Quantifying Transaction Costs in Online / Offline Grocery Channel Choice*

Pradeep Chintagunta, Robert Law Professor of Marketing
University of Chicago Booth School of Business
5807 South Woodlawn Avenue, Chicago, IL 60637-1610
Tel: (773) 702-8015
Pradeep.chintagunta@chicagobooth.edu

Junhong Chu, Assistant Professor of Marketing
NUS Business School
1 Business Link, Singapore 117592
Tel: (65)6516-6938
bizcj@nus.edu.sg

Javier Cebollada, Associate Professor
Public University of Navarra, Pamplona, Spain
Tel: (34) 948169736
cebollada@unavarra.es

* This project is partially supported by the Singapore Ministry of Education Research Project R-317-000-073-113. We 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 paper was previously titled as “What Drives Household Choice of Online and Offline Grocery Channels?”.
Quantifying Transaction Costs in Online / Offline Grocery Channel Choice

ABSTRACT

Households incur a number of transaction costs when choosing stores to make grocery purchases. When the online channel is available as an alternative to physical-store shopping, they may need to incur additional transaction costs. In this paper, we empirically quantify relative transaction costs when households choose between the online and offline channels of the same grocery chain. A key challenge to quantifying these costs is that several of them, such as picking items from the store and carrying them home, depend upon the items the household expects to purchase in the store; and unobserved factors that influence channel choice also likely influence the items purchased. Our econometric specification for channel choice accounts for observed and unobserved household heterogeneity, as well as the endogeneity of the items purchased via the “plausibly exogenous” approach in a hierarchical Bayesian framework. We find on average, the value to a household of avoiding 1 kilometer of travel and ordering online instead is €.47; the relative value of shopping online as opposed to offline on a weekday (in bad weather) compared to a weekend (good weather day) is €1.26 (€1.06). For every 10 items bought, the time savings in the online channel over the offline channel are equivalent to €.55; the costs of picking and putting 10 heavy/ bulky items into the shopping cart are about €.42; and the costs of carrying 10 heavy/ bulky items one kilometer are about €.83. On their online visits, households value the net transaction costs avoided by online shopping at €10.92, which exceeds the retailer’s delivery charges. We find considerable heterogeneity in these costs across households and characterize their distributions. We discuss the implications of our findings for the retailer in terms of product offerings, promotion and positioning for the two channels.

Keywords: Channel Choice, Online Grocery Shopping, Transaction Costs, Plausibly Exogenous, Hierarchical Bayesian
I. Introduction

Store choice for grocery products has been an enduring research topic for marketers. A majority of recent research in this area has focused on understanding the tradeoffs made by consumers between price, convenience and assortment. Researchers identify, characterize and quantify a household’s tradeoffs among the various stores along these three store choice drivers for a given visit or a given basket of items. Much of this research has looked at household choices among various supermarket chains (Briesch, Chintagunta and Fox 2009) or among chains of different formats such as EDLP vs. Hi-Lo stores (Bell and Lattin 1998); supermarkets vs. convenience stores vs. warehouse clubs (Fox, Montgomery and Lodish 2004), etc.

In markets where an internet grocery channel is available, one must consider a household’s tradeoffs between price, convenience and assortment for this channel as well. Prima facie, one can argue that the availability of the online channel should not fundamentally alter the kinds of tradeoffs necessary for store choice. Our objective in this paper is to first highlight some of the additional costs and benefits that need to be considered when looking at store choice that includes the internet option that might not be relevant in the absence of this option. More importantly, we then quantify and provide money metrics for the various elements of the tradeoffs involved when choosing between the online and offline grocery channels. As with the previous literature (e.g., Bell et al 1998), we focus on store choice conditional on making a store visit; we use “store choice” and “channel choice” interchangeably to mean choice of an offline store or an online store. Our discussion of money metrics therefore has to be construed as referring to relative metrics between shopping online and offline.

Online and offline channels provide varying levels of convenience and distribution services that entail different levels of direct and transaction costs. Direct costs refer to the sum of shelf
prices (less discounts) of the items in the shopping basket. Several types of transaction costs have been identified in the context of grocery shopping (see Betancourt 2005, Bell, Ho and Tang 1998, Lewis, Singh and Fay 2006): (1) opportunity costs of time - travel time to and from a store and in-store shopping time; (2) transportation costs to and from a store; (3) psychic costs associated with potentially undesirable characteristics of the retail environment; (4) adjustment costs due to product unavailability; (5) search costs associated with assortment differences across stores; (6) basket delivery costs; (7) physical costs such as costs of picking and putting items into the shopping cart and costs of carrying the basket home; and (8) other costs such as waiting costs for basket delivery and the inability to check product quality prior to paying. Costs (1) ~ (5) are relevant when considering store choice among offline stores. When the internet option is present, one also needs to factor costs (6) ~ (8) into the analysis.

Direct costs and transaction costs are a function of the shopped channel, the basket purchased, the duration and timing of the trip, and household characteristics. Hence, these costs vary across trips for the same household depending on when, what and how much it buys. For each shopping occasion, a household trades off these various costs and chooses the store-channel with the lowest costs.1 Direct costs and transaction costs for the same product bundle also differ by households, depending on their sensitivities to different cost components and valuation of convenience and time. Consequently, we would expect to see both within- and across-household variations in store choice. Traditionally, research that has included the internet channel in the analysis, i.e., in non-grocery contexts, has viewed the online channel as a mechanism to sort heterogeneous consumers (e.g. Goolsbee 2001 for online and offline PC purchases). A majority of these studies has been with cross-sectional and survey data. By contrast, we find that consumers “switch” between online and offline channels when buying groceries. This allows us

---

1 In principle, households could also account for future considerations, which are beyond the scope of this paper.
to quantify the impact of time-varying factors on the choice between online and offline channels.

The empirical context for quantifying the effects of the above costs on channel choice is a unique household panel dataset from Spain. The data consist of the same households choosing between shopping at a chain’s offline stores and its online store (which happens to be the biggest “store” in the chain) over a 6-month period. Using data from a single retail chain can have the disadvantage of not capturing all the purchases by a panelist. This concern, while legitimate, is mitigated in our case as the panelists’ annual expenditures in the focal chain are about 80% of grocery expenditures by households of similar socio-demographic characteristics in the same area. Meanwhile, focusing on purchases within a chain has several benefits. First, issues such as store image as a choice driver are no longer relevant. Second, the chain has the same prices and promotions online and offline. Thus, direct costs are identical across channels and will drop out from any discrete store choice model. Third, similar assortments are available everywhere so this factor will not play a role either. Taken together, our empirical context allows us to focus on the key cost factors that differ across online and offline channels and quantifying these factors.

Nevertheless, empirically quantifying the above costs is a challenging task because several of them, e.g., costs of putting items into the shopping cart and carrying them home, which are important in a market where some grocery stores have no parking lots and many people walk or take public transport to the store, depend on the specific categories and the amounts in these categories that the household expects to buy on that shopping trip. However, these factors are not exogenous since unobservable factors that influence the categories and quantities bought will also likely influence the choice of shopping channels (e.g., at the margin, a household shopping online might purchase an additional category and / or a larger pack size). Hence, the endogeneity of these channel choice drivers need to be accounted for. One way to deal with such 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). 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, i.e., shopping online does not encourage a household to buy a larger pack size than it would have bought at the offline store. 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 $87.0\% \sim 99.8\%$ 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 non-trivial since there is considerable variation across households in the categories that drive their basket expenditures. For example, only 15 categories overlap among each household’s top 30 categories. Thus, we cannot fix the set of categories to be analyzed as being common across households as in previous studies, nor can we focus on a smaller subset of categories since to account for at least 80% of a household’s basket expenditures we need to include 30 categories. Since our interest in the categories purchased stems from how they could influence transaction costs and since the choice of the channel does not affect quantities purchased, the identities of the specific categories purchased is not critical (other than for computing inventory levels). Thus, we can simply summarize the information contained in the expected categories purchased via metrics that will likely influence transaction costs and are also common across households. Specifically, we focus on the total number of items, the number of perishable items and the number of heavy / bulky items (defined later) that a household expects to purchase. One benefit

---

2 As online and offline channels have identical assortments, brand choice is unlikely correlated with channel choice.
3 This does not preclude the possibility that certain households systematically purchase large pack sizes than others. This will influence transaction costs. As discussed later, we account for this in our analysis.
of this classification is that transaction costs can be specified as functions of these numbers of items. 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 purchase.

Given the above set of 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 perspective where the numbers of items are instrumented for in the analysis of channel choice behavior. We choose the latter approach since this 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.

