 percent and that for 75-percent one is around 15 percent. Of course such promotions are far less frequent. The average UPC is sold at half the price only 5 percent of the time and 75-percent discounts only occur with .001 probability.

[7] Only the Retailer's discounts are systematically recorded in the data. Inclusion of discounts from manufacturer's coupons is more irregularly documented. However, the overwhelming majority of price discounts are issued by the retailer and tied to use of the chain's loyalty card. Paper coupons issued by the supermarket on their weekly flyer advertising prices are almost irrelevant. Importantly, paper coupons cannot be redeemed online. Therefore, if I were underestimating their importance I would be biasing the results against finding a positive effect of e-commerce on the relevance of promotions.

[8] The availability of a gross and a net price is convenient for the identification of sales and of the size of the discount. If we only observed the price posted by the chain, we would have to make assumptions to construct an hypothetical "reference price", that is the price we would expect if there were no promotions, in order to understand whether there is a sale and what its size is. Instead, the gross price provides this information directly as it represents the price that the retailer would charge in the absence of promotional discounts. The availability of such information in the data derives from the fact that only owners of the supermarket loyalty card are entitled to the discounts. All other customers will pay the sticker price.

In Figure 5 I plot the share of expenditure on promotional items in the six months leading to the rollout of Internet shopping and in the six month after that event. The data have been standardized so that each dot represents the difference from the average seven or more months before the service is introduced. Once again the pattern is consistent with a causal interpretation of the correlation between e-commerce and stockpiling on promotional items. For every definition of sale, the average expenditure share becomes significantly different from the baseline only from the month of rollout on.

That households spend a larger fraction of their grocery expenditure on items on sale once they can order online does not necessarily mean that they are stockpiling more. If goods were more often on sale on the Internet channel or if the frequency of sales increased once the online service is introduced, we could observe the same pattern. I can rule this out since both the website and the stores are operated by the same chain, which sets identical prices across channels. Furthermore, the pricing strategy of the chain did not change after the introduction of e-commerce as online shoppers are not a large enough fraction of the overall customer population to justify such adjustment. Even keeping promotions constant, the surge in the share of promotional items in the basket could be due to an increase in the number of times customers buy an item on promotion rather than to more stockpiling (i.e. a rise in the number of units bought on each instance). To address this concern, I look separately into the extensive and the intensive margin.

I start by assessing whether customers are more likely to buy an item in promotion when they can purchase on the supermarket website. On the one hand, collecting price information on the website should be easier and allow shoppers to more readily spot promotions. On the other hand, the label indicating discounts may not be as salient on the web. Furthermore, online customers can use a function that allows them to shop from the list of items purchased in the past; which makes it less likely that they would become aware of promotions for brands or package sizes other than those they usually pick. I estimate the following linear probability model where the dependent variable is an indicator function that takes value one if the purchased item was bought on promotion in a particular supermarket trip for a given household.

$$S_{uht} = a_u + a_1 Online\_Available_{ht} + a_2 X_h + e_{uht} \tag{3}$$

$$\tag{4}$$

where u indexes the product (UPC), h the household and t denotes a particular trip. The main regressor of interest is the indicator variable *Online_Available* that switches to one for all the products of a household living in an area reached by the Internet service. This specification can be seen as the reduced form of a model where I use availability of the Internet service to instrument for the household 's decision to shop online.

I select the 1,000 most sold products in the data. After excluding UPC's linked to gasoline and those whose product description is too generic to identify the product category, the final sample consists of nearly 900 UPC's belonging to a 151 different product categories. I include UPC fixed effects so that identification comes from variation in the frequency of purchases on sale for a same item, and control for time effects by adding a full set of month dummies in each specification. Finally, I include either demographic characteristics of the purchasing household or household fixed effects to account for customers' heterogeneity in taste for sales.

The results are reported in Table 4. For most specifications and definitions of a sale, the probability that an item is purchased on promotion turns out not to be significantly affected by the introduction of the online service. However, the fixed effect specification in Panel A and the baseline model in Panel B show a significant jump in the probability of purchasing an item on sale after the introduction of e-commerce. Nevertheless, these results do not seem to provide credible grounding for interpretations of the effect on stockpiling alternative to the one I provided. In fact, when I check for differential effects on a subset of goods which are clearly storable and, therefore, more likely to be subject to stockpiling purchases[9], I find consistently that such products are less likely to be bought on sale after the rollout of the website.

