



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Quantifying Transaction Costs in Online/Off-line Grocery Channel Choice

Pradeep K. Chintagunta, Junhong Chu, Javier Cebollada,

To cite this article:

Pradeep K. Chintagunta, Junhong Chu, Javier Cebollada, (2012) Quantifying Transaction Costs in Online/Off-line Grocery Channel Choice. Marketing Science 31(1):96-114. <https://doi.org/10.1287/mksc.1110.0678>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Quantifying Transaction Costs in Online/Off-line Grocery Channel Choice

Pradeep K. Chintagunta

Booth School of Business, University of Chicago, Chicago, Illinois 60637, pradeep.chintagunta@chicagobooth.edu

Junhong Chu

NUS Business School, National University of Singapore, Singapore 117592, bizcj@nus.edu.sg

Javier Cebollada

Public University of Navarra, 31006 Pamplona, Spain, cebollada@unavarra.es

Households incur transaction costs when choosing among off-line stores for grocery purchases. They may incur additional transaction costs when buying groceries online versus off-line. We integrate the various transaction costs into a channel choice framework and empirically quantify the relative transaction costs when households choose between the online and off-line channels of the same grocery chain. The key challenges in quantifying these costs are (i) the complexity of channel choice decision and (ii) that several of the costs depend on the items a household expects to buy in the store, and unobserved factors that influence channel choice also likely influence the items purchased. We use the unique features of our empirical context to address the first issue and the plausibly exogenous approach in a hierarchical Bayesian framework to account for the endogeneity of the channel choice drivers. We find that transaction costs for grocery shopping can be sizable and play an important role in the choice between online and off-line channels. We provide monetary metrics for several types of transaction costs, such as travel time and transportation costs, in-store shopping time, item-picking costs, basket-carrying costs, quality inspection costs, and inconvenience costs. We find considerable household heterogeneity in these costs and characterize their distributions. We discuss the implications of our findings for the retailer's channel strategy.

Key words: channel choice; online grocery shopping; transaction costs; plausibly exogenous; hierarchical Bayesian; green shopping

History: Received: November 23, 2010; accepted: July 28, 2011; Preyas Desai served as the editor-in-chief and Bart Bronnenberg served as associate editor for this article. Published online in *Articles in Advance* December 20, 2011.

1. Introduction

Researchers have identified a number of transaction costs such as opportunity costs of time and transportation costs as possible drivers of a consumer's choice of a *physical grocery* store (Bell et al. 1998, Pashigian et al. 2003, Fox et al. 2004, Briesch et al. 2009). In an online setting, researchers have found delivery charges and retailer reputation to be important factors influencing *online* store choices for *non-grocery* items (Smith and Brynjolfsson 2001). When consumers choose between online and off-line *grocery* channels, there could be additional transaction costs that are specific to this purchase setting.

Our objectives in this paper are (1) to identify the transaction costs from the previous literature as well as any additional transaction costs that could influence the choice of online and off-line grocery channels, (2) to incorporate these costs into a model of consumers' choice between online and off-line grocery channels, (3) to quantify and provide money metrics for these costs, and (4) to investigate the implications

of transaction costs for a retailer's marketing strategy. In addition, we explore the "green" aspect of online grocery shopping by quantifying possible societal benefits of such shopping. As with the previous literature (Bell et al. 1998), we focus on store choice conditional on a store visit. We use "store choice" and "channel choice" interchangeably to mean choice of an off-line or online store. Furthermore, given our focus on channel choice conditional on making a store visit, our money metrics represent the relative transaction costs of shopping online versus shopping off-line.

Online and off-line channels provide varying levels of distribution services that entail different levels of direct costs and transaction costs on consumers. Direct costs refer to the sum of shelf prices (less discounts, or net of discounts) of the items in the shopping basket; transaction costs are the costs needed to transfer market goods in a store into consumption goods at home. Transaction costs vary from trip to trip and differ across households. They play

an important role in household choice of shopping venues. However, many transaction costs are non-monetary and hard to measure and quantify, and thus it is difficult for retailers to factor them in when designing marketing strategies. Because understanding consumers' store and channel choice decisions is of fundamental importance to retailers, it is of considerable value to managers to understand the monetary implications of the various transaction costs.

The empirical context for quantifying transaction costs in channel choice is a unique household panel data set from Spain. The data consist of the same households choosing between a chain's off-line stores and its online store over six months. Using data from a single chain may seem limiting because it may not capture all the purchases of a panelist. This concern, although legitimate, is mitigated in our case, as the panelists' annual in-chain expenditures are about 80% of grocery expenditures by households of similar sociodemographics in the same area. Meanwhile, focusing on purchases within a chain has several advantages. First, issues such as store image as choice drivers are no longer relevant. Second, the chain has the same prices and promotions online and off-line. Thus, direct costs are identical across channels and will not affect channel choice. Third, similar assortments are available everywhere, so this factor will not play a role either. Our empirical context thus allows us to focus squarely on the role of transaction costs in driving channel choice. Conventional reasons for why people shop online, such as lower prices, sales tax avoidance, and so on, are ruled out directly and therefore do not confound our analysis.

Nevertheless, empirically quantifying transaction costs is a challenging task because several of them, e.g., costs of putting items into the shopping cart and carrying them home, depend on the specific categories and their quantities that a household expects to buy on that shopping trip. These factors need not be exogenous because unobserved factors that influence the choice of shopping channels could also influence the categories and quantities bought. For example, if a household member goes out purchasing other goods for consumption (e.g., clothes), he or she may decide to combine the trip with a visit to a physical grocery store. This unobserved factor that influences off-line purchase could also influence the categories purchased (e.g., if the household has limited time to spend in the store). Hence, the potential endogeneity of the channel choice drivers needs to be accounted for. One way to deal with endogeneity is to specify a full system of equations that characterize channel choice, categories purchased, and associated quantities (akin to Briesch et al. 2009, who model store

choice and category needs).¹ A formal treatment of all these factors simultaneously poses serious methodological and computational issues. A key feature of our data is that there is very little variation in quantity choices across channels within a household—decomposing the variance in quantity choices within a category into the variance explained by household fixed effects and channel fixed effects, we find that the former explain 87.0%–99.8% of the variance. This allows us to abstract away from considering purchase quantities in the analysis.

Next, accounting for each of the categories that a household expects to buy on a store visit is nontrivial because there is considerable variation across households in the categories that drive their basket expenditures. Thus, we cannot fix the set of categories under analysis as being common across households as in previous studies, nor can we focus on a smaller subset of categories because to account for at least 80% of a household's basket expenditures we need to include 30 categories. As our interest in the categories bought stems from how they influence transaction costs and as the choice of the channel does not affect quantities bought, the identities of the *specific* categories bought is not critical. Thus, we can simply summarize the information contained in the expected categories purchased via metrics that will likely influence transaction costs. Specifically, we focus on the total number of items, the number of perishable items, and the number of heavy/bulky items (defined in §3.4) that a household expects to purchase. One benefit of this classification is that transaction costs can be specified as functions of these numbers. Thus, instead of looking at channel choice and the expected purchase in each category, our endogenous variables are channel choice and the total number of items, the number of perishable items, and the number of heavy/bulky items that the household expects to buy.

Given the above set of potentially endogenous variables and given that channel choice depends on the various expected numbers of items purchased, one can adopt either a full information approach that specifies how each of these variables is determined by a consumer or a limited information approach where numbers of items are instrumented for in the analysis of channel choice. We choose the latter approach because it not only obviates the need to specify the exact data generating process for these endogenous variables but also avoids the problem of incorrect transaction cost estimates in the presence of a misspecified data generating process. We use as instruments (i) each household's inventory variables based on its own top expenditure categories and (ii) the

¹ Because online and off-line channels have similar assortments, *brand choice* is unlikely correlated with channel choice.

numbers of items bought by pure online and pure off-line households of similar demographics to the panelists (see §5 for a justification of these instruments). We combine these instruments with the Conley et al. (2010) plausibly exogenous approach in a hierarchical Bayesian (HB) framework to account for the endogeneity of the numbers of items purchased. A key advantage of this approach is that it allows us to work with potentially weak instruments. Finally, to provide money metrics for the various transaction costs, we exploit the fact that although direct costs do not play a role in channel choice, the presence of delivery costs allows us to compute the marginal utility of income. Using the posterior distributions of the parameter estimates, we quantify the distributions of various transaction costs across households.

The main contributions of our study are as follows: (1) We integrate the transaction costs identified in the literature and the additional transaction costs specific to the online and off-line grocery setting into a model of channel choice. (2) Methodologically, we apply the Conley et al. (2010) plausibly exogenous approach to a nonlinear and hierarchical setting. (3) We provide a strategy for reducing the complexity of the channel choice analysis by aggregating across categories in a manner that helps to simplify the problem substantially without losing relevant information. (4) Substantively, we quantify and provide money metrics for the important types of transaction costs in grocery shopping that involves online and off-line channels, which can help retailers to better design marketing strategies for the two channels. We also take some preliminary steps to explore the implications of online shopping for the environment in terms of reducing carbon emissions.

2. Transaction Costs and Online and Off-line Channel Choice

Transaction costs economics (Coase 1937; Williamson 1979, 1981; Williamson and Masten 1999) emphasizes the role of transaction costs in economic exchanges. The basic principle of transaction costs economics is that agents choose to conduct transactions in a way that minimizes their transaction costs. Grocery shopping is a canonical context that involves transactions between retailers and consumers. The basic economic function of retailers is to provide consumers with explicitly priced goods and a set of distribution services, including assortment, assurance of product delivery, information, accessibility of location, and ambiance (Betancourt and Gautschi 1986, 1988). These goods and distribution services generate customer value and are essential inputs into household production (Becker 1965). To carry out home production, consumers need to incur direct costs and a set

of transaction costs that map (not necessarily one-to-one) onto the set of goods and distribution services (Furubotn and Richter 1997). Stores differ in the goods and distribution services provided, and hence they differ in their direct and transaction costs to consumers. For the same level of home production, consumers choose stores with the lowest sum of direct and transaction costs for their shopping basket.

Transaction costs play a role in all stages of a consumer's store choice process. Researchers have applied the notion of transaction costs, in particular, the role of time costs and transportation costs, to analyze consumer's choice of conventional stores (Bell et al. 1998), online stores (Smith and Brynjolfsson 2001, Lee and Png 2004), and between online/direct retailers and conventional retailers (Liang and Huang 1998, Balasubramanian 1998, Keeny 1999, Sinai and Waldfogel 2004, Forman et al. 2009). Researchers have identified several types of transaction costs in grocery shopping in *physical* stores. Betancourt (2005) synthesizes these costs into categories that map into various distribution services: (1) *Opportunity costs of time* comprise travel time to and from a store and in-store shopping and waiting time. (2) *Transportation costs* to and from a store include travel time and costs, and are related to accessibility of store location. (3) *Psychic costs* are costs such as perceived difficulty of use, inconvenience, frustration, annoyance, anxiety, drudgery, dissatisfaction, disappointment, personal hassles, shopping enjoyment, or disagreeable social interactions that consumers are subject to in the store environment (Ingene 1984, Devaraj et al. 2002). (4) *Adjustment costs* as a result of the unavailability of products at the desired time or in the desired amount are costs that arise from additional time and transportation costs incurred because of forced search or lower utility associated with altering the consumption bundle of goods. (5) *Search costs* are the costs of time, transport, and other resources in the acquisition of information with respect to price, assortment, physical attributes, or performance characteristics of the goods provided in different stores. In looking at online shopping for nongrocery items, Smith and Brynjolfsson (2001) identified two additional costs as important transaction costs for choosing online stores: (6) *delivery costs* and (7) *waiting costs* for basket delivery.