What would be reasonable instruments for the numbers of items purchased? A key benefit of our data is that we have access to purchases in all product categories, including meats, fish, milk, and produce that make up a significant proportion of household market basket. This allows us, for each household, to use that household’s top-expenditure categories to construct its inventory variables. Importantly, these inventory variables do not directly influence channel choice and only do so via the numbers of items bought, i.e., lower inventory levels in many categories will result in greater numbers of items to be bought on that trip, which in turn increases, say, the costs of carrying the basket home. In addition, we also have information on the numbers of items bought by households that have similar demographics to the panelists but do not make a channel choice decision, i.e., “pure” online and offline households. Together with the inventory variables, we have information that is likely correlated with the numbers of items bought by a household, but unlikely correlated with that household’s online/offline choice. Finally, there are exogenous
factors such as whether shopping on a weekday or bad weather day that likely influence channel
choice, but not directly influence the numbers of items bought. We combine our instruments with
Conley et al.’s (2008) plausibly exogenous approach in a Hierarchical Bayesian (HB) framework
to account for the endogeneity of the numbers of items purchased.

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. On average, we find that the value of avoiding 1
kilometer of travel and ordering online instead is €.47; the relative value of shopping online as
opposed to offline on a weekday (in bad weather) compared to a weekend (good weather day) is
€1.26 (€1.06). For every 10 items bought, the time savings in the online channel over the offline
channel are equivalent to €.55; the costs of picking and putting 10 heavy/ bulky items into the
shopping cart are about €.42; and the costs of carrying 10 heavy/ bulky items one kilometer are
about €.83. On their online visits, households value the net transaction costs avoided by online
shopping at €10.92, on average, which exceeds the retailer’s delivery fees. We find considerable
heterogeneity in these costs across households. Based on our findings we provide implications
for the retailer in terms of product offerings, promotion and positioning for the two channels.

Our study can be seen as building on and complementing previous research such as Bell el al
(1998), Fox et al (2004) and Briesch et al (2009). In particular, we extend those studies to look at
the presence of the online channel and its associated transaction costs. Further, instead of
modeling store choice, purchase incidence and brand choice we show how focusing on channel
choice combined with a limited information approach allows us to quantify transaction costs.

The rest of the paper proceeds as follows. We describe the data in §2 and the econometric
model in §3. We detail variable operationalization in §4, report main results in §5 and conclude
with managerial implications in §6.

2. Data and Observed Shopping Patterns

2.1 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 offline stores and one 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 chain uses a centralized online ordering system. After receiving an online order, it assigns the order to one of the partner stores and notifies the household. The household then has two options: it can either go to the assigned partner store to pick up the order for free (not observed in the data) or have the basket delivered home within some chosen delivery time window (e.g. 7-9pm Monday) and incur delivery charges.

The retailer practices uniform pricing so prices are identical in all online and offline stores. It is a Hi-Lo chain that runs chain-wide promotions. The online product offerings are the same for everyone and available in all partner stores. People often walk or take public transport to buy groceries (only 68% of stores have parking lots). About 60% offline stores also offer delivery service. The retailer has the same delivery policy for online and offline 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 clients if the basket exceeds €100. Households with quarterly in-chain expenditure over €600 are eligible but need to apply for golden-card membership. In sum, online and offline channels have the same prices, price promotions, delivery and roughly same assortments.

2.2 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 3 to 8 weeks. 3-week, 4-
week and 5-week promotions account for 22.9%, 43.2% and 33.7% 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%.

2.3 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 offline channels. We observe when a household visited the chain, which store it visited, what items and how much it bought, whether the basket was home delivered, and delivery fees. 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, .68 preschool and .47 schoolchildren, 2.14 work-age adults, and .07 elders. 15.32%, 26.34% and 58.34% households live respectively in low, medium and high income / economic areas. We also obtained data on pure online and pure offline customers that we use to construct instruments.

2.4 Observed shopping patterns and implications for econometric modeling: The households exhibit the following shopping patterns (Table 2) that appear to be consistent with the transaction costs of shopping at each channel.

First, the availability of the Internet option does not act as a sorting mechanism whereby certain households mainly shop online and others primarily shop offline. All households shop in both channels. There is not the “20/80” phenomenon that 20% households account for 80% online trips or expenditure (Figure 1). The mean ratio of online trips to total trips is 41.8% (sd = 24.7%). The ratio is below 20% for 22% of households, above 80% for 11% of households, and between 20%-80% for 67% of households. The mean ratio of online expenditures to total
expenditures is 65.0% (sd = 24.9%). The ratio is below 20% for 7% households, between 20%-80% for 58% of households and above 80% for 35% of households. Channel “switching” is common in the data. The online-to-offline switching probability is 56.5% and the offline-to-online probability is 29.6%. This suggests that for the same households, transaction costs of buying online vs. offline vary from trip to trip. 22% online trips and 1% offline trips involve delivery charges, implying households derive less disutility from delivery fees on an online visit than an offline one. *Since 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 online shopping intensity across households.*

Second, households sort their shopping trips to online and offline channels based on basket size. They make major trips to the online channel and fill-in trips to the offline channel (Kahn & Schmittlein 1989). Online baskets (€155.8) on average are 3.5 times as large as offline baskets (€44.9). Households buy 28 (11) unique categories and 38 (14) unique items on an online (offline) trip. 95% of a household’s online trips have a basket larger than its mean offline basket, while 63% of its offline trips have below-average baskets. This suggests transaction costs of shopping online are low when the basket is large, but high when the basket is small. The opposite is true for shopping offline. *It also highlights the importance of accounting for unobservable factors that influence both channel choice and number of items purchased.*

Third, households also sort their trips to online and offline channels based on category characteristics. Online baskets consist of more heavy/bulky items, reflecting the convenience benefit of online shopping, while offline 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 CPGs (mineral water, beer, toilet paper) and liquid-rich non-CPGs (watermelon). Examples of these are 4 1-liter bottles of water, 12 rolls of toilet tissue, etc. Note that some categories can appear in both sets (e.g., watermelon).\textsuperscript{4} An online (offline) 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 offline trip expenditures. Further, a household on average bought 29.3 categories exclusively online, 32.4 categories exclusively offline, and 23.2 categories in both channels. The channel-specific category purchases suggest that the transaction costs of buying the same categories differ by channel. The transaction costs of buying heavy/bulky items are lower online, but higher offline. The opposite holds for buying perishables. It appears that types of categories play a role in channel choice 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 slightly different impacts on online and offline baskets. Figure 2 shows the distribution of shopping trips by basket size. The threshold does not affect offline trip expenditures at all because offline baskets do not peak above €100 and only 11.3% offline trips exceed €100. The threshold seems to boost online trip expenditures, but not too much, because 8% online trips are above €90 but below €100, and 70% online trips are larger than €110.

Fifth, there is considerable household heterogeneity in online shopping intensity. We regress the proportion of online trips on household demographics and characteristics of their most frequented offline stores. Downtown households are less likely to shop online. Households are less likely to shop online if their most frequented offline stores have larger assortments or provide parking facilities. If a household’s average basket has more heavy items, it is more likely to shop online; if its average basket has more perishables, it is more likely to shop offline.

\textsuperscript{4} We undertook extensive sensitivity analyses to group membership, results of which are available from the authors.
Sixth, there is a much less variation in online shopping than offline shopping both in basket size and trip interval. The coefficient of variation is .24 (sd = .16) for online baskets and .73 (sd = .32) for offline baskets; .52 (sd = .25) for online-online trip intervals and .86 (sd = .32) for offline-offline 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 offline 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 3.6% items bought online and 4.2% items bought offline. The figures are 4.9% and 3.9% for heavy items, and 1.2% and 1.7% for perishables. So we need to account for the differential effects of promotions across channels.

Seventh, there is not much evidence of household learning of the Internet as a grocery-shopping channel (Goettler and Clay 2010). The online basket size and basket shares of perishables and heavy/bulky items do not vary significantly from trip to trip. The online store was opened in 2001 and the data were collected in 2007. The sample households were not new customers to the chain, so channel preferences likely have reached steady state. This means that dynamic learning is less of an issue for our analysis.

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 analysis.

3. Econometric Specification

3.1 Description of shopping costs: A shopping trip involves direct costs and transaction costs. Direct costs are similar to the variable costs in Bell et al (1998). 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 the eight types of transaction costs identified previously. For a given basket bundle, search costs to find a store are not likely to play a role in our case, although in-store search costs to find the items to buy will differ, which are accounted for below.

(1) *Time costs* are the opportunity costs of travel time and in-store shopping time. Shopping time is an important factor in the store choice decision. Messinger and Narasimhan (1997) find that the growth of one-stop shopping has been a response to the growing demand for time-saving convenience. Pashigian, Peltzman and Sun (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.