While the data do not support the contention that introducing online shopping makes customers more likely to buy on sale, I can provide evidence that the quantity purchased conditional on buying on sale goes up. I do so by regressing the number of units purchased for a given UPC conditional on it being purchased in promotion on the usual set of regressors and UPC and time fixed effects. Table 5 shows that e-commerce availability hardly affects the average quantity bought on sale for the average good. However, once I condition on storable goods I identify an 8 percent increase in quantity purchased for products purchased in discounts of any size. As usual, the result is robust to restricting the definition of a sale only to large or extreme price cuts. In fact, the effect monotonically increases in the size of the discount, as we would expect. When I condition on purchasing the good at half its price, quantity purchased increases by 12 percent post e-commerce introduction; for 75% discount the estimated increase is nearly 55 percent but the effect is not statistically significant, most likely due to the low number of such large discounts. The estimated impact is even stronger in specifications where the demographic controls are replaced by household fixed effects.

5 Calculation of the savings

In this section, I perform some back-of-the-envelope calculations to quantify in monetary terms the savings that online commerce can generate by reducing the cost of stockpiling and by buying in bulk.

[9]The selected goods belong to the following categories: soda, bottled water, canned food and vegetables, bottled household supplies (laundry detergent, bleach, etc.), disposable diapers.

5.1 Saving from buying in bulk

I start by computing savings obtained through bulk purchases. I have shown that online commerce fosters the purchase of soda and laundry detergent in larger sizes. This allows the consumer to save money since big SKU's are sold at a higher total cost but at a lower price per unit. In the data, the average price per ounce of liquid laundry detergent in a 300 oz. bottle is 11.5 percent lower than for 200 oz. bottles and 15 percent lower than for 100 oz. bottles. Likewise, the average price of can of Coke or Pepsi sold in a 24-can package is 7 percent lower than the average price of a can in a 12-can package. The average price of a can sold as a single unit is almost five times as high.

I use the estimates from the specifications in columns 2 and 6 of Table 2, as well as information on the total amount spent in consumption of soda and laundry detergent by size of the SKU, to estimate the share of expenditure in bulk sizes due to the introduction of the online service.[10,11] I then divide this "extra expenditure" in bulk items by the average price of the 300 oz. bottle (for laundry detergent) or the 24-cans package (for soda). This gives me the number of units (liquid ounces or cans) which would not have been bought in bulk had online shopping not been available. Finally,

[10]The estimates used in the exercise are obtained using a different definition of bulky trip than the one used to obtain the results reported in columns 2 and 6 of Table 2. In the latter case, I considered a trip to be bulky if the total amount of laundry detergent or soda bought in the trip exceeded a certain threshold, *no matter the denomination of the sizes bought*. For the purpose of this exercise I re-estimate the model restricting the definition of bulky trips to those involving purchases of large amounts of the good *in the largest possible denomination*. In fact, whereas two trips involving a 24 cans pack or two 12 cans packs of soda may be equally bulky, only in the former case will the household take advantage of the nonlinear pricing. Therefore, what matters in order to compute savings is the increase in the share of expenditure in big SKU's rather than the overall bulkiness of the trip. Due to the rare occurrence of trips where the customer buys large amounts of laundry detergent of soda in small denominations, the results are virtually identical to those presented in Table 2.

[11]Given the effect of e-commerce on the share of expenditure in bulk (a_1) and the total expenditure in bulk ($Bulk_{it}$), the counterfactual expenditure in bulky items in a world without online shopping can be computed as $\frac{Bulk_{it}}{1+a_1}$. This approximation disregards the fact that not all customers in the sample had access to e-commerce throughout the sample period and, therefore, only their post-adoption expenditure should be used for the calculation. Furthermore, I assume that the introduction of online shopping would not have any effect on the total amount spent overall in laundry detergent and soda. The same caveats apply to the similar exercise I perform for sale purchases.

I compute the savings by multiplying this quantity by the difference in the average price per unit between non bulk and bulk SKUs.

Panel A in Table 6 reports the results as the percentage of the total expenditure of a household in the category. Total savings depend on the conjectured size of the SKU for the unit, had they not been bought in bulk. The Table provides figures for the two most popular non-bulk sizes as terms of comparison. Thanks to the improved ability to buy in bulk provided by online shopping, the average customer saves between one and one and a half percent of its yearly expenditure in laundry detergent. Savings for soda are somewhat smaller in percentage terms, though larger in number of dollars, if we assume that removing the Internet would have led consumers to buy 12-can packages rather than 24-can ones. Savings are much larger under the assumption that all this consumption would have occurred through the purchase of single cans. This is driven by the wide difference in unit price between the largest and the smallest SKU in soda.

5.2 Saving from stockpiling on sales

The computation of savings from stockpiling in bulk follows a similar procedure. Using the time series on prices over two years for the same 889 goods selected for the estimation of equation 3, I compute the average promotional discount associated with the category. Milk substitutes, pet food and granulated sugar are product categories with shallow price cuts: during promotion, price only drops 10 percent. Potato chips, fruit snack, and facial tissue are characterized by high discounts, in the range of 25 to 30 percent.