When it comes to the choice between online and off-line *grocery* stores, there could be additional transaction costs as well. Liang and Huang (1998) and Chircu and Mahajan (2006) mention these types of costs in their analysis, and the popular press has discussed them in some detail. One such cost that has been identified is (8) *physical costs*. Based on the experience of a Webvan customer, Chircu and Mahajan (2006, p. 909) note, "He lived on the 4th floor in a building with no

elevator, and the delivery person carried bulky grocery purchases up the stairs for him.” Some online blogs (e.g., Nick 2007) note the following when they compare costs across off-line and online channels—“You have to push a heavy cart” versus “Internet shopping carts weigh 0.003 grams,” “You have to maneuver through crowded aisles” versus “Internet aisles can fit any number of people,” and “You have to load your groceries in the car, unload them at home, and carry them into the kitchen” versus “The grocery delivery person does all of this.” Physical costs, therefore, include (a) the costs of picking and putting items into the shopping cart, which differ significantly across online and off-line stores as the Internet option eliminates these costs on consumers; and (b) the costs of carrying the basket home, which are important in markets where many people walk or take public transport to the store.

A second such cost associated with the Internet channel is (9) *quality inspection or product evaluation costs* (BusinessWeekOnline 2001). Liang and Huang (1998), in a nongrocery setting, refer to this as an “examination cost.” The Internet increases the costs of providing a given level of assurance of product delivery at the desired time, or of the desired quality, as a result of the consumer’s inability to inspect and acquire the product at the time and place of purchase. Quality inspection costs may play an important role when the shopping basket contains items such as perishables whose qualities vary from purchase occasion to purchase occasion. Delivery of perishable goods with quality guarantee is considered the biggest logistical problem for Internet grocers (Demby 2000). Internet retailers also take special actions to guarantee the quality of perishables ordered online: “If, for any reason relating to the freshness of a Perishable Good, you are less than 100% satisfied, COOLGREENANDSHADY.COM will arrange for the re-delivery of your order” (Cool Green & Shady 2011).

Direct costs and transaction costs are therefore a function of the shopping channel, basket (product) characteristics, shopping occasion (duration and timing of the trip) characteristics, and household characteristics. Together, they vary across trips for the same household, depending on when, what, and how much it buys. They also differ by households for the same product bundle, depending on household sensitivities to different cost components and valuation of convenience and time. For each shopping occasion, a household trades off these costs and chooses a store with the lowest sum of direct and transaction costs. Consequently, we would expect to see both within- and across-household variation in store choice as a function of these factors. Accordingly, our model of store choice needs to account for both direct and transaction costs.

3. Data and Observed Shopping Patterns

3.1. The Grocery Chain

Our data are from a major grocery retail chain in Spain. The data are for one metro area, where the retailer has about 200 physical stores and 1 online store. The online store is the retailer’s largest store by revenue. It partners with 18 of the chain’s physical stores for grocery supply. The partner stores were selected based on size (they are the largest stores), geographic location (dispersed around the metro area), and service ability such that they individually have the ability to fulfill online orders (e.g., have adequate assortments, personnel, delivery trucks) and collectively can do so efficiently.

When a consumer makes her first online purchase, she needs to register an account with her loyalty card. Her account information is then stored in the computer server and retrieved for subsequent orders. To place an online order, a consumer first goes to the grocer’s website and logs into her account. She then browses or searches the aisles, selects the items she wants to buy, and puts them into her basket. After filling her basket, she proceeds to checkout. The list of the basket items is stored in the computer system as “previous shopping lists” and can be used for later “express” orders. The chain uses a centralized online ordering system. After receiving an online order, the ordering system assigns it to one of the partner stores and notifies the household. The household then has two options: it can go to the assigned partner store to pick up the order for free (such purchases are not observed in our data) or have the basket delivered home within some chosen delivery time window (e.g., 7–9 P.M. Monday) and incur delivery charges.²

The retailer practices uniform pricing so prices are identical across all its online and off-line stores.³ It is a Hi-Lo chain and runs chainwide promotions. The online product offerings are the same for everyone and available in all partner stores. Thus, assortments are identical across the partner stores and the online store. Assortments in nonpartner stores can differ slightly because of store size differences and consequent differentials in stock-out rates. We were able to

² Recently, chains such as Farm Stores in the Miami–Dade area in the United States have started offering online ordering combined with free drive-through pickup services (see <http://www.miamiherald.com/2011/04/10/2159495/farm-stores-offers-online-grocery.html>).

³ We confirmed with the management and checked the data that the retailer indeed practices uniform pricing in all off-line stores and the online store. The retailer used to practice zone pricing in its off-line stores with two price zones, but it stopped off-line zone pricing prior to our data period (see Chu et al. 2008 and 2010 for details).

verify that 98.3% of the online items ordered were also purchased in nonpartner off-line stores in our data. This, in addition to confirmation from management, gave us the assurance that available assortments are similar across stores (see also §5.5).

People often walk or take public transport to buy groceries (only 68% of stores have parking lots). About 60% of off-line stores also offer delivery service. The retailer has the same delivery policies for online and off-line orders. It charges €6 for delivery if the basket is below €100 and €4 if the basket exceeds €100. Delivery is free for golden-card (a premium loyalty card awarded to big spenders) clients if the basket exceeds €100. Households with quarterly in-chain expenditure exceeding €600 are eligible but need to apply for golden-card membership. In sum, online and off-line channels have the same prices, price promotions, and delivery policies, as well as similar assortments.

3.2. Store Price Promotions

The retailer provided us with price promotion data, including categories and items promoted, promotion start and end dates, and depths of price cuts. Promotions occur in vastly differing sets of categories. Each day there are on average 85 categories and 419 items on promotion. Promotion durations range from three to eight weeks. Three-week, four-week, and five-week promotions account for 22.9%, 43.2%, and 33.7% of the promotion cycles, respectively. Such multiple-week promotions are quite different from weekly promotions commonly practiced by U.S. grocers. Promotion depths vary substantially across products, ranging from 4.5% to 25.0% with a mean of 8.7%.

3.3. Scanner Panel Data

We obtain the complete shopping records of 3,556 households between May and November 2007. This is a random sample of the retailer's online customers. The households shop interchangeably in the online and off-line channels (we observe purchases in 196 stores). We observe at what time a household visits the chain, which store it visits, what items and how much of each item it buys, whether the basket is home delivered, and delivery charges. We select a random sample of 1,025 households for model estimation. Table 1 presents major demographics and store characteristics for all households and for the chosen sample. An average household has 3.37 members, 0.68 preschool and 0.47 school-age children, 2.14 work-age adults, and 0.07 elders. Of these households, 15.32%, 26.34%, and 58.34% live, respectively, in low, medium, and high income/economic areas. We also obtained data on pure online and pure off-line customers that we use to construct instruments.

3.4. Observed Shopping Patterns and Implications for Econometric Modeling

The households exhibit the following shopping patterns (see Table 2) that appear to be consistent with the transaction costs of shopping at each channel. These patterns provide some model-free evidence of the importance of transaction costs in channel choice decisions.

First, the availability of the Internet option does not act as a sorting mechanism whereby certain households mainly shop online and others primarily shop off-line. All sample households shop in both channels. There is not the "20/80" phenomenon, where 20% of

Table 1 Household Demographics and Characteristics of Most Frequented Off-line Stores

	Entire sample		Random sample	
	Mean	Std. dev.	Mean	Std. dev.
<i>Family size</i>	3.39	2.20	3.37	2.05
<i>No. of preschool children (0–5)</i>	0.69	0.89	0.68	0.87
<i>No. of school-age children (6–18)</i>	0.44	0.85	0.47	0.89
<i>No. of working adults (19–65)</i>	2.18	1.89	2.14	1.74
<i>No. of elders (65+)</i>	0.07	0.47	0.07	0.35
<i>Distance to most frequented off-line store</i>	1.13	4.01	0.99	2.61
<i>Percentage of reside in low economic area</i>	15.43		15.32	
<i>Percentage of reside in medium economic area</i>	24.96		26.34	
<i>Percentage of reside in high economic area</i>	59.61		58.34	
<i>Percentage of downtown</i>	41.46		39.61	
<i>Characteristics of most frequented off-line stores</i>				
<i>Store square footage (m²)</i>	1,480.61	1,102.38	1,473.63	987.01
<i>Percentage of bazaar section</i>	63.86		64.39	
<i>Percentage of butcher shop</i>	91.68		90.83	
<i>Percentage of processed meat</i>	93.38		92.29	
<i>Percentage of fish shop</i>	83.05		82.05	
<i>Percentage of having parking lots</i>	68.03		68.68	
<i>Percentage of home delivery</i>	60.44		60.68	

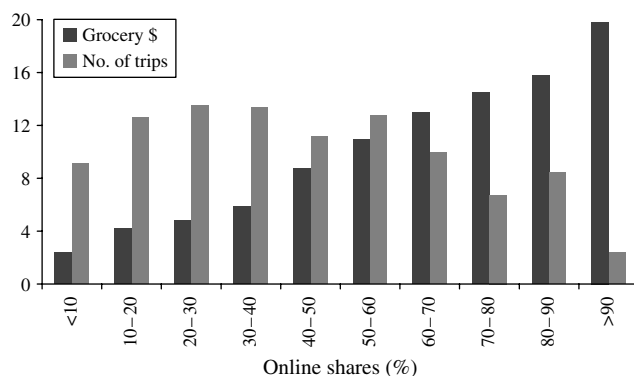
Table 2 Characteristics of Household Shopping Behavior by Channel

	Total		Online		Off-line	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Total trips	16.18	11.67	5.56	4.22	10.61	10.98
Mean trip interval (days)	10.97	12.34	14.46	14.29	9.14	10.75
Basket size (€)	82.99	82.59	155.77	83.48	44.85	50.12
Half-year spending (€)	1,342.60	874.50	866.56	697.73	476.04	534.10
No. of unique categories per trip	16.86	13.11	27.67	11.00	11.19	10.26
No. of unique perishable categories	4.11	4.82	4.87	5.57	3.72	4.32
No. of unique heavy/bulky categories	5.06	4.79	9.89	3.67	2.54	3.06
No. of unique items per trip	22.61	19.18	38.47	17.26	14.30	14.28
No. of unique perishable items	4.69	5.73	5.55	6.60	4.24	5.15
No. of unique heavy/bulky items	6.64	6.86	13.39	5.98	3.10	4.06
Percentage of perishables in trip expenditure	23.21	25.34	11.69	13.14	29.24	27.97
Percentage of heavy/bulky in trip expenditure	33.03	26.14	49.30	17.81	24.50	25.77
Coefficient of variation: Basket	0.49	0.35	0.24	0.16	0.73	0.32
Coefficient of variation: Trip interval	0.73	0.25	0.52	0.25	0.86	0.32
Channel switching: Household level (%)	58.69	33.70	68.92	31.51	48.50	32.72

the households account for 80% of the online trips or expenditures (see Figure 1). The mean ratio of online trips to total trips is 41.8% (std. dev. = 24.7%). The ratio is below 20% for 22% of the households, above 80% for 11% of the households, and between 20% and 80% for 67% of the households. The mean ratio of online expenditures to total expenditures is 65.0% (std. dev. = 24.9%). The ratio is below 20% for 7% of the households, between 20% and 80% for 58% of the households, and above 80% for 35% of the households. Channel “switching” is common in the data. The online-to-off-line switching probability is 56.5%, and the off-line-to-online probability is 29.6%. This suggests that for the same households, transaction costs of buying online versus off-line vary from trip to trip. Delivery charges are involved for 22% of the online trips and 1% of the off-line trips. *Because households visit both channels, the same household’s temporal channel choice decisions can only be explained by trip-level factors, even though demographics may explain the average online shopping intensity across households.*