*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 offline. For the same set of items, a household may need less time to shop online than offline, so we allow the marginal cost 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.

(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\(^5\).

(3) *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 offline shopping. We use the number of heavy/bulky items as a proxy for item-picking cost and allow the marginal cost to differ by channel. The costs of carrying a basket home are a

---

\(^5\) Travel time and transportation costs also depend on the means of transport. As we do not know the means of transport for, we use store distance as a proxy. Further, if we had information on the available public transportation in the panelists’ neighborhoods, we could include it to allow for heterogeneity across panelists in these costs.
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 it to zero for online shopping and offline 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 offline channels differently.

(4) **Basket delivery costs**: Smith and Brynjolfsson (2001) find consumers are more sensitive to delivery fees 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 online groceries. The shadow price of delivery charges differs for online and offline shopping. For online shopping, delivery charges cover the transportation costs to and from the store, the opportunity costs of travel time, and the physical costs of basket picking and carrying. For offline shopping, they only cover the physical costs of basket carrying. Thus, households may be more sensitive to delivery charges when shopping offline than when shopping online.

(5) **Costs of inability to verify product quality prior to paying** are primarily a function of perishables bought because product quality like freshness can vary a lot from purchase occasion to purchase occasion, while 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, while the online store lacks this ability because the actual quality can differ from what is shown on the computer screen. Therefore, households are more likely to visit the offline channel when buying more perishables.

(6) **Psychic costs** are difficult to measure, but are closely related to the characteristics of the shopping channels themselves. We use channel dummies to capture these costs.

(7) **Adjustment costs due to product unavailability, costs of waiting for basket delivery and in-store search costs**. As with psychic costs, these costs are channel specific and difficult to
measure. And they may vary over time because of different products not available in different channels at different times, instances of more urgent need of a product than others and familiarity with product layouts. We represent them by a household, trip and channel specific error term.

As noted previously, our data indicate that there is a greater variability in offline basket sizes, possibly due to the effect of promotions to differentially draw consumers into the offline stores. So while the financial consequences of promotions, conditional on a purchase in a category, are identical across channels and drop out of the channel choice decision, promotions themselves could be more likely to drive households to the offline channel. Further, promotions of heavy/bulky items and perishables may have different effects on visiting the two channels, so we include the effects of promotions and those of heavy/bulky items and perishables in the analysis.

3.2 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 fees, and travel distance. Given chain-wide promotions and the availability of the Internet option, it is also reasonable to assume rational expectations over promotions. Based on our discussion of the data characteristics in the previous section and the various transactions costs above, household $h$’s expected costs of shopping at channel $j$ on trip $t$ are:

$$ FC_{hjt} = \alpha_{hj0} + \alpha_{hj1}d_{hjt} + \alpha_{hj2}NI_{ht} + \alpha_{hj3}WK_t + \alpha_{hj4}NH_{ht} + \alpha_{hj5}NH_{ht} * d_{hjt} + \alpha_{hj6}W_t + \alpha_{hj7}DC_{ht} + \alpha_{hj8}NP_{ht} + \alpha_{hj9}P_{ht} + \alpha_{hj10}PP_{ht} + \alpha_{hj11}PH_{ht} + \epsilon_{hjt} $$ (1)

Where $d_{hjt}$ is the home-store distance, which changes from trip to trip because a household does not always visit the same offline store. $WK_t$ is a weekday dummy, $W_t$ is a bad weather dummy, and $DC_{ht}$ are the delivery charges. $NI_{ht}$, $NH_{ht}$ and $NP_{ht}$ are numbers of items, heavy/
bulky items and perishables. For identification purposes, we constrain the coefficients for all offline stores to be the same and subtract the costs of shopping at the offline channel from those at the online channel. We drop subscript $t$ for ease of exposition. Define $Y_h \equiv FC_{h, on} - FC_{h, off}$, $\varepsilon_h \equiv \varepsilon_{h, on} - \varepsilon_{h, off}$, and $\alpha_h \equiv \alpha_{h, on} - \alpha_{h, off}$. Let $X_h \equiv \{NI_{ht}, NP_{ht}, NH_{ht}, NH_{ht}*d_{ht}\}$ 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 $I_h$ the choice indicator. $\beta_h$ represents the subset of $\alpha_h$ that is associated with the variables $X_h$ and $\gamma_{1h}$ represents all the other coefficients in $\alpha_h$. Thus, we have the following relationships:

\[ Y_h = G_h\gamma_{1h} + X_h\beta_h + \varepsilon_h \]  \hspace{1cm} (2)

\[ I_h | Y_h \sim Binomial Probit \]  \hspace{1cm} (3)

3.3 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 unobservable factors that influence numbers of items bought also likely influence the choice of the shopping channel. 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 (Petrin and Train 2010); and (iii) the “plausibly exogenous” approach of Conley, Hansen and Rossi (2008). These approaches are similar in nature. (ii) and (iii) rely on finding exogenous instruments for the endogenous variables, and (i) requires the presence of

---

6 As noted before, households could systematically differ in their package size choices even within a particular group of items. To account for this, we also tried a different operationalization of the numbers of items by weighting the numbers by pack sizes relative to the smallest size in each group. This did not change our results materially. We also include family size as observable heterogeneity in the effects of the different numbers of items variables.
some excluded variables.\footnote{Other approaches such as the Latent Instrumental Variables (LIV) method are also available (Ebbes et al 2005).}

The plausibly exogenous approach 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 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 in Equation (2):

$$Y_h = G_h \gamma_{1h} + X_h \beta_h + Z_h \gamma + \epsilon_h$$

$$X_h = (G_h Z_h) \Pi + \nu_h$$

This approach essentially replaces the actual numbers of items, perishables and heavy/bulky items with their expected numbers, which are determined by exogenous instruments. Assume $(\epsilon_h, \nu_h) \sim N(0, \Sigma)$. Let $p(Data | \gamma_{1h}, \beta_h, \gamma, \Pi, \Sigma)$ 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 $\gamma$. The inference is based on the posterior distribution of $(\gamma_{1h}, \beta_h, \Pi, \Sigma)$, integrating out $\gamma$:

$$p(\beta_h, \gamma_{1h}, \Pi, \Sigma | Data) \propto p(Data | \beta_h, \gamma_{1h}, \gamma, \Pi, \Sigma) p(\gamma | \beta_h, \gamma_{1h}, \Pi, \Sigma) p(\beta_h, \gamma_{1h}, \Pi, \Sigma) d\gamma$$

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

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

where $D$ is household demographics and $\Delta$ is the 2nd-stage coefficients. The common priors are:

$$\gamma \sim N(\mu_\gamma, A_\gamma^{-1}), \quad \text{vec}(\Pi | V_\pi) \sim N(\text{vec}(\Pi), \Sigma_{22} \otimes A_{\Pi}^{-1})$$

$$\Sigma \sim IW(v_\Sigma, V_\Sigma), \quad \text{vec}(\Delta | V_\Delta) \sim N(\text{vec}(\Delta), V_\omega \otimes A_\Delta^{-1})$$

$$V_\omega \sim IW(v_{\omega 0}, V_{\omega 0})$$

We write out this model as a sequence of conditional distributions and run the Gibbs sampler to get the MCMC sequence (Chib 2001, Rossi, Allenby and McCulloch 2005, Conley et al 2008). The posterior distributions are (see the Appendix for details):
4. Variable Operationalization and Robustness Checks

4.1 Choice of instruments: To resolve the endogeneity issue, we need to find variables that affect the numbers of items, perishable or heavy/bulky items a household buys on trip \( t \), but do not influence the choice of the shopping channel. Inventory variables are valid instruments because they do not directly influence channel choice and only do so via the numbers of items bought, i.e., lower inventory levels in many categories will result in greater numbers of items to be bought on that trip. 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 offline households. We use a household’s numbers of low-inventory items, low-inventory perishables and low-inventory heavy/bulky items as instruments. We first identify the top 30 categories for each household by looking at the proportion of that household’s total grocery expenditure (across purchase occasions) that is accounted for by each category purchased. Since household-specific top 30 account for 81.1% of the basket value (sd = 9.4%), using the top 30 is nearly equivalent to using the entire basket. 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 offline households. We compute trip-level mean numbers of items (heavy/bulky items, 