The estimate in column 1 (Panel A) of Table 3 is used to calculate, for each household and each product category, the share of expenditure in promotional items induced by the easing of the physical cost of stockpiling introduced by the Internet. The depth of the price cuts in the category allows the estimation of the size of the savings, that is how much higher that expenditure would have been if the items had been bought at full price. Total savings for a household are computed by aggregating savings across all the categories.

Panel B in Table 6 displays the results. Unlike for the case of bulk purchases, where I was focusing on two specific categories, here I am looking at the whole set of grocery goods purchased by a household. Therefore, I report savings as percentages of the total household expenditure on grocery. The extra stockpiling triggered by the availability of e-commerce induces sizeable savings. The average household saves 3.7 percent of its annual grocery budget thanks to it. I also report the fraction of this figure due to some classic storable product categories (soda, bottled water, canned food and fruit, and diapers). These alone generate a reduction worth half a percentage point on overall grocery expenditure.

5.3 Interpretation of the results

To understand the importance of the savings allowed by e-commerce, it is useful to compare them with the calculations by Griffith, Leibtag, Leicester, and Nevo (2009). They find that for the average household the savings from buying in bulk and purchasing on sale accrue to 15.6 percent and 6.5 of its annual grocery budget, respectively. These figures are significantly larger than my estimates. However, they consider savings due to the *overall* share of bulky and promotional items in a consumer's basket; whereas my figures only refer to the lower expenditure linked to the *additional* share that such items gain thanks to the possibility of having home delivery. Furthermore, I exploit a sample of online shoppers, who tend to be wealthier than the average household and may be less bent on looking for ways to save money. Given the narrow margin under consideration, the effect is substantial in monetary terms.

The estimated savings should be taken with caution. They are constructed using the average effect of the introduction of e-commerce to calculate the increase in expenditure in bulk and discounted items. This would deliver the correct figure only if the effect is not heterogeneous across households with different propensities to buy in bulk and on sale. Also, some of the savings could be double counted. For example a promotional discount on bulky items would boost at the same time the savings from buying in bulk and those from buying during promotions. Finally, I observe grocery purchases at a single grocer and have no information on shopping trips households in my sample make at competing supermarket stores. The impact of online shopping could be overestimated if some of the extra consumption in bulk and promotional items were not incremental but substituting unobserved purchases at other stores. In the present setting, this is less likely to occur as I am dealing with a sample of online shoppers, who are likely to be more loyal to the chain and to shop there for a large fraction of their grocery needs.

The savings are also gross of some monetary costs associated with use of the online channel. First of all, customers pay a delivery fee when shopping online. However, the retailer frequently issues coupons for free or discounted delivery. In addition, trips involving stockpiling are likely to be large and such trips automatically get a fee reduction or a waiver. At any rate, the average household in the sample spends less than half of one percentage point of its total grocery expenditure on delivery fees. Therefore, the estimated savings would be positive even net of such charges. Other costs that may have an effect on the bottom line are not taken into account. For example, some of the evidence I presented is consistent with customers being worse at spotting promotional deals when shopping online; which could translate into foregone savings. However, the magnitude of this effect cannot be computed without estimating a fully fledged demand model to simulate the counterfactual brand choice of the household, had it not shopped on the web.

At the same time, the monetary savings triggered by the availability of home delivery capture only part of the benefits delivered by online shopping to stockpiling households. As I have shown, not only do customers buy more in bulk and stockpile more when shopping online but they also tend to sort stockpiling trips into the online channel. By revealed preferences, this suggests that they are better off shopping for bulky and heavy baskets once they can do so on the Internet. Once again, to measure the increase in welfare caused in this dimension by the introduction of e-commerce would require a more structural approach.

6 Conclusions

In this paper, I illustrated how online commerce, which is usually paired with home delivery of the goods purchased, can have an important effect on consumers' behavior in markets for frequently purchased goods. In particular, a reduction in the cost of carrying around the purchased items fosters shopping in bulk and stockpiling on items in promotion, which are major sources of savings in retail.

I exploit an ideal setup for identifying the causal effect of shopping online. In fact, I have data on a panel of consumers shopping both online and in traditional stores at a large supermarket chain. This reduces concerns about confounding factors such as unobserved heterogeneity across households, which I can control for using fixed effects, and differences in pricing and reputation between the online and the brick-and-mortar channel. Moreover, the staggered rollout of the online shopping service provides quasi-experimental variation in the use of the online distribution channel which further strengthen the identification.

I find that the share of bulky and promotional items in the average household's basket increases significantly once it becomes possible to shop online in a given market. This finding is in line with the prediction of a simple inventory model where both the storage cost and the physical cost of carrying around heavy baskets act as hurdles to stockpiling.