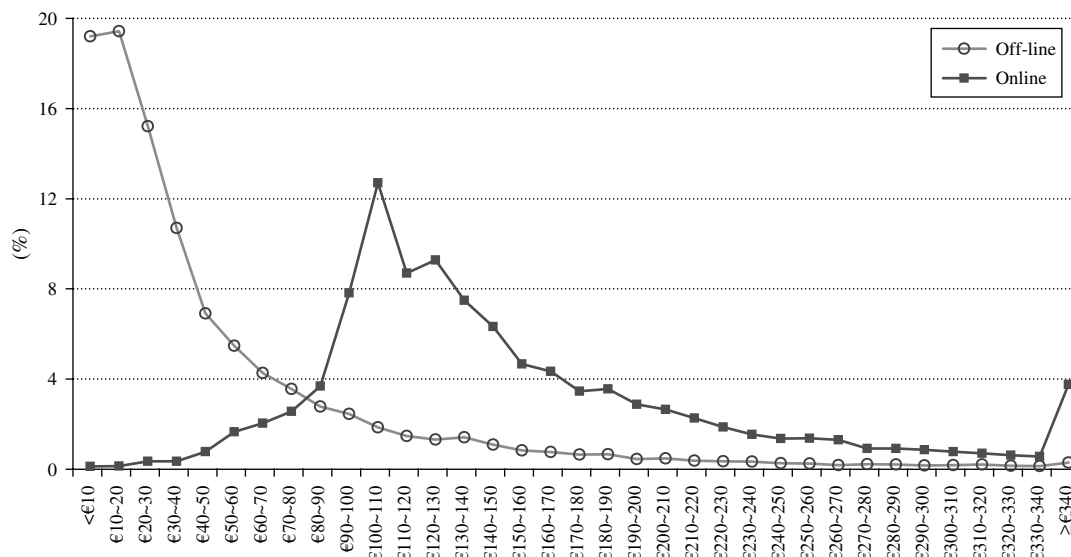
Second, households sort their shopping trips to online and off-line channels based on basket size.

Figure 1 Household Distribution by Shares of Online Grocery Expenditure and Trips



They make major trips to the online channel and fill-in trips to the off-line channel (Kahn and Schmittlein 1989). Online baskets (€155.8) on average are 3.5 times as large as off-line baskets (€44.9). Households buy 28 (11) unique categories and 38 (14) unique items on an online (off-line) trip; and 95% of a household’s online trips have a basket larger than its mean off-line basket, and 63% of its off-line trips have below-average-sized baskets. This suggests that relative to shopping off-line, transaction costs of shopping online are low when the basket is large but high when the basket is small. *It also raises the possibility that there exist unobservable factors that influence both channel choice and number of items purchased.*

Third, households also sort their trips to online and off-line channels based on category characteristics. Online baskets consist of more heavy/bulky items, reflecting the convenience benefit of online shopping, whereas off-line baskets have more perishables, reflecting the ability of the physical channel to allow for quality checking. Perishables refer to fresh produce, meat, seafood, and bakery items in a household’s basket. Heavy/bulky items refer to bottled, canned, or bagged consumer packaged goods (CPGs) (e.g., mineral water, beer, toilet paper) and liquid-rich non-CPGs (e.g., watermelon). Examples of these are 4 one-liter bottles of water, 12 rolls of toilet tissue, etc. Note that some categories can appear in both heavy and perishable sets (e.g., watermelon). An online (off-line) basket has 13.4 (3.1) unique heavy/bulky items and 5.6 (4.2) perishable items. Heavy/bulky (perishable) items account for 49.3% (11.7%) of online and 24.5% (29.2%) of off-line trip expenditures. Furthermore, a household, on average, buys 29.3 categories exclusively online, 32.4 categories exclusively off-line, and 23.2 categories in both channels. The channel-specific category purchases suggest that the transaction costs of buying the same categories differ by

Figure 2 Distribution of Shopping Trips by Basket Size

channel. The transaction costs of buying heavy/bulky items are lower online but higher off-line. The opposite holds for buying perishables. It appears that types of categories play a role and need to be accounted for when modeling household channel choice.

Fourth, the €100 threshold for reduced delivery charges and free delivery for golden-card members has different impacts on online and off-line baskets. Figure 2 shows the distribution of shopping trips by basket size. The threshold does not affect off-line trip expenditures at all because off-line baskets do not peak above €100, and only 11.3% off-line trips exceed €100. The threshold also seems to have a very small effect on online trip expenditures: 8% of online trips are for amounts above €90 but below €100, and 70% of online trips are for amounts larger than €110. This is at odds with what is reported in Lewis et al. (2006) based on the U.S. data. At the same time, it alleviates the endogeneity concern vis-à-vis delivery charges.

Fifth, there is considerable heterogeneity in households' average online shopping propensity. We regress their overall proportions of online trips on their demographics and characteristics of their most frequented off-line stores. Households are more likely to shop online if their most frequented off-line stores do not have parking facilities or are smaller, even though large but distant stores with parking lots are available. Downtown households are less likely to shop online than suburban households,⁴ even though their most frequented off-line stores are smaller and less likely to have parking facilities. If a household's

average basket has more heavy/bulky items, it is more likely to shop online; if its average basket has more perishables, it is more likely to shop off-line.

Sixth, there is a much less variation in online shopping than in off-line shopping both in basket size and trip interval. The coefficient of variation is 0.24 (std. dev. = 0.16) for online baskets and 0.73 (std. dev. = 0.32) for off-line baskets; 0.52 (std. dev. = 0.25) for online-online trip intervals and 0.86 (std. dev. = 0.32) for off-line-off-line trip intervals. These statistics imply the more regular nature of online shopping. Customized shopping lists like the previous shopping lists created by households in the online channel might help reduce variation in the online basket size. The greater variability in off-line trips may reflect the role of fill-in trips, as well as the role of promotions that draw consumers into the store only to cherry pick.

Promoted items account for 4.9% heavy items bought online and 3.9% heavy items bought off-line, and the figures are 1.2% and 1.7% for perishables, respectively. Although the economic benefits of price promotions across channels are the same, these promotions seem to affect channel preferences via differences in their *perceived* benefits because of differences in the products bought in the two channels. To further investigate this finding, we created a daily time series (by aggregating across all households) of the proportion of store visits in each day that result in online purchasing. We then regressed this proportion on the total number of items promoted, the number of perishables promoted, and the number of heavy/bulky items promoted that day (along with other controls). This regression revealed that the number of heavy/bulky items promoted significantly increased the online trip proportion, whereas

⁴ Downtown households have a lower car ownership than suburban households (Matas et al. 2009), so they should be *more* likely to shop online if not having a car is an important factor for a household to shop online. We find the opposite, which indicates that car ownership may not play an important role in channel choice.

the other two decreased the proportion. This seems to suggest that promotions of different items differentially influence the propensity to visit the different channels, e.g., observing promotions of perishable items encourages customers to visit the off-line store. *Therefore, we need to account for the differential effects of promotions across channels, even though the economic benefits of these promotions are the same across channels.*

In sum, the observed shopping patterns appear to show that transaction costs of shopping differ by channel, basket size and composition, shopping occasion, and household characteristics. Our econometric specification tries to incorporate these into the analysis.

4. Econometric Specification

4.1. Description of Shopping Costs

A shopping trip involves direct costs and transaction costs. As the retailer has uniform pricing in the two channels, direct costs drop out in any discrete model of channel choice, so we do not consider them. We now describe how we measure the nine types of transaction costs identified. For a given basket, search costs to find a store are unlikely to play a role in our case, although in-store search costs will differ, which are accounted for as follows:

(1) *Time costs* are the opportunity costs of travel time and in-store shopping time. Shopping time is an important factor in store choice decision. Messinger and Narasimhan (1997) find that the growth of one-stop shopping has been a response to the growing demand for timesaving convenience. Pashigian et al. (2003) examine how firms respond to the increasing cost of a shopper's time in ways that economize on shopping time. Consistent with the literature (e.g., Bell et al. 1998), we use home-store distance as a proxy for travel time. As did Hoch et al. (1995), we scale distance in the downtown area by 2 to reflect its greater congestion. We also scale distance by two if the shopping trip occurs at peak weekday hours (7:30–9 A.M. and 6–8 P.M.) to reflect the longer time needed to travel the same distance.

In-store shopping time primarily is a function of the number of items purchased. We use the number of basket items as a proxy for in-store shopping time. Online item selection and checkout differ from that off-line. For the same set of items, a household may need less time to shop online than off-line, so we allow the marginal costs of shopping time to differ by channel.

The opportunity costs of time also depend on when a household visits a store. We include a weekday dummy to capture the higher opportunity costs of time on weekdays. Furthermore, a consumer may not be able to leave her office to buy groceries in a physical store during office hours but can still do so in the

online store, implying that the costs of off-line shopping during office hours may be high compared with online shopping. We include a dummy variable for trips that occurred during office hours (8 A.M.–noon, and 2–6 P.M.).

(2) *Transportation costs* are directly related to the home-store distance. Consistent with the literature (e.g., Bell et al. 1998), we use home-store distance as a proxy for transportation costs. Thus, home-store distance captures the effect of both travel time and transportation costs. To more accurately quantify transportation costs and travel time, we need trip-level mode of transportation used, which is not available to us. Using information at some aggregate level, such as zip-code car ownership, is not useful because this information does not vary over our short data duration. Furthermore, there is limited variation from one zip code to another. In the empirical analysis, we interact distance with household income level to capture differences in time costs across households.

(3) *Psychic costs* are difficult to measure but are closely related to the characteristics of the shopping channels themselves. We use channel dummies and a household-, trip-, and channel-specific error term to reflect these costs.

(4) *Adjustment costs as a result of product unavailability* and (5) *in-store search costs*: As with psychic costs, these costs are channel specific and difficult to measure. They may vary over time because of different products not available in different channels at different times and familiarity with product layouts. We represent them by a household-, trip-, and channel-specific error term.