---

\[ Y_h \mid G_h, X_h, Z_h, \gamma_{th}, \beta_h, \gamma, \Sigma \sim \text{Truncated Normal} \]
\[ \{\gamma_{th}, \beta_h\} \mid Y_h, G_h, X_h, Z_h, D, \gamma, \Pi, \Sigma, \Delta, V_\omega \sim \text{Normal} \]
\[ \gamma \mid Y_h, G_h, X_h, Z_h, \{\gamma_{th}, \beta_h\}, \Pi, \Sigma, \mu_\gamma, A_\gamma \sim \text{Normal} \]
\[ \Pi \mid G_h, X_h, Z_h, \{\gamma_{th}, \beta_h\}, \Sigma, \Pi_0, A_{\Pi_0} \sim \text{Normal} \]
\[ \Sigma \mid Y_h, G_h, X_h, Z_h, \{\gamma_{th}, \beta_h\}, \gamma, \Pi \sim \text{Inverted Wishart} \]
\[ \Delta \mid \{\gamma_{th}, \beta_h\}, V_\omega, \Delta_0, A_\Delta \sim \text{Normal} \]
\[ V_\omega \mid \{\gamma_{th}, \beta_h\}, \omega_0, V_\omega_0 \sim \text{Inverted Wishart} \]
perishables) on each day for pure online and pure offline shoppers, broken down by observed demographics and match them to each sample household. The rationale is that variations 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; however since these other households do not make a channel choice, their numbers of items are unlikely to be correlated with the unobservables that drive channel choice for our panelist households. These instruments, together with exogenous variables in Equation (1) and household characteristics, explain 22.2%, 24.1% and 28.1% of the variation in the endogenous variables (table 3). We checked the sensitivity of our results to including both sets of instruments or using one set (inventory or other households’ numbers) at a time. We find the qualitative nature of results to be consistent although there is some variation in the actual point estimates.

4.2 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.

We also tried common (across households) promotion indices, computed as the weighted average of category promotion depths with the rationale that promotions in general bring consumers to stores. The weights are retail sales of the top 30 categories that account for the
largest expenditure shares across all households. We find that household-specific promotion indices result in smaller standard errors although the magnitudes of the effects are similar.

4.3 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 offline option is the value for that household’s most frequented physical store. The distance is zero for online shopping.

The delivery charges variable ($DC_{ht}$) is defined as follows. For all offline visits entailing delivery, the variable drops out since the charges are the same in both cases. For offline visits with no delivery, $DC_{ht} = 0$ for the offline 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 offline option corresponds to no delivery, so $DC_{ht} = 0$ for the offline channel and $DC_{ht} =$ actual charges for the online channel. We recognize that this is an approximation in that, to deal with this structurally, we need to consider offline 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 offline visits as a single choice. Further, as noted before a very small fraction of offline shoppers avail of the delivery option.

4.4 Demographics and store characteristics: These variables allow us to account for observable heterogeneity across households and they appear in the hierarchy as $D$ in equation (7).

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 (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 partnership.

5. Main Findings

We estimate the model parameters via the MCMC approach. We use diffuse priors to let data dictate the estimates. We take 100,000 draws, keep every 10th and use the first 20,000 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. The results in Table 4 represent the effects of the variables on the choice of the online channel relative to the offline store.

5.1 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 offline 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 offline 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 households are usually more time constrained and have higher opportunity costs of time during weekdays than on weekends.

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 offline 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 offline stores. Households are more likely to shop online during bad weather days, as it is 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 offline channel in order 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 offline channel, but 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 offline channel when they need to buy more perishables.

The channel-specific effects of buying perishables offline 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 tradeoffs in transaction costs that households have to make when shopping at the two channels.

Delivery charges. Even though the benefit of delivery charge for an online order is bigger than an offline order, delivery charges have a statistically significant effect in discouraging households from visiting the online channel.
**Effects of price promotions.** Overall, promotions are more likely to drive households to the offline channel than the online channel. Promotions of different categories have channel-specific traffic building effects. Compared to the offline option, promotions of perishables discourage households from visiting the online channel, but the effect is only marginally significant. Promotions of heavy/bulky items significantly increase online channel visits. The channel-specific effects of promotions as well as promotions of perishables and heavy/bulky items reflect the specific transaction costs associated with shopping at online and offline channels.

**5.2 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 1% (Table 4), indicating the importance of accounting for unobserved household heterogeneity.

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 make intuitive sense. For example, households in the wealthy areas have a higher intrinsic preference for the online channel than those in the less wealthy area. Households in the suburbs have a higher preference for the online channel than those downtown because of the more scattered locations of physical stores in the suburbs. Households have a lower preference for the online channel if their most frequented offline stores sell fresh seafood because they prefer the freshness physical stores can provide with. Households that tend to buy more perishables have a stronger preference for the offline channel than for the online channel, while households that tend to buy more heavy/bulky items have a stronger preference for the online channel than for the offline channel.

Price promotions will encourage households to visit the offline stores when they tend to buy
more perishables. Promotions of perishables more likely encourage households to visit the online channel when their most frequented offline stores sell fresh seafood, while promotions of heavy/bulky items more likely encourage households to shop online when their average baskets consist of more heavy/bulky items. When a household’s average basket has more perishables (heavy/bulky items), it is more (less) sensitive to online delivery charges. Households residing in the wealthy areas are more likely to shop online during weekdays and less sensitive to online delivery fees because of their higher opportunity costs of time.

5.3 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 the delivery fee, which is interpreted as the marginal utility of income (see for example Smith and Brynjolfsson 2001). Note that these monetary values are relative to the option of shopping offline and cannot be interpreted directly as a household’s willingness-to-pay since for that we need to include an outside option. For example, the coefficient of store distance in Table 4 is .580 and the delivery cost coefficient is 1.229. Therefore, on average, relative to shopping offline, shopping online provides a value of $(€.580/1.229) = €.47 / km$ that the household has to travel. On average, it is worth €1.26 (€1.06) to a household for shopping online as opposed to offline on a weekday (bad weather day) compared to a weekend (good weather day). For every 10 items bought, the time savings in the online channel over the offline option is equivalent to €.55; the costs of picking and putting 10 heavy/bulky items into the shopping cart are €.42; and the costs of carrying 10 heavy/bulky items

---

9 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. Since direct costs drop out of the model, the coefficient of delivery fee represents the price coefficient. If consumers are more sensitive to delivery charges than to product prices as in Brynjolfsson and Smith (2000), 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 since it does not serve the role of price in the model.
one kilometer are €.83 (in addition to the €.47 above). The cost a household attaches to not being able to check product quality prior to paying in the online store is €.28 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 show the estimated empirical distributions of these costs in Figures 3 ~ 5. Figure 3 refers to 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 1-kilometer without any heavy/bulky items. A majority of households values weekday online shopping by €.5~€2; 13.3% value it at €2 to €5 and 3.4% value it at more than €5. About 16% households are not bothered by bad weather, while 18.5% households attach more than €2 to online shopping in bad weather relative to the alternative. Some households do not seem to have any disutility of travel, while 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 average value of the online channel relative to the offline channel for each store visit made 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 €10.92 (sd = €6.21). Figure 6 shows the distribution of the relative value across households. 59% of households value the internet option at €7-12 per online trip, which more than offsets the delivery fees for online shopping. Only 2.7% households have a value lower than €6, the delivery charges for baskets smaller than €100. This implies that the delivery charge the retailer has set is reasonable for the value it provides relative to the offline channel 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 channel be if an offline trip occurred online? We find that 83.42% of these trips had a value at €1-3 and 13.19% had a value at €3-4. That is, for the great majority of offline trips, the value provided by the Internet channel does not cover the delivery charges. This explains why the households did not visit the online channel for these trips. This is also consistent with the observed sorting strategy households employ when deciding between the two channels on each shopping trip.

6. Discussion and Managerial Implications

Our findings have important implications for managerial practice. We explore some of them below. Since we do not incorporate competitors’ information due to lack of data, it is important to interpret these results as directional and suggestive.

6.1 Enterprise design for online and offline grocery stores: There is a growing trend for conventional grocery retailers (e.g. Safeway) to start online operations (www.safeway.com). Our study provides some guidance on product offerings for online and offline channels. In particular, our findings have implications for the positioning of the online channel. Peapod’s slogan is “smart shopping for busy people”. 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”. 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. 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. Actually, some online grocers have adopted this strategy. For example, Peapod.com advertises “You order online. We shop. We deliver. Order heavy items.”