Since items sold in bulk and on sale feature a lower unit price, the possibility of buying them more easily allows households to save money. I provide an estimate of the monetary savings achieved by customers thanks to online shopping. Additional bulk purchases in laundry detergent and soda allow them to save around one percent of the total expenditure in those categories. Increased stockpiling on items on discount triggers even more significant savings, equal to almost 4 percent of the total yearly expenditure on grocery of the average household in the sample.

Besides shedding light on an understudied characteristic of Internet commerce, this study opens up a few avenues for further research. First, the findings of this paper should apply beyond grocery, to any market characterized by regularly repeated purchases. Investigating the relationship between stockpiling and online shopping in other settings would provide some external validity to my findings and illustrate whether this effect depends on the characteristics of the good. Second, my results have implications for the supply side too. In particular, online retailers could devise strategies for pricing the online service so that they appropriate the monetary gains which are currently pocketed by consumers.

References

Boizot, C., J.-M. Robin, and M. Visser (2001): "The Demand for Food Products: An Analysis of Interpurchase Times and Purchased Quantities," *Economic Journal*, 111, 391–419.

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., Y. J. Hu, and D. Simester (2011): "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on Concentration of Product Sales," *Management Science*, 55(11), 1755–1765.

Brynjolfsson, E., and M. D. Smith (2000): "Frictionless Commerce? A Comparison of Internet and Conventional Retailers.," *Management Science*, 46(4), 563–.

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.

Cameron, C. A., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference with Multiway Clustering," *Journal of Business and Economic Statistics*, 29(2), 238–249.

Chintagunta, P., J. Chu, and J. Cebollada (2012): "Quantifying Trasnsaction Costs in Online/Off-line Grocery Channel Choice," *Marketing Science*, 31(1), 96–114.

Chu, J., P. Chintagunta, and J. Cebollada (2008): "A Comparison of Within-Household Price Sensitivity across Online and Offline Channels," *Marketing Science*, 27, 283–299.

Danaher, P. J., I. W. Wilson, and R. A. Davis (2003): "A Comparison of Online and Offline Consumer Brand Loyalty.," *Marketing Science*, 22(4), 461–476.

Ellison, G., and S. Ellison (2009): "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77, 427–452.

Erdem, T., S. Imai, and M. Keane (2003): "Price and Promotion Expectations: Capturing Consumer Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5–64.

Goldfarb, A., R. C. McDevitt, S. Samila, and B. Silverman (2012): "The Effect of Social Interaction on Economic Transactions: An Embarrassment of Niches," working paper.

Griffith, R., E. Leibtag, A. Leicester, and A. Nevo (2009): "Consumer Shopping Behavior: How Much Do Consumers Save?," *Journal of Economic Perspectives*, 23(2), 99–120.

Hendel, I., and A. Nevo (2004): "Intertemporal Substitution and Storable Products," *Journal of the European Economic Association*, 2(2-3), 536–547.

——— (2006a): "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica*, 74(6), 1637–1673.

——— (2006b): "Sales and Consumer Inventory," *The RAND Journal of Economics*, 37(3), 543–561.

Neslin, S. A., C. Henderson, and J. Quelch (1985): "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4, 147–165.

Neslin, S. A., and L. Schneider Stone (1996): "Consumer Inventory Sensitivity and the Postpromotion Dip," *Marketing Letters*, 7, 77–94.

Perrone, H. (2009): "Inventories, Unobservable Heterogeneity and Long Run Price Elasticities," working paper.

Pozzi, A. (2012): "Shopping Cost and Brand Exploration in Online Grocery," *American Economic Journal: Microeconomics*, 4(3), 96–120.

Zentner, A., M. D. Smith, and C. Kaya (forthcoming): "Bricks, Clicks, Blockbusters, and Long Tails: How Video Rental Patterns Change as Consumers Move Online," *Management Science*.