(6) *Basket delivery costs*: Smith and Brynjolfsson (2001) find consumers are more sensitive to delivery charges than to product prices when shopping online. Lewis et al. (2006) find a significant effect of nonlinear shipping and handling fees on purchase incidence and expenditure decisions for dry groceries online. The shadow price of delivery charges differs for online and off-line shopping. For online shopping, delivery charges cover transportation costs to and from the store, opportunity costs of travel time, and physical costs of basket picking and carrying. For off-line shopping, they only cover physical costs of basket carrying. Thus, households may be less sensitive to delivery charges when shopping online than off-line.

(7) As with adjustment costs, *waiting costs for basket delivery* are difficult to measure and may vary over time because of instances of more urgent need of a product than others. We represent them by the channel dummy and by the household-, trip-, and channel-specific error term.

(8) *Physical costs* of picking items are related to the number of items, particularly the number of

heavy/bulky items purchased. It is a simple mouse click for online shopping but can be high for off-line shopping. We use the number of heavy/bulky items as a proxy for item-picking costs and allow the marginal costs to differ by channel. The costs of carrying a basket home are a function of the number of heavy items in the basket and the distance traveled. We use their interaction as a proxy for basket-carrying costs and set them to 0 for online shopping and off-line shopping with home delivery. The costs of carrying a basket also depend on weather conditions, so we allow weather to influence the choice of online and off-line channels differently.

(9) *Costs of inability to verify product quality prior to paying* are primarily a function of perishables bought because product qualities such as freshness can vary a lot from one purchase occasion to the next, whereas the quality of packaged goods does not change much over purchase occasions. We use the number of perishables bought as a proxy for these costs. Physical stores allow households to check product quality prior to paying; the online store lacks this ability because the actual quality can differ from what is shown on the computer screen. Thus, households may be more likely to visit the off-line channel when buying more perishables.

Intrinsic channel preferences and an error term capture, respectively, the time-invariant (channel image, environmental concerns, etc.) and time-varying aspects of unobserved transaction costs. Thus, given our data, we are able to quantify the important types of transaction costs. To account for differences in the *perceived* benefits of promotions across channels for the different products purchased by the households, we include overall promotions, promotions of heavy/bulky items, and promotions of perishable items as drivers of channel choice. Promotions could result in households choosing the off-line channel because of lower cherry-picking costs in the off-line stores.

4.2. Econometric Model

On each shopping trip, a household compares the expected costs of different channels and chooses the one with the lowest expected costs to maximize its utility. We assume households have rational expectations over shopping costs, so expected costs are the same as actual costs. It is reasonable to assume rational expectations on weather, delivery charges, and travel distance. Given chainwide promotions and the availability of the Internet option, it is also reasonable to assume rational expectations over promotions. Household h 's expected total transaction costs of shopping at channel j on trip t are

$$FC_{hjt} = \alpha_{hj0} + \alpha_{hj1}d_{hjt} + \alpha_{hj2}WK_t + \alpha_{hj3}OH_t \\ + \alpha_{hj4}NI_{ht} + \alpha_{hj5}NP_{ht} + \alpha_{hj6}NH_{ht}$$

$$+ \alpha_{hj7}NH_{ht} \cdot d_{hjt} + \alpha_{hj8}W_t + \alpha_{hj9}DC_{ht} \\ + \alpha_{hj10}P_{ht} + \alpha_{hj11}PP_{ht} + \alpha_{hj12}PH_{ht} + \varepsilon_{hjt}, \quad (1)$$

where FC_{hjt} is total transaction costs and d_{hjt} is the home-store distance, which changes from trip to trip because a household does not always visit the same off-line store. Variable WK_t is a weekday dummy and OH_t is an office hour dummy; NI_{ht} , NP_{ht} , and NH_{ht} are numbers of items, perishables, and heavy/bulky items, respectively; and $NH_{ht} \cdot d_{hjt}$ is the interaction between heavy items and distance. W_t is a dummy for bad weather and DC_{ht} is the delivery charge; P_{ht} , PP_{ht} , and PH_{ht} are the overall promotion index and promotion indices of perishables and heavy/bulky items; and ε_{hjt} stands for unobserved search, adjustment, and waiting costs that we assume are normally distributed.

We constrain the coefficients for the 196 off-line stores to be the same and subtract the costs of shopping at the off-line channel from those at the online channel. Two factors necessitate our assumption on the parameters across off-line stores. First, we do not have a sufficient number of observations for each store to estimate separate effects. Second, even if we did, it would be computationally infeasible to estimate store-specific parameters for each of the 196 stores. At the same time, the assumption is not restrictive for the following reasons: (1) All off-line stores belong to the same chain and thus share the same or similar chain image, service quality, product layout, returns policy, etc. The effects of other factors such as store distance and size are accounted for. (2) We estimate coefficients for each household. Even though there are 196 stores, 32.9%, 31.1%, and 20.2% of the households only visit one, two, or three off-line stores, respectively. (3) For 74.1% of the households, the most frequented off-line stores account for over 60% of off-line trips. This combined with household-level parameters should capture some of the heterogeneity across stores as well. In §5.5, we try to assess the impact of this assumption. Nevertheless, we acknowledge this as a possible limitation of our analysis.

Next, we drop subscript t for ease of exposition. Define $Y_h \equiv FC_{h, \text{on}} - FC_{h, \text{off}}$, $\varepsilon_h \equiv \varepsilon_{h, \text{on}} - \varepsilon_{h, \text{off}}$, and $\alpha_h \equiv \alpha_{h, \text{on}} - \alpha_{h, \text{off}}$. Let $X_h \equiv \{NI_{ht}, NP_{ht}, NH_{ht}, NH_{ht} \cdot d_{hjt}\}$ be the vector of numbers of items, perishable items, heavy/bulky items, and its interaction with distance; let G_h be other costs in Equation (1) and I_h be the choice indicator. β_h represents the subset of α_h that is associated with the variables X_h , and γ_{1h} represents all other coefficients in α_h . Thus, we have the following relationships:

$$Y_h = G_h \gamma_{1h} + X_h \beta_h + \varepsilon_h, \quad (2)$$

$$I_h | Y_h \sim \text{Binomial Probit}. \quad (3)$$

4.3. Endogeneity of Channel Choice Drivers

Numbers of items, perishables, and heavy/bulky items are household decision variables on trip t and thus may be endogenous to the channel choice decision because unobserved factors (as discussed previously) that influence numbers of items bought also likely influence the choice of shopping channels (i.e., X_h could be correlated with ε_h). There are several approaches to address this issue: (i) explicitly model each of these decisions—a full-information structural approach, (ii) the control function approach of Petrin and Train (2010), and (iii) the plausibly exogenous approach of Conley et al. (2010). These approaches are similar in nature. The second and third approaches rely on finding exogenous instruments for the endogenous variables, and the first approach requires the presence of some excluded variables.

The plausibly exogenous approach, of which the conventional instrumental variables technique is a special case, relaxes the exclusion restriction that instrumental variables have no correlation with the unobservables; thus it allows for the use of less-than-perfect instruments. It works well with both strong and weak instruments, so we adopt this approach. We find a set of instruments Z_h for endogenous variables X_h and include them into Equation (2) as follows:

$$Y_h = G_h \gamma_{1h} + X_h \beta_h + Z_h \gamma + \varepsilon_h, \quad (4)$$

$$X_h = (G_h Z_h) \Pi + v_h. \quad (5)$$

This approach essentially replaces the actual numbers of items, perishables, and heavy/bulky items with their expected numbers, which are determined by exogenous instruments ($G_h Z_h$). When $\gamma = 0$, this model reduces to the conventional instrumental variables model. Assume $(\varepsilon_h, v_h) \sim N(0, \Sigma)$. Let $p(\text{Data} | \gamma_{1h}, \beta_h, \gamma, \Pi, \Sigma)$ be the likelihood of the data, $p(\gamma_{1h}, \beta_h, \Pi, \Sigma)$ the prior distribution of the model parameters, and $p(\gamma | \gamma_{1h}, \beta_h, \Pi, \Sigma)$ the prior distribution of γ . The inference is based on the posterior distribution of $(\gamma_{1h}, \beta_h, \Pi, \Sigma)$, integrating out γ :

$$p(\beta_h, \gamma_{1h}, \Pi, \Sigma | \text{Data}) \propto \int [p(\text{Data} | \beta_h, \gamma_{1h}, \gamma, \Pi, \Sigma) \cdot p_\gamma(\gamma | \beta_h, \gamma_{1h}, \Pi, \Sigma) p(\beta_h, \gamma_{1h}, \Pi, \Sigma) d\gamma]. \quad (6)$$

To overcome the computational burden of this model, we cast it in an HB framework as

$$B = D\Delta + U_\omega, \quad U_\omega \sim N(0, V_\omega), \quad B \equiv \{\gamma_{1h}, \beta_h\}, \quad (7)$$

where D is household demographics and Δ is the second-stage coefficients. The common priors are

$$\begin{aligned} \gamma &\sim N(\mu_\gamma, A_\gamma^{-1}), \quad \text{vec}(\Pi | V_h) \sim N(\text{vec}(\bar{\Pi}), \Sigma_{22} \otimes A_\Pi^{-1}), \\ \Sigma &\sim \text{IW}(v_0, V_0), \quad \text{vec}(\Delta | V_\omega) \sim N(\text{vec}(\bar{\Delta}), V_\omega \otimes A_\Delta^{-1}), \\ V_\omega &\sim \text{IW}(v_{\omega 0}, V_{\omega 0}). \end{aligned} \quad (8)$$

We write out this model as a sequence of conditional distributions and run the Gibbs sampler to get the Markov chain Monte Carlo (MCMC) sequence (Chib 2001, Rossi et al. 2005, Conley et al. 2010). The posterior distributions are

$$\begin{aligned} Y_h | G_h, X_h, Z_h, \gamma_{1h}, \beta_h, \gamma, \Sigma &\sim \text{Truncated Normal}, \\ \{\gamma_{1h}, \beta_h\} | Y_h, G_h, X_h, Z_h, D, \gamma, \Pi, \Sigma, \Delta, V_\omega &\sim \text{Normal}, \\ \gamma | Y_h, G_h, X_h, Z_h, \{\gamma_{1h}, \beta_h\}, \Pi, \Sigma, \mu_\gamma, A_\gamma &\sim \text{Normal}, \\ \Pi | G_h, X_h, Z_h, \{\gamma_{1h}, \beta_h\}, \Sigma, \bar{\Pi}, A_\Pi &\sim \text{Normal}, \\ \Sigma | Y_h, G_h, X_h, Z_h, \{\gamma_{1h}, \beta_h\}, \gamma, \Pi &\sim \text{Inverted Wishart}, \\ \Delta | \{\gamma_{1h}, \beta_h\}, V_\omega, \bar{\Delta}, A_\Delta &\sim \text{Normal}, \\ V_\omega | \{\gamma_{1h}, \beta_h\}, v_{\omega 0}, V_{\omega 0} &\sim \text{Inverted Wishart}. \end{aligned} \quad (9)$$

The details are available in the electronic companion to this paper that is available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

5. Variable Operationalization and Robustness Checks

5.1. Choice of Instruments

To resolve the endogeneity issue, we need to find variables that affect the numbers of items, perishables, or heavy/bulky items a household buys on trip t but that do not directly influence the choice of the shopping channel. What would be reasonable instruments for the numbers of items bought? A key benefit of our data is that we have access to purchases in all categories (about 700 categories, as defined by the retailer, were bought by the households), including meat, fish, milk, and produce, that make up a significant proportion of a household market basket. This allows us to use each household's own top expenditure categories to construct its inventory variables. Importantly, these inventory variables influence channel choice primarily via influencing the numbers of items bought; i.e., lower inventory levels in many categories will result in greater numbers of items being bought on that trip.⁵ To check whether this is indeed the case, we run homogeneous probit models of channel choice that, respectively, include (1) inventory levels of all items, heavy/bulky items, and perishables; (2) numbers of items, heavy/bulky items, and perishables; and (3) both sets of variables in (1) and (2); and we look at changes in the model fit. We find that numbers of items have a much larger explanatory power than inventory levels. Furthermore, when numbers of items are included in the model, two of the three

⁵ We assume that there is no serial correlation in the channel choice equation errors as is typical in these models.

inventory coefficients are no longer significant (see the electronic companion for details). We therefore conclude that a household's numbers of low-inventory items, low-inventory perishables, and low-inventory heavy/bulky items are reasonable instruments for our endogenous variables (recall that our method does not require perfect instruments). We first identify the top 30 categories for each household by looking at the proportion of that household's total grocery expenditure (across all purchase occasions) accounted for by each category purchased. Because household-specific top 30 account for 81.1% of the basket value (std. dev. = 9.4%), using the top 30 is nearly equivalent to using the entire basket. This aggregation strategy substantially reduces the complexity of the model to a manageable scale without loss of relevant information. Next, we compute category inventory for each trip as in Gupta (1988) and find categories within each household's top 30 that have below-average inventories for that household. We then sum up the mean numbers of items across the low-inventory categories, perishables, or heavy/bulky categories for each household.