6.2 Encouraging dual channel shopping among single-channel customers: Of the retailer’s customers, dual-channel shoppers (that are the focus of our analysis) on average spend €1,343 (sd = €875) in the chain, which is nearly twice as much as the spending by single-channel customers – €734 (sd = €725) for pure online shoppers and €771 (sd = €799) for pure offline shoppers. Other studies (e.g. Ansari et al 2008) also find that multi-channel customers spend significantly more than single-channel customers. Further, given their low in-chain expenditure, single-channel shoppers probably also shop at competing retailers. Therefore, it will be profitable for the retailer to encourage pure online shoppers to visit its physical stores and pure offline customers to also shop online. Here, we quantify the benefits of cross-channel shopping for the chain’s single channel shoppers by using our estimation results for the dual-channel shoppers discussed above. We recognize that there are unobserved sources of heterogeneity that could distinguish the single channel shoppers from their dual-channel counterparts. We try to mitigate this concern by using the average transaction cost values of mixed shoppers that have similar characteristics (see Table 5) to each pure online or pure offline shopper (matched in the same fashion as we did to create instruments).
Next, we compute the value of the Internet option (physical channel) to the pure offline (online) shoppers for each offline (online) trip. We find that pure offline shoppers would be better off on 5.81% of their trips if they actually shopped online – the value provided by the Internet option exceeds delivery fees; and pure online shoppers would be better off on 10.12% of their trips if they actually shopped offline – the physical channel provides equivalent value and delivery charges can be avoided. On average, the net incremental value is €9.89 per trip to pure offline shoppers and €3.12 plus avoided delivery fees to pure online shoppers from this “re-allocation” of trips. The retailer can use targeted couponing and channel-specific price cuts to promote dual channel shopping among single-channel customers, explicitly emphasizing the benefits of shopping in the other channel.

6.3 Channel-specific and category-specific promotions: We find that promoting heavy/bulky items and perishables has different traffic building effects for online and offline channels. Retailers can adopt category-specific and channel-specific promotions to manage channel traffic. They can promote perishables in the offline channel to increase offline 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 the grocery store. They could even use this strategy to attract new customers and retain existing customers.

6.4 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 33.46% households that value the Internet over €12 per 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, 21% households have a high disutility of travel and online shopping is a very attractive option for these shoppers. The retailer can use email promotions and targeted couponing to attract these households to the online store.

6.5 Quantifying societal benefits of online shopping: Online grocery shopping is greener way of shopping than driving to shop offline. Let us assume for simplicity that on an offline grocery-shopping visit, the household does not engage in other activities such as going to the post-office, etc. Thus saving an offline 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 offline orders by basket size. If one truckload can fulfill the delivery of 20 online orders, the internet store will reduce offline shopping trips by the magnitude of 70, thereby reducing carbon emissions. The households made 5,703 online trips. Without the internet store, they would have to make 19,960 more offline trips. If half of them were made by driving, 9,980 driving trips would occur, as compared to 285 delivery truckloads. As long as the 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 eco-friendly delivery trucks. For example, Peapod 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.” Given worldwide environmental concerns, 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.

6.6 Summary: The main contribution of this study is in quantifying the transaction costs of offline 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 offline shopping when buying a large number of items, on bad weather days, on weekdays and when the physical store is located far from home; segment consumers and target them based on these costs; and encourage cross-channel shopping of single-channel shoppers. Methodologically, we show how the “plausibly exogenous” approach can be applied in a marketing context in the presence of a nonlinear model.

While 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 €.58/km compared to our estimate of €.47/km. So even though our data are for a different country and a different context, the results appear to be in the same “ballpark.” 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.

As the online and offline 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 offline channels have different prices. A limitation of our analysis is that we do not have competing retailers’ information, but our research framework can easily accommodate that case as well. Further, an advantage of our data is that we did not need to look at the effect of channel choice...
on purchase quantities and 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.

References


Table 1 Household Demographics & Characteristics of Most Frequentened Offline Stores

<table>
<thead>
<tr>
<th></th>
<th>Entire sample</th>
<th></th>
<th>Random sample</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Family size</td>
<td>3.39</td>
<td>2.20</td>
<td>3.37</td>
<td>2.05</td>
</tr>
<tr>
<td># pre-school children</td>
<td>.69</td>
<td>.89</td>
<td>.68</td>
<td>.87</td>
</tr>
<tr>
<td># school-age children</td>
<td>.44</td>
<td>.85</td>
<td>.47</td>
<td>.89</td>
</tr>
<tr>
<td># working adults</td>
<td>2.18</td>
<td>1.89</td>
<td>2.14</td>
<td>1.74</td>
</tr>
<tr>
<td># of elders (65+)</td>
<td>.07</td>
<td>.47</td>
<td>.07</td>
<td>.35</td>
</tr>
<tr>
<td>Distance to most freq</td>
<td>1.13</td>
<td>4.01</td>
<td>.99</td>
<td>2.61</td>
</tr>
<tr>
<td>offline store</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% reside in low economic area</td>
<td>15.43</td>
<td>15.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% reside in medium economic area</td>
<td>24.96</td>
<td>26.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% reside in high economic area</td>
<td>59.61</td>
<td>58.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% downtown</td>
<td>41.46</td>
<td>39.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Characteristics of most frequented offline stores

| Store square footage (m²) | 1,480.61 | 1,102.38 | 1,473.63 | 987.01 |
| % of bazaar section      | 63.86     | 64.39     |          |        |
| % of butcher shop        | 91.68     | 90.83     |          |        |
| % of processed meat      | 93.38     | 92.29     |          |        |
| % of fish shop           | 83.05     | 82.05     |          |        |
| % of having parking lots | 68.03     | 68.68     |          |        |
| % of home delivery       | 60.44     | 60.68     |          |        |

Table 2 Characteristics of Household Shopping Behavior by Channel

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th></th>
<th>Online</th>
<th></th>
<th>Offline</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Total trips</td>
<td>16.18</td>
<td>11.67</td>
<td>5.56</td>
<td>4.22</td>
<td>10.61</td>
<td>10.98</td>
</tr>
<tr>
<td>Mean trip interval (days)</td>
<td>10.97</td>
<td>12.34</td>
<td>14.46</td>
<td>14.29</td>
<td>9.14</td>
<td>10.75</td>
</tr>
<tr>
<td>Basket size (€)</td>
<td>82.99</td>
<td>82.59</td>
<td>155.77</td>
<td>83.48</td>
<td>44.85</td>
<td>50.12</td>
</tr>
<tr>
<td>Half-year spending (€)</td>
<td>1,342.60</td>
<td>874.50</td>
<td>866.56</td>
<td>697.73</td>
<td>476.04</td>
<td>534.10</td>
</tr>
<tr>
<td>No. unique categories per trip</td>
<td>16.86</td>
<td>13.11</td>
<td>27.67</td>
<td>11.00</td>
<td>11.19</td>
<td>10.26</td>
</tr>
<tr>
<td>No. unique perishable categories</td>
<td>4.11</td>
<td>4.82</td>
<td>4.87</td>
<td>5.57</td>
<td>3.72</td>
<td>4.32</td>
</tr>
<tr>
<td>No. unique heavy/bulky categories</td>
<td>5.06</td>
<td>4.79</td>
<td>9.89</td>
<td>3.67</td>
<td>2.54</td>
<td>3.06</td>
</tr>
<tr>
<td>No. unique items per trip</td>
<td>22.61</td>
<td>19.18</td>
<td>38.47</td>
<td>17.26</td>
<td>14.30</td>
<td>14.28</td>
</tr>
<tr>
<td>No. unique perishable items</td>
<td>4.69</td>
<td>5.73</td>
<td>5.55</td>
<td>6.60</td>
<td>4.24</td>
<td>5.15</td>
</tr>
<tr>
<td>No. unique heavy/bulky items</td>
<td>6.64</td>
<td>6.86</td>
<td>13.39</td>
<td>5.98</td>
<td>3.10</td>
<td>4.06</td>
</tr>
<tr>
<td>% of perishables in trip expenditure</td>
<td>23.21</td>
<td>25.34</td>
<td>11.69</td>
<td>13.14</td>
<td>29.24</td>
<td>27.97</td>
</tr>
<tr>
<td>% of heavy/bulky in trip expenditure</td>
<td>33.03</td>
<td>26.14</td>
<td>49.30</td>
<td>17.81</td>
<td>24.50</td>
<td>25.77</td>
</tr>
<tr>
<td>Coefficient of variation: basket</td>
<td>.49</td>
<td>.35</td>
<td>.24</td>
<td>.16</td>
<td>.73</td>
<td>.32</td>
</tr>
<tr>
<td>Coefficient of variation: trip interval</td>
<td>.73</td>
<td>.25</td>
<td>.52</td>
<td>.25</td>
<td>.86</td>
<td>.32</td>
</tr>
<tr>
<td>Channel switching: household level (%)</td>
<td>58.69</td>
<td>33.70</td>
<td>68.92</td>
<td>31.51</td>
<td>48.50</td>
<td>32.72</td>
</tr>
</tbody>
</table>
Table 3  Regression of Endogenous Variables on Instruments