Tables and Figures

Figure 1: Rollout of the online service

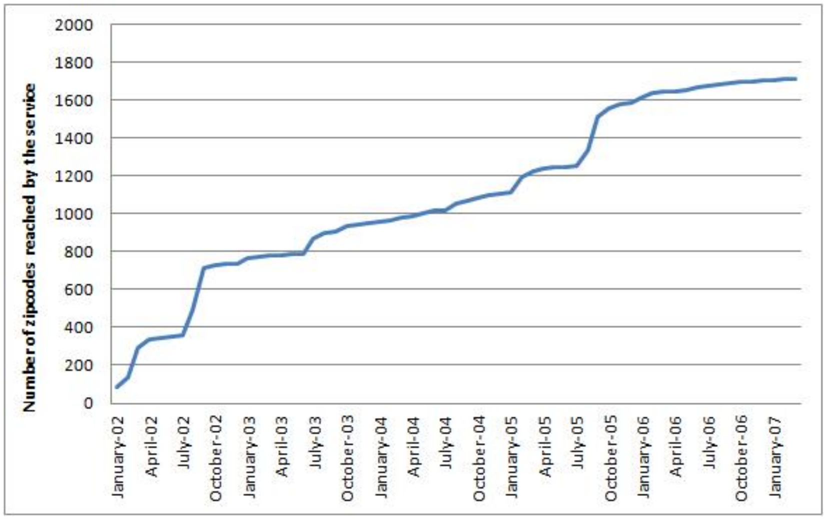
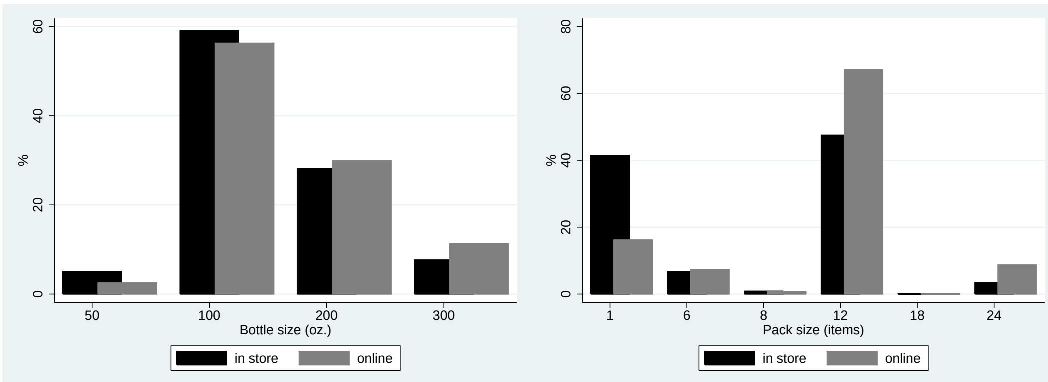


Figure 2: Market share of different package sizes, by channel of purchase

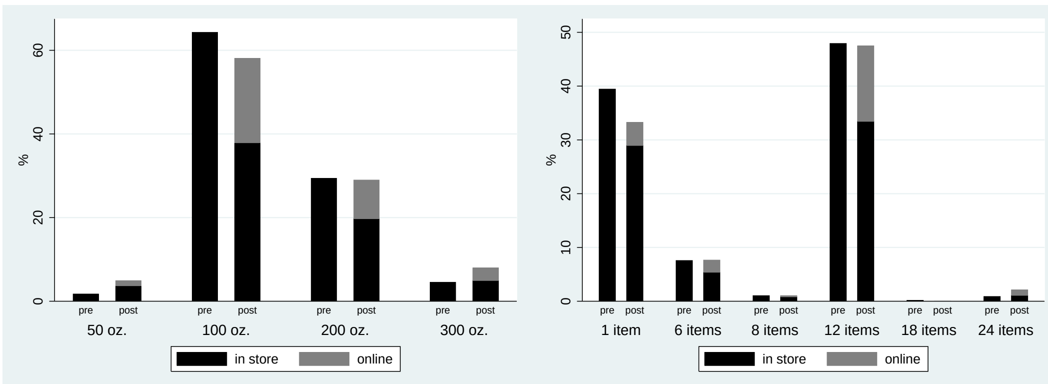
(a) Laundry detergent (b) Soda



Notes: Market shares are computed as the fraction of purchases of a particular package size over total purchases in the product category.

Figure 3: Market share of different package sizes before and after introduction of the online shopping channel, by channel of purchase