We also have information on the numbers of items bought by households that have similar demographics to the panel households but do *not* make a channel choice decision, i.e., "pure" online and "pure" off-line households. We compute trip-level mean numbers of items (heavy/bulky items, perishables) on each day for pure online and pure off-line shoppers, broken down by observed demographics, and match them to each sample household. The rationale for using these variables as instruments is that variation in numbers of items bought can be partly explained by unobserved consumer characteristics. As long as these characteristics are correlated with observed demographics, other households' purchases will be correlated with those of the target household. Because these other households do not make a channel choice, their numbers of items are unlikely correlated with the unobservables that drive channel choice for the households in our panel.

These two sets of instruments explain, respectively, 19.1%, 20.4%, and 23.9% of the variation in the number of items, number of perishables, and number of heavy/bulky items. Together with demographics, they explain 19.8%, 23.6%, and 24.7% of the variation; and in combination with exogenous variables in Equation (1), they explain 23.2%, 24.9%, and 28.7% of the variation in the endogenous variables (see Table 3). These instruments appear to have independent explanatory power. Although we recognize the potential consequences of using multiple weak instruments, this is less likely in our case because of the low correlations between instruments (see the electronic companion for details).

5.2. Operationalization of Promotion Variables

We first calculate category promotion depths as weighted averages of price cut depths across all items in the category, using each household's mean item expenditures as weights. We then compute an overall promotion index as the weighted average of category promotion depths, using each household's mean basket shares of its top 30 categories as the weights. Similarly, we compute promotion indices for perishables and heavy/bulky items as weighted averages of promotion depths. The weights are each household's mean basket shares of the respective items among its top 30 categories. Thus, the promotion indices are household specific.

5.3. Operationalization of Other Cost Variables

We use a weather indicator for rainy, stormy, and windy days. The home-store distance refers to zip-code-level distance between a household's home and the store visited on that occasion. On occasions when the online channel is visited, the distance for the off-line option is the value for that household's most frequented physical store.⁶ The distance is zero for online shopping.

The delivery charges variable (DC_{ht}) is defined as follows. For all off-line visits entailing delivery, the variable drops out because the charges are the same in both cases. For off-line visits with no delivery, $DC_{ht} = 0$ for the off-line channel and the amount corresponding to the basket size for the online channel, and $DC_{ht} = 0$ for golden-card members if the basket exceeds €100. For online visits, we assume that the off-line option corresponds to no delivery, so $DC_{ht} = 0$ for the off-line channel and DC_{ht} equals the actual charges for the online channel. We recognize that this is an approximation in that, to deal with this structurally, we need to consider off-line visits with and without delivery as two options. However, other than delivery charges, there is nothing else in the data that informs us of this choice, so we treat off-line visits as a single choice. Furthermore, as noted before, a very small fraction of off-line shoppers avail of the delivery option.

5.4. Demographics and Store Characteristics

These variables appear in the hierarchy as D in Equation (7) and allow us to account for observable household heterogeneity. Demographics include family size, number of children, and the residential area's economic status or income (low/medium/high). Household basket characteristics include mean basket shares

⁶ This is the best approximation as it is the cost for most off-line trips. Households visit their most frequented off-line stores for major trips and fill-in trips. We also used the distance to the closest off-line store (the two are the same for many households) and obtained similar results.

Table 3 Regression of Endogenous Variables on Instruments

	No. of items		No. of perishables		No. of heavy/bulky items	
	Estimate	Std. error	Estimate	Std. error	Estimate	Std. error
<i>Intercept</i>	−38.358	3.727	−4.121	0.675	−34.932	1.905
<i>Pure online shoppers' mean no. of items</i>	0.103	0.019	0.013	0.003	−0.006	0.010
<i>Pure off-line shoppers' mean no. of items</i>	0.374	0.031	0.049	0.006	0.197	0.016
<i>Pure online shoppers' mean no. of perishables</i>	0.477	0.082	0.248	0.015	0.058	0.042
<i>Pure off-line shoppers' mean no. of perishables</i>	−0.386	0.102	−0.009	0.018	−0.274	0.052
<i>Pure online shoppers' mean no. of heavy/bulky items</i>	0.527	0.035	−0.005	0.006	0.404	0.018
<i>Pure off-line shoppers' mean no. of heavy/bulky items</i>	−0.420	0.081	−0.051	0.015	−0.255	0.042
<i>Pure online shoppers heavy items · Distance</i>	−0.443	0.075	−0.019	0.014	−0.221	0.039
<i>Pure off-line shoppers heavy items · Distance</i>	0.706	0.246	−0.034	0.045	0.452	0.126
<i>No. of low inventory items</i>	0.558	0.029	0.035	0.005	0.127	0.015
<i>No. of low inventory perishables</i>	−0.071	0.094	0.350	0.017	−0.141	0.048
<i>No. of low inventory heavy/bulky items</i>	−0.067	0.044	−0.036	0.008	0.281	0.023
<i>Low inventory heavy/bulky items · Distance</i>	−0.533	0.081	0.025	0.015	−0.379	0.041
<i>Family size</i>	−0.095	0.262	−0.028	0.047	−0.062	0.134
<i>Preschool children</i>	−1.204	0.517	−0.063	0.094	−0.921	0.264
<i>Parking</i>	5.632	0.978	1.481	0.177	2.369	0.500
<i>Partner store (yes/no)</i>	8.682	1.100	1.159	0.199	3.495	0.562
<i>MBS: Ratio of heavy/bulky^a</i>	−9.750	4.710	−1.308	0.853	22.744	2.408
<i>MBS: Ratio of perishables</i>	7.965	6.036	19.806	1.093	1.613	3.086
<i>Medium economic status</i>	−1.117	1.411	0.069	0.255	0.521	0.721
<i>High economic status</i>	−0.060	1.274	0.309	0.231	0.316	0.651
<i>Overall promotion</i>	0.392	0.281	0.106	0.051	0.052	0.144
<i>Promotion of perishables</i>	−2.949	1.077	0.260	0.195	−1.234	0.551
<i>Promotion of heavy/bulky</i>	1.195	0.520	−0.029	0.094	0.604	0.266
<i>Delivery charges</i>	4.968	0.310	−0.203	0.056	3.293	0.158
<i>Weekday dummy</i>	−0.470	1.212	−0.701	0.219	−0.216	0.620
<i>Bad weather dummy</i>	−8.911	3.115	−0.314	0.564	−4.553	1.593
<i>Store distance</i>	0.134	0.024	0.007	0.004	0.077	0.012
<i>Office hours</i>	17.477	0.942	1.139	0.170	9.004	0.481
<i>R²</i>	0.232		0.249		0.287	
<i>N</i>	16,796		16,796		16,796	

^aMBS, mean basket share for each household across all store visits.

(across store visits) of perishables and heavy/bulky items. Characteristics of the most frequented physical store include store square footage (a proxy for in-store shopping time but also correlated with the store location—a large mall, etc.) and indicators for seafood shop, parking lots, delivery service, and Internet store partnership.

5.5. Robustness Checks

A model without instruments: We run an HB probit model of channel choice with the same channel choice drivers, but not accounting for the endogeneity of some channel choice drivers. We find that, consistent with the literature on endogeneity, not accounting for endogeneity leads to biased estimates. Some coefficients (e.g., coefficient of delivery charges) are biased toward zero, and some coefficients (e.g., weather dummy) are wrongly signed (see the electronic companion for details).

Zip-code-specific coefficients: To check how restrictive it is to constrain all off-line stores to have the same coefficients, we grouped off-line stores into 20 different bins based on their zip-code location—stores in the

same zip code are grouped into the same bin, and all those zip codes with less than 200 observations are grouped into one bin. We estimated different sets of parameters for different bins for the homogeneous probit model. We find that it is not possible to precisely identify these coefficients, and the great majority of the coefficients are not statistically significantly different from each other.

To check whether possible differences in store assortments play a role in store choice, we estimate an HB model with households that primarily shop at partner stores and stores of the same size as the partner stores. We find that the results obtained from this subsample are very similar to those from the entire sample, which alleviates concerns about store assortments influencing store choice.

Households could systematically differ in their package size choices even within a particular group of items (perishables, heavy/bulky items). To account for this variation, we also tried a different operationalization of the number of items by weighting the numbers by pack sizes relative to the smallest size in each group. This did not change our results materially.

Table 4 Population Parameter Estimates and Standard Deviations

	Mean		Std. dev.	
	Estimate ^a	Std. error	Estimate	Std. error
<i>Intercept</i>	−0.393	0.434	1.961	0.334
<i>Store distance</i>	0.512	0.209	0.591	0.189
<i>Weekday dummy</i>	1.411	0.420	0.350	0.067
<i>Office hours</i>	2.359	1.119	2.342	0.525
<i>No. of items/10</i>	0.725	0.411	0.506	0.085
<i>No. of perishables/10</i>	−0.361	0.115	0.314	0.064
<i>No. of heavy/bulky items/10</i>	0.581	0.277	0.668	0.115
<i>No. of heavy/bulky items · Distance/10</i>	1.095	0.460	1.145	0.189
<i>Bad weather dummy</i>	1.303	0.781	0.623	0.085
<i>Delivery charges</i>	−1.115	0.380	0.499	0.103
<i>Overall promotion index</i>	−0.120	0.064	0.127	0.017
<i>Promotion index of perishables</i>	−0.069	0.041	0.420	0.069
<i>Promotion index of heavy/bulky items</i>	0.125	0.069	0.268	0.031

^aA positive number means it is less costly to shop online, and a negative number means it is more costly to shop online.