<table>
<thead>
<tr>
<th></th>
<th>No. Items</th>
<th></th>
<th>No. Perishables</th>
<th></th>
<th>No. Heavy/bulky</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std Err</td>
<td>Estimate</td>
<td>Std Err</td>
<td>Estimate</td>
<td>Std Err</td>
</tr>
<tr>
<td>Intercept</td>
<td>-42.031</td>
<td>3.808</td>
<td>-4.279</td>
<td>.685</td>
<td>-37.390</td>
<td>1.945</td>
</tr>
<tr>
<td>Online shoppers’ mean no. items</td>
<td>.095</td>
<td>.020</td>
<td>.013</td>
<td>.004</td>
<td>-.013</td>
<td>.010</td>
</tr>
<tr>
<td>Offline shoppers’ mean no. items</td>
<td>.355</td>
<td>.031</td>
<td>.048</td>
<td>.006</td>
<td>.186</td>
<td>.016</td>
</tr>
<tr>
<td>Online shoppers’ mean no. perishables</td>
<td>.492</td>
<td>.083</td>
<td>.250</td>
<td>.015</td>
<td>.070</td>
<td>.043</td>
</tr>
<tr>
<td>Offline shoppers’ mean no. perishables</td>
<td>-.354</td>
<td>.104</td>
<td>-.007</td>
<td>.019</td>
<td>-.255</td>
<td>.053</td>
</tr>
<tr>
<td>Online shoppers’ mean no. heavy/bulky</td>
<td>.572</td>
<td>.036</td>
<td>-.005</td>
<td>.006</td>
<td>.440</td>
<td>.018</td>
</tr>
<tr>
<td>Offline shoppers’ mean no. heavy/bulky</td>
<td>-.370</td>
<td>.083</td>
<td>-.049</td>
<td>.015</td>
<td>-.220</td>
<td>.042</td>
</tr>
<tr>
<td>Overall promotion</td>
<td>.006</td>
<td>.003</td>
<td>.000</td>
<td>.000</td>
<td>.003</td>
<td>.001</td>
</tr>
<tr>
<td>Promotion of perishables</td>
<td>-.285</td>
<td>1.096</td>
<td>-.253</td>
<td>.197</td>
<td>-1.149</td>
<td>.560</td>
</tr>
<tr>
<td>Promotion of heavy/bulky</td>
<td>1.260</td>
<td>.529</td>
<td>-.018</td>
<td>.095</td>
<td>.629</td>
<td>.270</td>
</tr>
<tr>
<td>No. low-inventory categories</td>
<td>.572</td>
<td>.029</td>
<td>.037</td>
<td>.005</td>
<td>.130</td>
<td>.015</td>
</tr>
<tr>
<td>No. low-inventory perishables</td>
<td>-.058</td>
<td>.095</td>
<td>.354</td>
<td>.017</td>
<td>-.136</td>
<td>.049</td>
</tr>
<tr>
<td>No. low-inventory heavy/bulky</td>
<td>-.055</td>
<td>.045</td>
<td>-.038</td>
<td>.008</td>
<td>.300</td>
<td>.023</td>
</tr>
<tr>
<td>No. low-inventory heavy*store dist/10</td>
<td>-.066</td>
<td>.009</td>
<td>.003</td>
<td>.002</td>
<td>-.050</td>
<td>.004</td>
</tr>
<tr>
<td>Delivery fee</td>
<td>5.217</td>
<td>.313</td>
<td>-.188</td>
<td>.056</td>
<td>3.449</td>
<td>.160</td>
</tr>
<tr>
<td>Weekday dummy</td>
<td>7.781</td>
<td>1.147</td>
<td>-.230</td>
<td>.206</td>
<td>4.183</td>
<td>.386</td>
</tr>
<tr>
<td>Weather</td>
<td>-8.855</td>
<td>3.156</td>
<td>-.360</td>
<td>.568</td>
<td>-4.410</td>
<td>1.612</td>
</tr>
<tr>
<td>Store distance</td>
<td>.356</td>
<td>.061</td>
<td>.004</td>
<td>.011</td>
<td>.275</td>
<td>.031</td>
</tr>
<tr>
<td>Family size</td>
<td>-.054</td>
<td>.265</td>
<td>-.021</td>
<td>.048</td>
<td>-.051</td>
<td>.136</td>
</tr>
<tr>
<td>Preschool children</td>
<td>-.793</td>
<td>.525</td>
<td>-.048</td>
<td>.095</td>
<td>-.701</td>
<td>.268</td>
</tr>
<tr>
<td>Parking</td>
<td>5.528</td>
<td>.993</td>
<td>1.475</td>
<td>.179</td>
<td>2.345</td>
<td>.507</td>
</tr>
<tr>
<td>Internet partner store (Y/N)</td>
<td>9.026</td>
<td>1.116</td>
<td>1.197</td>
<td>.201</td>
<td>3.614</td>
<td>.570</td>
</tr>
<tr>
<td>MBS perishables*</td>
<td>-9.095</td>
<td>4.790</td>
<td>-.123</td>
<td>.862</td>
<td>22.985</td>
<td>2.447</td>
</tr>
<tr>
<td>MBS heavy/bulky</td>
<td>10.377</td>
<td>6.146</td>
<td>20.367</td>
<td>1.106</td>
<td>11.867</td>
<td>3.139</td>
</tr>
<tr>
<td>Medium income</td>
<td>-1.256</td>
<td>1.434</td>
<td>.032</td>
<td>.258</td>
<td>.419</td>
<td>.733</td>
</tr>
<tr>
<td>High income</td>
<td>.056</td>
<td>1.296</td>
<td>.302</td>
<td>.233</td>
<td>.333</td>
<td>.662</td>
</tr>
<tr>
<td>R²</td>
<td>.222</td>
<td>.241</td>
<td>.281</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* MBS: Mean Basket Share for each household across all store visits