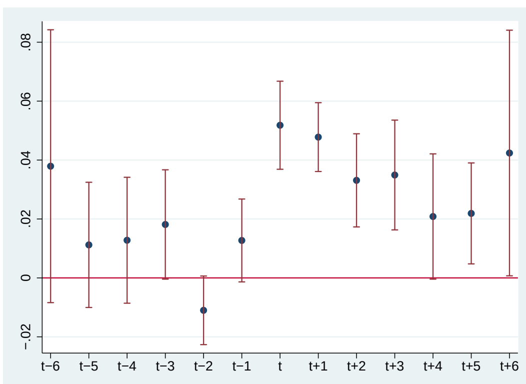
(a) Laundry detergent (b) Soda



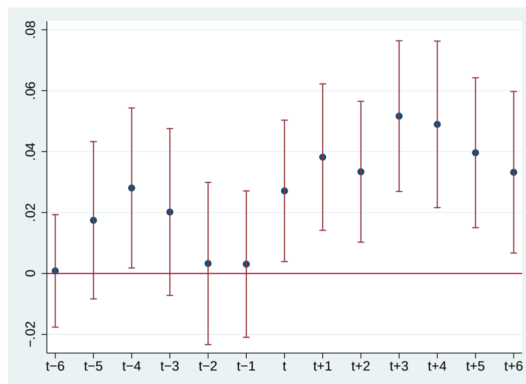
Notes: Market shares are computed as the fraction of purchases of a particular package size over total purchases in the product category. Market share "pre" is computed considering only trips which occurred when online shopping had not yet been introduced by the retailer in the zipcode of residence of the household. Market share "post" include only trips that took place after the introduction of the service. The sample only includes households living in zipcodes where the online service was rolled out between June 2004 and June 2006.

Figure 4: Expenditure share on bulk sizes as a function of time to/since introduction of e-commerce

(a) Laundry detergent



(b) Soda



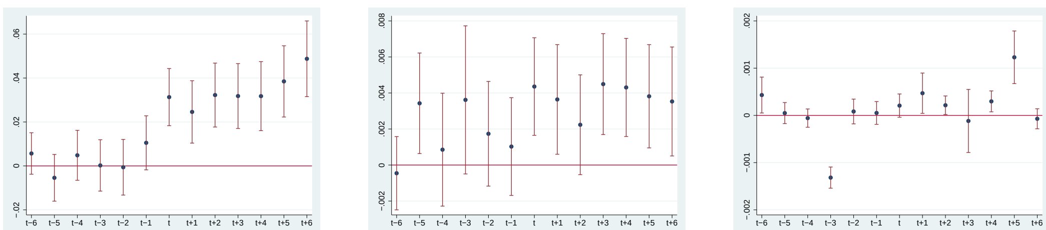
Notes: The dots represent the estimates of the β coefficients for laundry (figure (a)) and soda (figure (b)) purchases from the following model

$$share\_Bulk_{ht} = \alpha_h + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{t = R + k\} + T_t + e_{ht}$$

where R indicates the time of rollout of electronic commerce in the zipcode of residence of household h. Data for months later than the sixth after the introduction of the service are not used so that the coefficients represent differences from the average expenditure in bulk seven or more months before the introduction of e-commerce. The whiskers denote the 90 percent confidence interval. Robust standard errors are clustered at the household level. The sample only includes households living in zipcodes where the online service was rolled out between June 2004 and June 2006.

Figure 5: Expenditure share on items on sale as a function of time to/since introduction of e-commerce

(a) Sale=Discount>0 (b) Sale=Discount>50% (c) Sale=Discount>75%



Notes: The dots represent the estimates of the β coefficients for different definition of "sale" from the following model

$$share\_Sale_{ht} = \alpha_h + \sum_{k=-6}^{6} \beta_k \mathbb{I}\{t = R + k\} + T_t + e_{ht}$$

where R indicates the time of rollout of electronic commerce in the zipcode of residence of household h. Data for months later than the sixth after the introduction of the service are not used so that the coefficients represent differences from the average expenditure in items on promotion seven or more months before the introduction of e-commerce. The whiskers denote the 90 percent confidence interval. Robust standard errors are clustered at the household level. The sample only includes households living in zipcodes where the online service was rolled out between June 2004 and June 2006.

Table 1: Characteristics of zipcodes belonging to different waves of rollout of the e-commerce service

	First wave	Second wave	Difference First vs. Second	Third wave	Difference First vs. Third
Hispanic	15.1	17.6	-2.5**	9.2	6.9***
	(13.11)	(17.81)		(12.66)	
Black	4.9	5.4	-0.97	11.1	-6.3***
	(7.64)	(9.58)		(19.63)	
Age 18 to 24	8.3	8.6	-0.82	8.3	0.003
	(4.53)	(5.95)		(7.36)	
Age 25 to 34	16.1	15.1	2.7**	15.1	2.09**
	(5.94)	(5.85)		(8.09)	
Age 35 to 44	17.8	17.5	1.5	17.5	1
	(2.84)	(3.60)		(4.24)	
Age 45 to 54	14.7	14	3.1***	15	-1.4
	(3.19)	(3.54)		(4.23)	
Age 55 to 64	8.7	8.5	0.74	9.4	-3.2***
	(2.54)	(3.12)		(3.82)	
Age 65 or older	11.6	11.3	0.54	11.8	-0.41
	(6.25)	(9.71)		(7.41)	
Family	65.9	69.7	-3.9***	66.5	-0.49
	(15.71)	(15.07)		(18.4)	
College degree	45.8	39.9	5.7***	47.3	-1.2
	(16.34)	(16.32)		(20.21)	
Employed	63.6	62.6	1.8*	64.7	-1.9*
	(7.51)	(10.27)		(10.10)	
Median household income	64,645	63,750	0.07	64,582	0.04
	(22,638)	(22,599)		(27,519)	
Observations	462	556		431	