We note that we include “family size” as observed heterogeneity in the effects of different numbers of items variables to account for heterogeneity along this dimension.⁷

6. Main Findings

We estimate the model parameters using MCMC methods. We use diffuse priors to let the data dictate the estimates. We take 50,000 draws, keep every 10th, and use the first 10,000 draws as the burn-in period. We check the MCMC sequences to ensure they reach their equilibrium distributions. We present our model estimates in Tables 4 and 5. Table 4 displays the effects of the variables on the choice of the online over the off-line channel. A positive number means it is less costly to shop online, and a negative number indicates that it is more costly to shop online.

6.1. Effects of Shopping Costs on Channel Choice

Travel costs: Consistent with the literature (Forman et al. 2009, Bell et al. 1998), we find travel time and transportation costs discourage households from visiting the off-line channel and encourage them to visit the online channel. The farther away a household lives to the physical store, the more likely it will visit the online channel. This points to one particular attractiveness of the Internet channel, i.e., the elimination of travel and transportation costs.

In-store shopping time: We find that the more items a household needs to buy on a shopping trip, the more likely it will visit the online channel than the off-line

channel. This reflects the time efficiency of in-store shopping at the online channel. When a household needs to buy a large number of items, the efficiency gain of shopping online can be substantial. Thus, households are more likely to shop online when they want to buy large baskets. Households’ valuation of the time efficiency of online shopping is also evidenced by their higher tendency to shop online on weekdays than on weekends, because people are usually more time constrained and have higher opportunity costs of time during weekdays than weekends. Households also have a much greater propensity to shop online over off-line during office hours.

Delivery charges: Even though the benefit of delivery charges for an online order is bigger than an off-line order, delivery charges still have a statistically significant effect in discouraging households from visiting the online channel. Although customers currently bear the expense of going to stores themselves, they are reluctant to pay for their shopping baskets to be brought to them instead. This is consistent with the literature (e.g., Kacen et al. 2003) and with the Nielsen’s finding that delivery costs are one of the two most common reasons Americans give for not buying groceries online (Economist 2010).

Physical costs of item picking and basket carrying: We find that the need to buy a larger number of heavy/bulky items drives households to the online channel and discourages them from visiting the off-line channel. This effect becomes stronger when travel distance is taken into account. The farther away a household lives to the physical store, the less likely it will buy heavy/bulky items from the physical store, and the more likely it will buy them from the online channel. This is further evidence for the higher physical costs of picking items and carrying the basket when shopping in the off-line stores. Households are more likely to shop online during bad weather days,

⁷In addition to these, we also carried out robustness checks on the store distance, the time window for peak and office hours, the classification of various items as heavy/bulky, the sensitivity of our results to dropping the promotion variables, and the use of subsets of our instruments. Some results are in the electronic companion, and others are available from the authors upon request.

Table 5 Household Heterogeneity Parameter Estimates: Channel Choice

	Intercept	Store distance	Weekday	Office hours	No. of items/10	No. of perishables	No. of heavy	Heavy Distance	Weather	Delivery charge	Overall promo	Perish promo	Heavy promo
Estimates													
<i>Intercept</i>	−0.448**	0.635*	0.902**	2.953*	0.635**	−0.702**	0.511**	1.311**	1.434**	−0.194*	−0.092**	−0.043**	0.124**
<i>Family size</i>	0.019*	0.039*	0.003	−0.076	0.022**	−0.099**	0.016**	−0.038	0.012	−0.033**	0.024*	0.050	0.049**
<i>Preschool children</i>	0.011*	0.062**	0.103*	0.367**	0.019*	0.136**	−0.010	0.094*	0.107**	−0.023	0.073*	−0.173*	−0.045
<i>Fish shop</i>	−0.042*	−0.048	0.406**	−1.276*	−0.058*	−0.026	−0.053*	−0.315**	0.644	−0.122*	0.066	−0.207*	−0.115*
<i>Parking</i>	0.068	0.135*	−0.182*	−0.861*	−0.041*	−0.077**	−0.021*	−0.019	0.779**	−0.054*	−0.204**	−0.263*	0.260*
<i>Partner store</i>	−0.003	−0.010	0.134	1.015**	−0.019	−0.202	0.002	0.090*	0.881*	−0.023	−0.155**	−0.032	0.165*
<i>Delivery service</i>	−0.006	−0.065**	−0.207*	−0.729**	−0.007	−0.169*	−0.008	0.011	−0.984**	0.092**	0.179**	−0.357*	−0.065
<i>Store size/1,000</i>	0.203*	−0.369*	−0.718*	−0.216	−0.240*	−0.197*	−0.184	−0.639*	0.519*	−0.251*	0.187	0.833*	0.172*
<i>Medium income</i>	0.039*	0.230*	0.265**	1.758*	0.055*	−0.050	0.061	0.248*	0.862**	−0.044	−0.056	−0.494*	0.167*
<i>High income</i>	0.073**	0.214**	0.306**	2.970**	0.084**	−0.168*	0.126**	0.209*	0.694*	0.022	−0.116**	−0.431*	0.187**
<i>Downtown</i>	−0.028*	0.407**	0.281	1.347*	0.016	0.026	0.020	0.108*	1.136**	0.083	−0.126**	−0.156**	0.201**
<i>MBS perishables</i>	−0.615**	−0.645	−0.970**	−0.510	−0.547**	−0.806**	−0.408**	−1.115**	−2.053*	−1.181**	−0.913**	−0.948**	−0.148*
<i>MBS heavy/bulky</i>	0.380**	0.423**	0.799*	0.868**	0.429**	−0.073	0.477**	0.445*	1.788**	0.870**	0.290	0.050	0.982**
Standard errors													
<i>Intercept</i>	0.136	0.320	0.358	1.561	0.099	0.299	0.069	0.257	0.660	0.108	0.039	0.028	0.052
<i>Family size</i>	0.010	0.020	0.027	0.181	0.007	0.032	0.005	0.030	0.103	0.015	0.013	0.060	0.024
<i>Preschool children</i>	0.005	0.029	0.063	0.141	0.011	0.031	0.011	0.048	0.051	0.040	0.038	0.102	0.067
<i>Fish shop</i>	0.021	0.116	0.161	0.729	0.032	0.077	0.027	0.137	0.748	0.067	0.078	0.127	0.068
<i>Parking</i>	0.051	0.075	0.105	0.480	0.022	0.034	0.013	0.119	0.372	0.028	0.079	0.149	0.156
<i>Partner store</i>	0.057	0.115	0.188	0.355	0.030	0.278	0.027	0.049	0.498	0.085	0.075	0.314	0.102
<i>Delivery service</i>	0.046	0.022	0.122	0.230	0.027	0.096	0.022	0.091	0.461	0.040	0.069	0.183	0.120
<i>Store size/1,000</i>	0.109	0.202	0.401	2.558	0.135	0.119	0.105	0.378	0.268	0.155	0.298	0.426	0.102
<i>Medium income</i>	0.020	0.125	0.118	0.931	0.031	0.095	0.033	0.152	0.326	0.112	0.090	0.261	0.088
<i>High income</i>	0.021	0.105	0.117	0.909	0.027	0.103	0.028	0.121	0.351	0.108	0.057	0.223	0.091
<i>Downtown</i>	0.016	0.126	0.205	0.730	0.032	0.139	0.025	0.059	0.408	0.089	0.059	0.078	0.080
<i>MBS perishables</i>	0.202	0.326	0.484	0.358	0.125	0.306	0.099	0.341	1.228	0.319	0.303	0.441	0.082
<i>MBS heavy/bulky</i>	0.188	0.200	0.476	0.360	0.102	0.499	0.080	0.261	0.850	0.303	0.319	0.047	0.398

Note. MBS, mean basket share for each household across all store visits.

*Indicates significance at the 5% level; **indicates significance at the 1% level.

because it is more inconvenient to carry the basket home on rainy or windy days.

Costs of inability to verify product quality prior to purchase: We find that the inability of the online channel to facilitate product quality verification prior to paying is a hurdle for households to purchase online. When households need to buy a larger number of perishable items, they are less likely to visit the online channel and more likely to visit the off-line channel to inspect the quality of perishables prior to purchase. This ability reduces a household's risk of buying lower-than-expected-quality products in the off-line channel, but it increases the risk of buying them in the online channel. When households need to buy more perishables, this cost difference across the two channels can be large. Therefore, households are more likely to visit the off-line channel when they need to buy more perishables.

The channel-specific effects of buying perishables off-line and buying heavy/bulky items online reflect the different transaction costs of shopping in the two channels. The household-level coefficients of perishables and heavy/bulky items are negatively correlated, reflecting the trade-offs in transaction costs that

households have to make when shopping at the two channels.

Effects of price promotions: Overall, promotions are more likely to drive households to the off-line channel than the online channel. Promotions of different categories have channel-specific traffic building effects. Compared with the off-line option, promotions of perishables discourage households from visiting the online channel, but the effect is only marginally significant. On the other hand, promotions of heavy/bulky items significantly increase online channel visits.

6.2. Household Heterogeneity

There exists considerable household heterogeneity in the intrinsic preference for the online channel and in the valuation of and sensitivity to the different shopping cost components. All household heterogeneity parameters are significant at the 1% level of significance (see Table 4).

Table 5 shows how a household's intrinsic channel preference and sensitivity to transaction costs depend on household demographics, store characteristics, and mean basket characteristics. Most coefficients are significant at the 5% or 1% level and make intuitive sense. Family size influences the intrinsic preference

for the online channel—all else equal, larger families have a higher preference for the online store and are more likely to visit the online store when living farther away from the physical stores, when buying more items or more heavy items, or when the overall promotion level and promotion level of heavy/bulky items are high. However, larger families are more sensitive to delivery charges. Households with more preschool children have a stronger preference for the online channel and are more likely to shop online during weekdays, office hours, bad weather days, or when the overall promotion level is high. Households have a lower preference for the online channel if their most frequented off-line stores have a fish shop or their baskets consist of more perishables. Downtown households have lower preference for the online channel than suburban households do, but they are more likely to shop online during office hours, bad weather days, or when buying more heavy items and traveling longer distance. A household's basket share of heavy/bulky items also drives many of the effects. Households with larger basket shares of such items have a stronger preference for the online channel and are more likely to shop online during weekdays, office hours, bad weather days, when traveling longer distance, when buying more heavy items, and when heavy items are promoted. They are also less sensitive to delivery charges. In contrast, a larger basket share of perishables increases a household's intrinsic preference for the off-line channel and makes a household a lot more sensitive to delivery charges and more responsive to promotions of perishables in the off-line channel.

6.3. Quantifying the Monetary Value of Transaction Costs

The monetary value a household attaches to each type of transaction cost can be calculated by dividing the coefficient of the transaction cost by the coefficient of delivery charges. The latter coefficient is interpreted as the marginal utility of income (see, for example, Smith and Brynjolfsson 2001).⁸ Note that these monetary values are relative to the option of shopping off-line and cannot be interpreted directly as a household's willingness to pay because for that we need to include an outside option.