Table 4  Population Parameter Estimates and Standard Deviations

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th></th>
<th>Standard Deviation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Std. Error</td>
<td>Estimate</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Intercept</td>
<td>.143</td>
<td>.219</td>
<td>1.778</td>
<td>.335</td>
</tr>
<tr>
<td>Store distance</td>
<td>.580</td>
<td>.224</td>
<td>.949</td>
<td>.230</td>
</tr>
<tr>
<td>No. of Items/10</td>
<td>.676</td>
<td>.314</td>
<td>.352</td>
<td>.072</td>
</tr>
<tr>
<td>Weekday</td>
<td>1.549</td>
<td>.389</td>
<td>.459</td>
<td>.063</td>
</tr>
<tr>
<td>No. of Heavy/bulky/10</td>
<td>.520</td>
<td>.149</td>
<td>.228</td>
<td>.046</td>
</tr>
<tr>
<td>No. heavy/bulky*distance/10</td>
<td>1.014</td>
<td>.426</td>
<td>.792</td>
<td>.161</td>
</tr>
<tr>
<td>Weather</td>
<td>1.298</td>
<td>.636</td>
<td>.826</td>
<td>.179</td>
</tr>
<tr>
<td>No. of Perishables/10</td>
<td>-.348</td>
<td>.176</td>
<td>.252</td>
<td>.076</td>
</tr>
<tr>
<td>Delivery charge</td>
<td>-.129</td>
<td>.643</td>
<td>.391</td>
<td>.104</td>
</tr>
<tr>
<td>Overall promotion index</td>
<td>-.131</td>
<td>.070</td>
<td>.192</td>
<td>.018</td>
</tr>
<tr>
<td>Promotion index of perishables</td>
<td>-.084</td>
<td>.047</td>
<td>.458</td>
<td>.069</td>
</tr>
<tr>
<td>Promotion index of heavy/bulky items</td>
<td>.127</td>
<td>.073</td>
<td>.313</td>
<td>.033</td>
</tr>
<tr>
<td>Estimates</td>
<td>Store intercept</td>
<td>No. items</td>
<td>week day</td>
<td>No. heavy</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>Intercept</td>
<td>-4.535</td>
<td>.538</td>
<td>.627</td>
<td>.849</td>
</tr>
<tr>
<td>Family size</td>
<td>.186</td>
<td>-.039</td>
<td>-.021</td>
<td>.003</td>
</tr>
<tr>
<td>Preschool children</td>
<td>.108</td>
<td>.067</td>
<td>-.018</td>
<td>-.096</td>
</tr>
<tr>
<td>Fish shop</td>
<td>-.388</td>
<td>-.039</td>
<td>-.058</td>
<td>.414</td>
</tr>
<tr>
<td>Parking</td>
<td>.700</td>
<td>.118</td>
<td>-.041</td>
<td>-.198</td>
</tr>
<tr>
<td>Partner store</td>
<td>-.073</td>
<td>-.012</td>
<td>-.018</td>
<td>.154</td>
</tr>
<tr>
<td>Delivery service</td>
<td>-.018</td>
<td>-.054</td>
<td>-.007</td>
<td>-.217</td>
</tr>
<tr>
<td>Store size/1000</td>
<td>1.985</td>
<td>-.251</td>
<td>-.229</td>
<td>-.726</td>
</tr>
<tr>
<td>Medium Income</td>
<td>.436</td>
<td>.208</td>
<td>-.053</td>
<td>-.274</td>
</tr>
<tr>
<td>High Income</td>
<td>.728</td>
<td>.212</td>
<td>-.053</td>
<td>.318</td>
</tr>
<tr>
<td>Downtown</td>
<td>-.266</td>
<td>.384</td>
<td>.014</td>
<td>.296</td>
</tr>
<tr>
<td>MBS perishables*</td>
<td>-3.833</td>
<td>-.590</td>
<td>-.539</td>
<td>1.004</td>
</tr>
<tr>
<td>MBS heavy/bulky</td>
<td>6.106</td>
<td>.538</td>
<td>.627</td>
<td>.881</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard errors</th>
<th>Store intercept</th>
<th>No. items</th>
<th>week day</th>
<th>No. heavy</th>
<th>Heavy *dist</th>
<th>weather perish</th>
<th>delivery overall charge</th>
<th>Promo perish</th>
<th>Promo heavy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.306</td>
<td>.289</td>
<td>.096</td>
<td>.507</td>
<td>.069</td>
<td>.233</td>
<td>1.071</td>
<td>.216</td>
<td>.224</td>
</tr>
<tr>
<td>Family size</td>
<td>.094</td>
<td>.036</td>
<td>.006</td>
<td>.041</td>
<td>.005</td>
<td>.028</td>
<td>.091</td>
<td>.029</td>
<td>.015</td>
</tr>
<tr>
<td>Preschool children</td>
<td>.240</td>
<td>.089</td>
<td>.014</td>
<td>.093</td>
<td>.011</td>
<td>.084</td>
<td>.236</td>
<td>.030</td>
<td>.039</td>
</tr>
<tr>
<td>Fish shop</td>
<td>.227</td>
<td>.216</td>
<td>.032</td>
<td>.251</td>
<td>.027</td>
<td>.128</td>
<td>.666</td>
<td>.066</td>
<td>.100</td>
</tr>
<tr>
<td>Parking</td>
<td>.505</td>
<td>.186</td>
<td>.029</td>
<td>.249</td>
<td>.027</td>
<td>.083</td>
<td>.664</td>
<td>.090</td>
<td>.084</td>
</tr>
<tr>
<td>Delivery service</td>
<td>.420</td>
<td>.168</td>
<td>.027</td>
<td>.178</td>
<td>.021</td>
<td>.089</td>
<td>.415</td>
<td>.140</td>
<td>.075</td>
</tr>
<tr>
<td>Store size/1000</td>
<td>2.094</td>
<td>.838</td>
<td>.128</td>
<td>.815</td>
<td>.101</td>
<td>.331</td>
<td>1.442</td>
<td>.288</td>
<td>.355</td>
</tr>
<tr>
<td>Medium Income</td>
<td>.278</td>
<td>.283</td>
<td>.040</td>
<td>.300</td>
<td>.033</td>
<td>.132</td>
<td>.575</td>
<td>.090</td>
<td>.112</td>
</tr>
<tr>
<td>High Income</td>
<td>.336</td>
<td>.243</td>
<td>.035</td>
<td>.167</td>
<td>.026</td>
<td>.107</td>
<td>.687</td>
<td>.112</td>
<td>.106</td>
</tr>
<tr>
<td>Downtown</td>
<td>.560</td>
<td>.213</td>
<td>.032</td>
<td>.311</td>
<td>.025</td>
<td>.084</td>
<td>.409</td>
<td>.087</td>
<td>.083</td>
</tr>
<tr>
<td>MBS perishables</td>
<td>1.977</td>
<td>.759</td>
<td>.122</td>
<td>.710</td>
<td>.096</td>
<td>.309</td>
<td>.202</td>
<td>.268</td>
<td>.311</td>
</tr>
<tr>
<td>MBS heavy/bulky</td>
<td>1.844</td>
<td>.690</td>
<td>.102</td>
<td>.448</td>
<td>.080</td>
<td>.249</td>
<td>1.412</td>
<td>.277</td>
<td>.300</td>
</tr>
</tbody>
</table>

* MBS: Mean Basket Share for each household across all store visits. Bold numbers refer to significant at 5% or 10% level.
Figure 1  Household Distribution (%) by Shares of Online Grocery Expenditure and Trips

Figure 2  Distribution of Shopping Trips by Basket Size
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 1 Km of Travel

Figure 6 Distribution of Relative Value of Internet Channel
Appendix Derivation of Posterior Distributions

In this appendix, we derive the various conditional distributions for the Gibbs sampler in the MCMC approach. The models are:

\[ Y_h = G_h \gamma_{1h} + X_h \beta_h + Z_h \gamma + \varepsilon_h \]  
(\text{Channel choice model})  
(A1)

\[ X_h = (G_h Z_h) \Pi + v_h \]  
(\text{Endogenous variables})  
(A2)

\[ B = D \Delta + U_o, \quad U_o \sim N(0, V_o) \]  
(2\text{nd stage of the hierarchical model})  
(A3)

\[ \left( \begin{array}{c} \varepsilon_h \\ v_h \end{array} \right) \sim N(0, \Sigma), \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \Sigma^{-1} = V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}, \quad \varepsilon_h | v_h \sim N(-V_{11}^{-1}V_{12}v_h, V_{11}^{-1}) \]  
(A4)

Where \( G_h \) is the exogenous variables (including intercepts), \( X_h \) is the endogenous variables and \( Z_h \) is additional instruments other than \( G_h \). \( B \equiv \{ \gamma_{1h}, \beta_h \} \) and \( D \) is the matrix of household demographics. Priors for the common parameters are:

\[ \gamma \sim N(\mu_\gamma, A_\gamma^{-1}), \quad \text{vec}(\Pi|\Sigma) \sim N(\text{vec}(\Pi), \Sigma_{22} \otimes A_\Pi^{-1}) \]
\[ \Sigma \sim IW(v_\Sigma, V_\Sigma), \quad \text{vec}(\Delta|V_o) \sim N(\text{vec}(\Delta), V_o \otimes A_\Delta^{-1}) \]
\[ V_o \sim IW(v_{o0}, V_{o0}) \]  
(A5)

The posteriors are:

\[ Y_h \mid G_h, X_h, Z_h, \beta_h, \gamma_{1h}, \gamma, \Sigma \sim \text{Truncated Normal} \]
\[ \{ \beta_h, \gamma_{1h} \} \mid Y_h, G_h, X_h, Z_h, \beta_h, \gamma_{1h}, \gamma, \Sigma, \Delta, V_o \sim \text{Normal} \]
\[ \gamma \mid Y_h, G_h, X_h, Z_h, \{ \beta_h, \gamma_{1h} \}, \Pi, \Sigma, \mu_\gamma, A_\gamma \sim \text{Normal} \]
\[ \Pi \mid G_h, X_h, Z_h, \{ \beta_h, \gamma_{1h} \}, \Sigma, \Pi, A_\Pi \sim \text{Normal} \]
\[ \Sigma \mid Y_h, G_h, X_h, Z_h, \beta_h, \gamma_{1h}, \gamma, \Pi \sim \text{Inverted Wishart} \]
\[ \Delta \mid \{ \beta_h, \gamma_{1h} \}, V_o, \bar{\Delta}, A_\Delta \sim \text{Normal} \]
\[ V_o \mid \{ \beta_h, \gamma_{1h} \}, v_{o0}, V_{o0} \sim \text{Inverted Wishart} \]  
(A6)

Next we provide the detail distributions of these posteriors.

(1) \[ Y_h \mid G_h, X_h, Z_h, \beta_h, \gamma_{1h}, \gamma, \Sigma \sim \text{Truncated Normal} \]

Let \( I_i \) be household \( h \)'s \( i \)th choice observation and \( N \) be the total number of observations across all households. Given \( \beta_h, \gamma_{1h}, \gamma \) and \( \Sigma \), the \( \{ Y_h \} \) are independent, univariate truncated normal random vectors, and the regions of truncation \( R_{ij} \) are defined as (Rossi et al 2005, P116):
\[
Pr(I_t | Y_{ih}, \beta_h, \gamma_{1h}, \gamma, \Sigma) = \int_{R_h} \phi(Y_h | G, X, Z, \beta_h, \gamma_{1h}, \gamma, \Sigma) dY_h
\]

\[
R_{t_i} = \begin{cases} 
\{Y_h : Y_{ih} > Y_{h2}, \text{ if } I_i = 1 \text{ (online)} \\
\{Y_h : Y_{ih} < Y_{h2}, \text{ if } I_i = 2 \text{ (offline)}
\end{cases}
\]

(A7)

Here \( \phi(\cdot) \) is the univariate normal density function.