Notes: The table reports averages of demographic characteristics for zipcodes belonging to three different waves of rollout of the Internet service. The variables are zipcode level statistics on shares from Census 2000. The waves are identified by sizeable jumps in the number of zipcodes reached by online grocery. The first wave includes zipcodes where the service was implemented before July 2002; the second wave groups those where e-commerce become available between August 2002 and July 2005; all the other zipcodes belong to the third wave. The columns titled "Difference" report the t-statistic and the significance of a test for the difference in means between the first and second wave and the first and third wave, respectively. ***: Significant at 1% **: Significant at 5% *: Significant at 10%

Table 2: Effect of e-commerce availability on the weight of bulk purchases in consumers' shopping baskets

Panel A: All trips

	Laundry detergent				Soda			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. variable	*Expenditure share*		*Total expenditure*	*Number of trips*	*Expenditure share*		*Total expenditure*	*Number trips*
Mean (St.dev.)	*0.096 (0.29)*		*1.85 (5.90)*	*0.104 (0.32)*	*0.084 (0.25)*		*1.48 (6.09)*	*0.166 (0.55)*
Online available	0.029***	0.014*	0.621***	0.492***	0.067***	0.075***	1.21***	1.43***
	(0.010)	(0.009)	(0.202)	(0.166)	(0.005)	(0.005)	(0.107)	(0.151)
Demographic controls	Yes	No	No	Yes	Yes	No	No	Yes
Household f.e.	No	Yes	Yes	No	No	Yes	Yes	No
Observations	17,604	25,457	25,457	17,604	47,869	70,707	70,707	47,869

Panel B: Only trips worth at least $50 and no single-serving soda

	Laundry detergent				Soda			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. variable	*Expenditure share*		*Total expenditure*	*Number of trips*	*Expenditure share*		*Total expenditure*	*Number trips*
Mean (St.dev.)	*0.142 (0.34)*		*2.99 (7.99)*	*0.152 (0.38)*	*0.523 (0.47)*		*9.39 (16.31)*	*0.829 (0.96)*
Online available	0.048***	0.037**	1.09***	0.485***	0.050***	0.024**	2.18***	.107**
	(0.014)	(0.019)	(0.320)	(0.146)	(0.018)	(0.009)	(0.582)	(0.045)
Demographic controls	Yes	No	No	Yes	Yes	No	No	Yes
Household f.e.	No	Yes	Yes	No	No	Yes	Yes	No
Observations	15,312	22,105	22,105	15,312	39,982	59,025	59,025	39,982

Notes: The unit of observation is a household-month pair. The specifications including demographic controls use only the random subsample of households for which the Retailer provided information on the census block group of residence. Demographic controls include ethnicity, family status, age, income, income squared (all matched from Census 2000) and distance from the retailer's closest store (information provided by the retailer). The panel is unbalanced since not all households have positive expenditure in soda or laundry in every month in the data. Month fixed effects are included in all columns. Robust standard errors are in parenthesis and are clustered at the household level. ***: Significant at 1% **: Significant at 5% *: Significant at 10%

Table 3: Effect of e-commerce availability on the weight of items on sale in consumers' shopping baskets

Panel A: Sale = Discount>0
Share of monthly expenditure on promotional items: Mean=.40; st.dev.=.16
Total monthly expenditure on promotional items: Mean=149; st.dev.=123

	Dependent variable			
	(1) expenditure share	**(2)** expenditure share	**(3)** expenditure share	**(4)** total expenditure
Online available	0.042***	0.038***	0.037***	30.8***
	(0.004)	(0.004)	(0.002)	(3.28)

Panel B: Sale = Discount>50%
Share of monthly expenditure on promotional items: Mean=.02; st.dev.=.03
Total monthly expenditure on promotional items: Mean=6.94; st.dev.=10.51

	Dependent variable			
	(1) expenditure share	**(2)** expenditure share	**(3)** expenditure share	**(4)** total expenditure
Online available	0.004***	0.004***	0.001***	1.84***
	(0.001)	(0.001)	(0.000)	(0.238)

Panel C: Sale = Discount>75%
Share of monthly expenditure on promotional items: Mean=.001; st.dev.=.002
Total monthly expenditure on promotional items: Mean=0.05; st.dev.=0.42

	Dependent variable			
	(1) expenditure share	**(2)** expenditure share	**(3)** expenditure share	**(4)** total expenditure
Online available	0.0002*	0.0001*	0.0002*	0.010*
	(0.000)	(0.001)	(0.000)	(0.005)
Observations	132,842	119,281	208,560	132,842
Household f.e.	No	No	Yes	No
Demographic controls	Yes	Yes	No	Yes
Sample restrictions		Trips worth over $50		