⁸ Researchers usually use the price coefficient to compute the monetary value of attributes (Smith and Brynjolfsson 2001). In our case, the price coefficient applies to the sum of direct and delivery costs. Because direct costs drop out of the model, the coefficient of delivery charge represents the price coefficient. If consumers are more sensitive to delivery charges than to product prices as in Smith and Brynjolfsson (2001), using the delivery charge coefficient will underestimate the monetary value of the various transaction costs. Nevertheless, this method still gives us the monetized value of transactions relative to the value of the delivery charges. Note also that we cannot use the coefficient of the promotion index because it does not serve the role of price in the model.

The Bayesian approach allows us to compute each household's valuation of each transaction cost and characterize its distribution. Relative to shopping off-line, shopping online provides a value of €0.59 (std. dev. = €1.96) for one kilometer that a household has to travel. It is worth €1.43 (std. dev. = €2.14) to a household for shopping online rather than off-line on a weekday versus weekend, worth €1.17 (std. dev. = €2.15) for shopping online on a bad weather day as compared with a good weather day, and worth €1.98 (std. dev. = €2.86) for shopping online during office hours vis-à-vis nonoffice hours. For every 10 items bought, the time savings in the online channel over the off-line option is equivalent to €0.69 (std. dev. = €1.26), the costs of picking and putting 10 heavy/bulky items into the shopping cart are €0.53 (std. dev. = €0.93), and the costs of carrying 10 heavy/bulky items one kilometer are €0.98 (std. dev. = €1.82) (in addition to the €0.59 above). The cost a household attaches to not being able to check product quality prior to paying in the online store is €0.34 (std. dev. = €0.64) for every 10 perishables. The relatively low value attached to quality inspection might be because households have been assured about relative quality equivalence across channels.

There exists considerable household heterogeneity in the monetary value attached to various transaction costs. We present the estimated empirical distributions for some types of these costs in Figures 3–5. Figure 3 is the distribution of the value of shopping online on weekdays, Figure 4 gives the distribution of the value of shopping online on a bad weather day, and Figure 5 provides the distribution of the amount a household values traveling one kilometer without any heavy/bulky items. A majority of households values weekday online shopping by €0.5 to €2; 13.5% value it at €2 to €5, and 3.6% value it at more than €5. About 15% of households are not bothered by bad weather, whereas 20.3% of households attach more than €2 to

Figure 3 Distribution of Relative Value of Weekday/Weekend Time

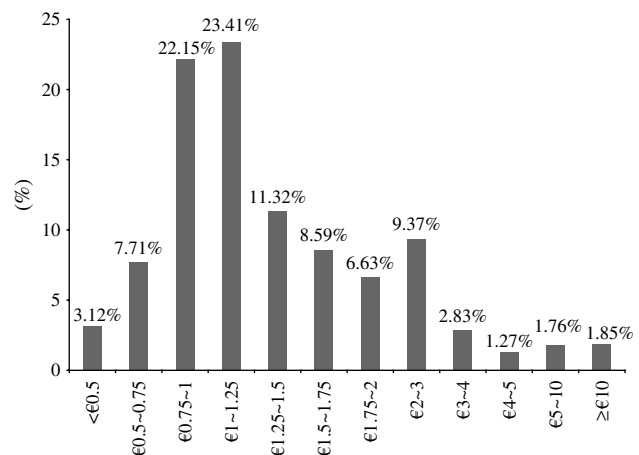


Figure 4 Distribution of Relative Value in Good/Bad Weather

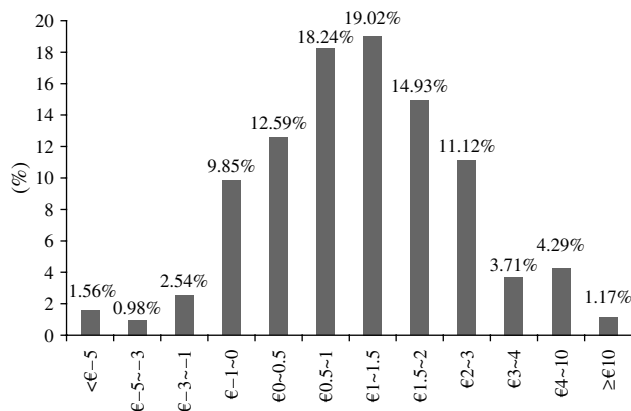


Figure 5 Distribution of Value Attached to One Kilometer of Travel

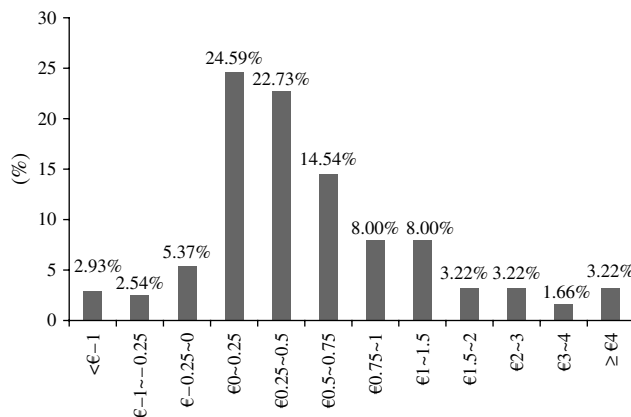
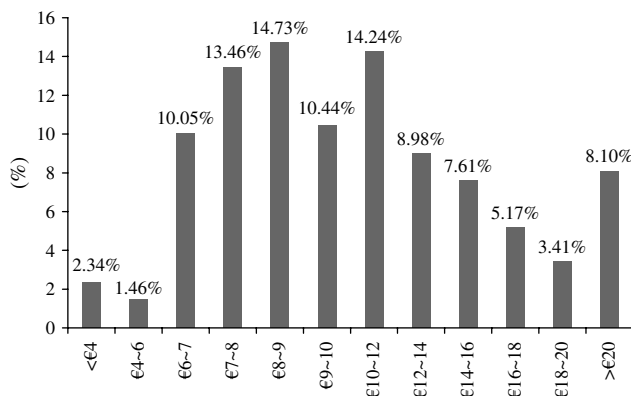


Figure 6 Distribution of Relative Value of Internet Channel



online shopping in bad weather relative to the alternative. Some households do not seem to have any disutility of travel, whereas others need more than €4 compensation for one kilometer of travel.

With the value a household attaches to each type of transaction cost, we next compute the value of the online channel relative to the off-line channel for each store visit by the households by summing up the costs on each trip and computing the average value.

Focusing *only* on the *online* trips, we find the relative average value per trip of the Internet option is €11.22 (std. dev. = €6.40). Figure 6 shows the distribution of the relative value across households. A total of 62.9% of households value the Internet option at €6 to €12 per online trip, which more than offsets the delivery charges for online shopping. Only 3.8% households have a value below €6—the delivery charges for baskets smaller than €100, and only 2.3% households have a value lower than €4—the delivery charges for baskets larger than €100. Because households do not seem to increase their basket sizes to take advantage of the delivery charge schedule, this implies that the delivery charge the retailer has set is reasonable for the value it provides relative to the off-line option, and households obtain a net surplus by shopping at the online channel for these trips. The retailer can use this as a selling point for promoting the online channel to households.

What would the value of the Internet option be if an *off-line* trip occurred *online*? We find that 95.2% of these trips had a value of €1 to €4, which is lower than the delivery charges. The major contributors for the low value are smaller basket size, higher proportion of perishables in the basket, weekend trips, and nonoffice hour trips. This explains why the households did not visit the online channel on these trips. This is also consistent with the observed sorting strategy households employ when deciding between the two channels on each shopping trip.

7. Discussion and Managerial Implications

Our findings have important implications for managerial practice. Because we do not incorporate competitors' information because of a lack of data, it is important to interpret these results as directional and suggestive.

7.1. Enterprise Design for Online and Off-line Grocery Stores

There is a growing trend for conventional grocery retailers (e.g., Safeway) to start online operations (<http://www.safeway.com>). Our study provides some guidance on product offerings for online and off-line channels, such as more varieties of heavy/bulky items for the online channel and more varieties of perishables for the off-line channel. In particular, our findings have implications for the positioning of the online channel. The American online grocer Peapod's tagline is "Smart shopping for busy people," which is more of a sorting strategy for attracting customers. We find that the online channel is not only for busy people but also for busy days. Retailers can promote the online channel as "smart shopping for busy people and on busy days," which is a combination of both

sorting and pooling strategies. This can help enlarge the customer base for the online channel.

Although the benefits of shopping online appear to more than offset the delivery charges with the Internet option actually having a net surplus for households' online trips, delivery charges are still a deterrent to shopping online. One reason might be that the benefits of shopping online are implicit in that households might have difficulty in converting the various hard-to-measure-and-quantify transaction costs, e.g., how much time is saved and the money value of the time saved, into concrete money metrics. Delivery charges on the other hand, are upfront out-of-pocket money. The retailer can make the implicit benefits explicit to increase the attractiveness of the online channel. It can educate customers that delivery charges are only nominal if they factor in what they save on time, gas, bus fare, and other transaction costs as well as the avoided hassles and cumbersomeness of getting heavy/bulky items off the shelf, putting them into the cart, getting them out onto the checkout counter, and so on. For example, they can advertise the difficulties in carrying a heavy basket on a rainy day and emphasize the ease in having heavy items delivered to the house. Some online grocers have actually adopted this strategy: <http://www.peapod.com> advertises "You order online. We shop. We deliver. Order heavy items" (Peapod 2011).

7.2. Channel-Specific and Category-Specific Promotions

We find that promoting heavy/bulky items and perishables has different traffic building effects for online and off-line channels. Retailers can adopt category-specific and channel-specific promotions to manage channel traffic. They can promote perishables in the off-line channel to increase off-line traffic and promote heavy/bulky items, particularly among those households living farther away from physical stores to increase online traffic. Our results suggest that online traffic can increase significantly if retailers run price discounts on heavy items to their customers, especially in markets such as ours where not everyone drives to do grocery shopping. If the retailer promotes all heavy/bulky items only at the online channel at the observed promotion depths, online traffic can increase by 8.33%; if it promotes perishables only at the off-line channel at the observed promotion depths, off-line traffic can increase by 4.89%.

7.3. Customer Segmentation and Targeting

Our results provide some useful bases for customer segmentation and targeting. One way to segment customers is by the overall value the Internet option provides to customers. The 32.1% of households that

value the Internet over €12 per online trip have larger online baskets, order more heavy/bulky items, and live far away from physical stores. They have stronger preferences for the online channel and are more responsive to promotions of heavy/bulky items. The retailer can consider promoting the online channel to large-basket households and households with baskets dominated by heavy/bulky items. Another way to segment customers is based on the values associated with specific transaction costs. For example, 19.3% of households have a high disutility of travel, and online shopping is a very attractive option for these shoppers. The retailer can use e-mail promotions and targeted coupon dropping to attract these households to the online store.

7.4. Quantifying Societal Benefits of Online Shopping

Although people drive to shop for groceries at very high frequencies, the environmental consequences of such activities often go unheeded. According to the U.S. Environmental Protection Agency (2005), the mean number of miles driven per year for all passenger vehicles is 12,000 miles, and the annual emissions of greenhouse gases (GHG) from a typical passenger vehicle equate to 5.5 metric tons of carbon dioxide (CO₂) equivalent. According to newly released IRI data, U.S. households on average make 104 grocery shopping trips per year (Chintagunta and Chu 2011). If the average driving distance to a grocery store is 3 miles, the total annual driving distance for grocery shopping will be 624 miles, and the GHG emissions per vehicle year will be 0.29 metric tons of CO₂ equivalent, which is a large enough number to warrant our attention.