(2) \( \{\beta_h, \gamma_{1h}\} | Y_h, G_h, X_h, Z_h, D, \gamma, \Pi, \Sigma, \Delta, V_{\omega} \sim \text{Normal} \)

Given \( \Pi \), we can “observe” \( \nu_h \). We can condition our analysis of (2) on \( \nu_h \):

\[
Y_h = G_h \gamma_{1h} + X_h \beta_h + Z_h \gamma + \varepsilon_h | \nu_h
\]

\[
= G_h \gamma_{1h} + X_h \beta_h + Z_h \gamma - V^{-1}_{11} V_{12} \nu_h + \varepsilon_{2|1} \quad \text{var}(\varepsilon_{2|1}) \equiv v_{2|1}^2 = V_{11}^{-1}
\]

\[
\frac{Y_h - Z_h \gamma + V_{11}^{-1} V_{12} \nu_h}{v_{2|1}} = \frac{G_h}{v_{2|1}} \gamma_{1h} + \frac{X_h}{v_{2|1}} \beta_h + \frac{\varepsilon_{2|1}}{v_{2|1}} = N(0,1)
\]

\[
\{\beta_h, \gamma_{1h}\} | Y_h, \tilde{Y}_h, G_h, \tilde{X}_h, D, \gamma, \Pi, \Delta, V_{\omega} \sim N(\tilde{\beta}, (\tilde{X}_h' \tilde{X}_h + (V_{\omega})^{-1})^{-1})
\]

\[
\tilde{\beta} = (\tilde{X}_h' \tilde{X}_h + (V_{\omega})^{-1})^{-1} (\tilde{X}_h' \tilde{Y}_h + (V_{\omega})^{-1}(\Delta)' D_h)
\]

\[
\tilde{X}_h = (G_h' X_h), \quad \tilde{Y}_h = \frac{Y_h - Z_h \gamma + V_{11}^{-1} V_{12} \nu_h}{v_{2|1}}
\]

(A8)

(3) \( \gamma | Y_h, G_h, X_h, Z_h, \{\beta_h, \gamma_{1h}\}, \Pi, \Sigma, \mu_j, A_j \sim \text{Normal} \)

Similarly, given \( \Pi \), we can “observe” \( \nu_h \) and thus can condition our analysis of (3) on \( \nu_h \):

\[
\frac{Y_h - G_h \gamma_{1h} - X_h \beta_h + V_{11}^{-1} V_{12} \nu_h}{v_{2|1}} = \frac{Z_h}{v_{2|1}} \gamma + \frac{\varepsilon_{2|1}}{v_{2|1}} \sim N(0,1)
\]

\[
\gamma | \tilde{Y}_h, \tilde{Z}_h, \{\beta_h, \gamma_{1h}\}, \Pi, \Sigma, \mu_j, A_j \sim N(\tilde{\gamma}, (\tilde{Z}_h' \tilde{Z}_h + A_j^{-1})^{-1})
\]

\[
\tilde{\gamma} = (\tilde{Z}_h' \tilde{Z}_h + A_j^{-1})^{-1} (\tilde{Z}_h' \tilde{Y}_h + A_j \mu_j)
\]

\[
\tilde{Z}_h = \frac{Z_h}{v_{2|1}}, \quad \tilde{Y}_h = \frac{Y_h - G_h \gamma_{1h} - X_h \beta_h + V_{11}^{-1} V_{12} \nu_h}{v_{2|1}}
\]

(A9)

(4) \( \Pi | G_h, X_h, Z_h, \{\beta_h, \gamma_{1h}\}, \Sigma \sim \text{Normal} \)

This conditional can be handled by transforming to the reduced form that can be written as a regression model. As there are \( n_x \) endogenous variables, we need to manipulate the equations as:

\[
\tilde{Y}_h = Y_h - G_h \gamma_{1h} - Z_h \gamma = X_h \beta_h + \varepsilon_h = (G_h' Z_h) \Pi \beta_h + (v_h \beta_h + \varepsilon_h)
\]
$X_h = (G_h Z_h) \Pi + v_h$

Let $C_h = \begin{bmatrix} 1 & \beta_h' \\ 0 & I_n \end{bmatrix}$, then $(v_h \beta_h + e_h) \sim N \left( 0, \begin{bmatrix} \beta_h' \Sigma_{22} \beta_h + 2 \Sigma_{12} \beta_h + \Sigma_{11} & \beta_h' \Sigma_{22} + \Sigma_{12} \\ \Sigma_{22} \beta_h + \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) = N(0, C_h \Sigma C_h')$

Let $L_h$ be the transpose of the Cholesky root of $C_h \Sigma C_h'$, then

$L_h^{-1} \left( \begin{array}{c} \tilde{Y}_h \\ \tilde{X}_h \end{array} \right) = L_h^{-1} \left( \begin{array}{c} (G_h Z_h) \Pi \beta_h \\ (G_h Z_h) \Pi \end{array} \right) + L_h^{-1} \left( v_h \beta_h + e_h \right), \quad \Psi \equiv L_h^{-1} \left( v_h \beta_h + e_h \right) \sim N(0, I)$

$L_h^{-1} \left( \begin{array}{c} (G_h Z_h) \Pi \beta_h \\ (G_h Z_h) \Pi \end{array} \right) = L_h^{-1} \left( \begin{array}{c} (G_h Z_h) \Pi \beta_h \\ (G_h Z_h) \Pi \end{array} \right)$

vec($\Pi$)

Let $\tilde{Y}_h \equiv L_h^{-1} \left( \begin{array}{c} \tilde{Y}_h \\ \tilde{X}_h \end{array} \right)$ and $\tilde{X}_h = L_h^{-1} \left( \begin{array}{c} G_h Z_h \beta_1 \\ G_h Z_h \beta_2 \\ \cdots \\ G_h Z_h \beta_{nx} \\ 0 \\ \cdots \\ G_h Z_h \\ 0 \\ \cdots \\ G_h Z_h \end{array} \right)$. We have:

vec($\Pi$) | $G_h, X_h, Z_h, \{\beta_h, \gamma_{ih}\}, \Sigma \sim N(\text{vec}(\tilde{\Pi}), (\tilde{X}' \tilde{X} + \Sigma_{22}^{-1} \otimes A_{11})^{-1})$

$\tilde{\Pi} = (\tilde{X}' \tilde{X} + \Sigma_{22}^{-1} \otimes A_{11})^{-1} (\tilde{X}' \tilde{Y} + (\Sigma_{22}^{-1} \otimes A_{11}) \tilde{\Pi})$

(A10)

(5) $\Sigma | Y_h, G_h, X_h, Z_h, \beta_h, \gamma_{ih}, \gamma, \Pi \sim \text{Inverted Wishart}$

$\Sigma | Y_h, G_h, X_h, Z_h, \beta_h, \gamma_{ih}, \gamma, \Pi \sim IW(v_0 + N, S + V_0)$

$S = \sum_{i=1}^{N} e_i e_i'$, $e_i = \begin{bmatrix} e_{hi} \\ v_{hi} \end{bmatrix}$

$e_{hi} = Y_h - (G_h \gamma_{ih} + X_h \beta_h + Z_h \gamma)$

$v_{hi} = X_h - (G_h Z_h) \Pi$

(A11)

(6) $\Delta | (\beta_h, \gamma_{ih}), \omega, \Lambda, \omega, \omega$ (Rossi et al, 2005, P34)

$\Delta | (\beta_h, \gamma_{ih}), \omega, \Lambda, \omega \sim N(\text{vec}(\tilde{\Delta}), \omega, \omega \otimes (D' D + A_{\omega})^{-1})$

$\tilde{\Delta} = (D' D + A_{\omega})^{-1} (D' B + A_{\omega} \Lambda)$

(A12)

(7) $V_\omega | (\beta_h, \gamma_{ih}), D, V_{\omega 0}, V_{\omega 0}$ (Rossi et al, 2005, P34)

$V_\omega | (\beta_h, \gamma_{ih}), D, V_{\omega 0}, V_{\omega 0} \sim IW(v_{\omega 0} + n, V_{\omega 0} + S)$

$S = (B - D\tilde{\Delta})' (B - D\tilde{\Delta}) + (\Lambda - \tilde{\Lambda})' A_{\omega} (\Lambda - \tilde{\Lambda})$

(A13)