Notes: The unit of observation is a household-month pair. The specifications including demographic controls use only the random subsample of households for which the Retailer provided information on the census block group of residence. Demographic controls include ethnicity, family status, age, income, income squared (all matched from Census 2000) and distance from the retailer's closest store (information provided by the retailer). The panel is unbalanced since some of the households do not shop at the chain in every month in the data. Month fixed effects are included in all columns. Robust standard errors are in parenthesis and are clustered at the household level. ***: Significant at 1% **: Significant at 5% *: Significant at 10%

Table 4: Effect of e-commerce availability on the probability of buying on sale

Panel A: Sale = Discount>0

Dep. variable: Dummy for purchase on sale. Mean=0.43; st.dev.=0.50

	(1)	(2)	(3)
Online available	-0.001	0.020***	0.021***
	(0.002)	(0.003)	(0.003)
Available*storable			-0.023***
			(0.003)

Panel B: Sale = Discount≥50%

Dep. variable: Dummy for purchase on sale. Mean=0.05; st.dev.=0.22

	(1)	(2)	(3)
Online available	0.012***	-0.000	0.001
	(0.001)	(0.001)	(0.001)
Available*storable			-0.018***
			(0.003)

Panel C: Sale = Discount≥75%

Dep. variable: Dummy for purchase on sale. Mean=0.001; st.dev.=0.03

	(1)	(2)	(3)
Online available	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Available*storable			-0.001*
			(0.000)
Observations	5,528,989	5,528,989	5,528,989
Product f.e.	Yes	Yes	Yes
Time f.e.	Yes	Yes	Yes
Household f.e.	No	Yes	Yes
Demographic controls	Yes	No	No

Notes: An observation is a household-UPC-trip triplet. Specifications featuring demographic controls include as regressors ethnicity, family status, age, income, income squared (all matched from Census 2000) and distance from the retailer's closest store (information provided by the retailer) as well as month fixed effects. Robust standard errors are in parenthesis and are clustered at the household-trip level using the procedure proposed in Cameron, Gelbach, and Miller (2011). ***: Significant at 1% **: Significant at 5% *: Significant at 10%

Table 5: Effect of e-commerce availability on quantity purchased on sale

	Sale = Discount>0		**Sale = Discount>0**		**Sale = Discount>0**	
	Dependent variable: Quantity (number of units) purchased conditional on buying on sale.					
	Mean=1.40; st.dev.=1.11		*Mean=1.35; st.dev.=1.11*		*Mean=1.56; st.dev.=1.66*	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Online available	-0.009**	0.005	-0.070***	-0.072***	0.089	n.a.
	(0.005)	(0.003)	(0.025)	(0.031)	(0.060)	
Available*storable	.108***	.404***	0.229***	0.737***	0.871	n.a.
	(0.033)	(0.042)	(0.079)	(0.101)	(0.780)	
Observations	2,399,022	2,399,022	281,743	281,743	5,097	5,097
Household f.e.	No	Yes	No	Yes	No	Yes
Demographic controls	Yes	No	Yes	No	Yes	No

Notes: An observation is a household-UPC-trip triplet. Observations are used conditional on the UPC being on promotion, which explains the large variation in the number of observations as the definition of a sale becomes more restrictive. Specifications featuring demographic controls include as regressors ethnicity, family status, age, income, income squared (all matched from Census 2000) and distance from the retailer closest store (information provided by the retailer) as well as month fixed effects. Robust standard errors are in parenthesis and are clustered at the household-trip level using the procedure proposed in Cameron, Gelbach, and Miller (2011). ***: Significant at 1% **: Significant at 5% *: Significant at 10%

Table 6: Savings from the introduction of e-commerce

	Mean	St.dev.	10th pctile	50th pctile	90th pctile
Panel A: Savings through bulk purchases					
Laundry detergent					
Moving from 200 oz to 300 oz. bottles	1.01%	.52	.3%	1%	1.7%
Moving from 100 oz to 300 oz. bottles	1.5%	.73	.4%	1.4%	2.4%
Soda					
Moving from 12-can to 24-can packs	0.67%	.62	.1%	.5%	1.5%
Moving from single cans to 24-can packs	76%	70	13.2%	52.1%	174.6%
Panel B: Savings through stockpiling on sales					
All categories	3.7%	1.16	2.4%	3.6%	5.3%
Subset of storable categories	.43%	.42	.1%	.3%	.8%

Notes: The dollar savings are reported as a fraction of the total expenditure in the product category in Panel A and as a fraction of the overall expenditure in grocery in Panel B.