Online grocery shopping is a greener way of shopping than driving to shop off-line. Let us assume for simplicity that on an off-line grocery shopping visit, the household does not engage in other activities such as going to the post office. Thus, saving an off-line trip would avoid driving to the grocery store and back. The Internet grocery store could then benefit society by reducing driving trips. One online order is equivalent to 3.5 off-line orders by basket size. If one truckload can fulfill the delivery of 20 online orders, the Internet store will reduce off-line shopping trips by the magnitude of 70, thereby reducing carbon emissions. The households made 5,721 online trips. Without the Internet store, they would have to make 20,024 more off-line trips. If half of them were made by driving, 10,012 driving trips would occur, compared with 286 delivery truckloads. As long as the GHG emissions associated with the latter are lower than those with the former, there would be a net societal benefit from online shopping. We acknowledge that this

is a rather simplistic, “back-of-the-envelope” computation. Nevertheless, this could be a potential benefit to increased online grocery shopping, especially if the online store uses ecofriendly delivery trucks. For example, <http://www.peapod.com> is promoting a greener delivery policy. It appeals its customers to “help us reduce carbon emissions by choosing a ‘Green Delivery Window’ which allows us to consolidate orders in a specific area thereby reducing the mileage between orders” (Peapod 2011). Given worldwide environmental concerns over GHG emissions, it will be very appealing to position and advertise the green aspect of online shopping. Shoppers may be more willing to shop online and pay for delivery if they are informed of the positive environmental effect their choice has.

7.5. Summary

The main contributions of this study are in formulating a transaction costs model of grocery channel choice in the presence of the Internet option and quantifying the transaction costs of off-line shopping relative to online shopping for a grocery retailer. Substantively, we show how these costs can be used to make explicit the relative costs of off-line shopping when buying a large number of items, on bad weather days, on weekdays, and when the physical store is located far from home. We then show how to segment consumers and target them based on these costs. Methodologically speaking, we demonstrate how the “plausibly exogenous” approach of Conley et al. (2010) can be applied to a marketing context in the presence of a nonlinear and hierarchical model. We propose a data aggregation strategy that helps reduce the model complexity substantially with a limited loss of relevant information. We also highlight the environmental aspect of grocery shopping activities and hope this can stimulate more discussion among marketers regarding the effects of shopping on the environment.

Although we have quantified transaction costs for the retailer/market of interest, a natural question is whether the results have “face validity.” To assess this, we looked at the estimates in Bell et al. (1998). We find that the panel households in their data value one mile of travel at \$1.40 or €0.62/km compared with our estimate of €0.59/km. So even though our data are for a different country and in a different context, the results appear to be in the same “ball park.” This gives us some confidence in our results. At the same time, we are able to quantify other transactions costs as well, which is a unique contribution of our study.

Because the online and off-line channels have the same prices, we did not have to incorporate direct costs in the model. However, it is straightforward to include these costs if the online and off-line channels

have different prices. A limitation of our analysis is that we do not have competing retailers’ information, but our research framework can accommodate that case as well. Furthermore, an advantage of our data is that we did not need to look at the effect of channel choice on purchase quantities, and we were thus able to summarize category information via the numbers of items. In the presence of data that do not have this feature, extending our framework would be a useful endeavor. To the extent that previous research in the grocery context has looked largely at the choice of stores across chains, our research can be seen as complementing the literature.

Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mktsci.journal.informs.org/>.

Acknowledgments

This project is partially supported by the Singapore Ministry of Education Research Project R-316-000-073-112. The authors are very grateful to a Spanish grocery chain for providing us with the data for this study and to the Kilts Center at the Booth School of Business, University of Chicago, for financial assistance. The authors thank the editor in chief, the area editor, and two anonymous reviewers for their constructive comments and suggestions. The usual disclaimer applies.

References

- Balasubramanian, S. 1998. Mail versus mall: A strategic analysis of competition between direct marketers and conventional retailers. *Marketing Sci.* 17(3) 181–195.
- Becker, G. S. 1965. A theory of the allocation of time. *Econom. J.* 75(299) 493–517.
- Bell, D., T. H. Ho, C. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *J. Marketing Res.* 35(3) 352–369.
- Betancourt, R. R. 2005. *The Economics of Retailing and Distribution*. Edward Elgar Publishing, Cheltenham, UK.
- Betancourt, R. R., D. A. Gautschi. 1986. The evolution of retailing: A suggested economic interpretation. *Internat. J. Res. Marketing* 3(4) 217–232.
- Betancourt, R. R., D. A. Gautschi. 1988. The economics of retail firms. *Managerial Decision Econom.* 9(2) 133–144.
- Briesch, R., P. Chintagunta, E. Fox. 2009. How does assortment affect grocery store choice? *J. Marketing Res.* 46(2) 176–189.
- BusinessWeekOnline. 2001. Q&A with Tesco.com’s John Browett. (October 1), http://www.businessweek.com/magazine/content/01_40/b3751626.htm.
- Chib, S. 2001. Markov chain Monte Carlo methods: Computation and inference. J. J. Heckman, E. E. Leamer, eds. *Handbook of Econometrics*, Vol. 5. North-Holland, Amsterdam, 3569–3649.
- Chintagunta, P., J. Chu. 2011. Category selection and store choice modeling. Working paper, Booth School of Business, University of Chicago, Chicago.
- Chircu, A. M., V. Mahajan. 2006. Managing electronic commerce retail transaction costs for customer value. *Decision Support Systems* 42(2) 898–914.

- Chu, J., P. Chintagunta, J. Cebollada. 2008. A comparison of within-household price sensitivity across online and off-line channels. *Marketing Sci.* **27**(2) 283–299.
- Chu, J., M. Arce-Urriza, J.-J. Cebollada-Calvo, P. K. Chintagunta. 2010. An empirical analysis of shopping behavior across online and offline channels for grocery products: The moderating effects of household and product characteristics. *J. Interactive Marketing* **24**(4) 251–268.
- Coase, R. H. 1937. The nature of the firm. *Economica* **4**(16) 386–405.
- Conley, T. G., C. B. Hansen, P. E. Rossi. 2010. Plausibly exogenous. *Rev. Econom. Statist.*, ePub ahead of print October 7.
- Cool Green & Shady. 2011. Satisfaction guarantee. Accessed April 13, 2011, <http://www.coolgreenandshady.com/cgsLegalSG.aspx>.
- Demby, E. R. 2000. Pricey and perishable, online supermarkets struggle to deliver on their retail promise. Internet Retailer (December 26), <http://www.internetretailer.com/2000/12/26/pricey-and-perishable-online-supermarkets-struggle-to-deliver-o?p=1>.
- Devaraj, S., M. Fan, R. Kohli. 2002. Antecedents of B2C channel satisfaction and preference: Validating e-commerce metrics. *Inform. Systems Res.* **13**(3) 316–333.
- Economist, The. 2010. Online grocers: Keep on trucking. (July 15), <http://www.economist.com/node/16595179>.
- Forman, C., A. Ghose, A. Goldfarb. 2009. Competition between local and electronic markets: How the benefits of buying online depends on where you live. *Management Sci.* **55**(1) 47–57.
- Fox, E., A. Montgomery, L. Lodish. 2004. Consumer shopping and spending across retail formats. *J. Bus.* **77**(2) S25–S60.
- Furubotn, E. G., R. Richter. 1997. *Institutions and Economic Theory*. University of Michigan Press, Ann Arbor.
- Gupta, S. 1988. Impact of sales promotions as when, what, and how much to buy. *J. Marketing Res.* **25**(4) 342–355.
- Hoch, S. J., B. Kim, A. L. Montgomery, P. E. Rossi. 1995. Determinants of store-level price elasticity. *J. Marketing Res.* **32**(1) 17–29.
- Ingene, C. A. 1984. Productivity and functional shifting in spatial retailing. *J. Retailing* **60**(Fall) 15–36.
- Kacen, J. J., J. D. Hess, W.-Y. K. Chiang. 2003. Bricks or clicks? Consumer attitude toward traditional stores and online stores. Working paper, University of Illinois at Urbana-Champaign, Champaign.
- Kahn, B., D. Schmittlein. 1989. Shopping trip behavior: An empirical investigation. *Marketing Lett.* **1**(December) 55–70.
- Keeny, R. L. 1999. The value of Internet commerce to the customer. *Management Sci.* **45**(4) 533–542.
- Lee, S.-Y. T., I. P. L. Png. 2004. Buyer shopping costs and retail pricing: An indirect empirical test. *Rev. Marketing Sci.* **2**(1), Article 6. <http://www.bepress.com/romsjournal/vol2/iss1/art6>.
- Lewis, M., V. Singh, S. Fay. 2006. An empirical study of the impact of nonlinear shipping and handling fees on purchase incidence and expenditure decisions. *Marketing Sci.* **25**(1) 51–64.
- Liang, T.-P., J.-S. Huang. 1998. An empirical study on consumer acceptance of products in electronic markets: A transaction cost model. *Decision Support Systems* **24**(1) 29–43.
- Matas, A., J.-L. Raymond, J.-L. Roig. 2009. Car ownership and access to jobs in Spain. *Transportation Res. Part A* **43**(6) 607–617.
- Messinger, P. R., C. Narasimhan. 1997. A model of retail formats based on consumers' economizing on shopping time. *Marketing Sci.* **16**(1) 1–23.
- Nick. 2007. Why in the world are you not grocery shopping online? *Punny Money* (blog), October 31, <http://www.punny.org/money/why-in-the-world-are-you-not-grocery-shopping-online/>.
- Pashigian, B., S. Peltzman, J. M. Sun. 2003. Firm responses to income inequality and the cost of time. *Rev. Indust. Organ.* **22**(4) 253–273.
- Peapod. 2011. Peapod growing greener. Accessed November 18, <http://www.peapod.com/brownGrowingGreenerDeliveryTimes-Page.jhtml>.
- Petrin, A., K. Train. 2010. A control function approach to endogeneity in consumer choice models. *J. Marketing Res.* **47**(1) 3–13.
- Rossi, P., G. Allenby, R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons, Hoboken, NJ.
- Sinai, T., J. Waldfogel. 2004. Geography and the Internet: Is the Internet a substitute or a complement for cities? *J. Urban Econom.* **56**(1) 1–24.
- Smith, M., E. Brynjolfsson. 2001. Consumer decision making at an Internet shopbot: Brand still matters. *J. Indust. Econom.* **49**(4) 541–558.
- U.S. Environmental Protection Agency. 2005. Emission facts. Report EPA420-F-05-004, USEPA, Washington, DC. <http://www.epa.gov/oms/climate/420f05004.pdf>.
- Williamson, O. E. 1979. Transaction cost economics: The governance of contractual relations. *J. Law Econom.* **22**(2) 233–261.
- Williamson, O. E. 1981. The economics of organization: The transaction cost approach. *Amer. J. Sociol.* **87**(3) 547–577.
- Williamson, O. E., S. E. Masten. 1999. *The Economics of Transaction Costs*. Edward Elgar Publishing, Cheltenham, UK.