A Structural Model of Channel Choice with Implications for Retail Entry

Scott Shriver and Bryan Bollinger *

February 15, 2015

Abstract

In this paper, we propose a structural model of channel choice and product purchases, in which consumers base their channel choice decision on both the utility inherent in visiting a channel and the expected utility from purchasing products upon visiting. Specifically, we allow channels to have different abilities to resolve consumer uncertainty about product fit that varies by product category. We use this model to evaluate and predict the demand impact of retail entry for a seasonal goods provider that uses both retail and online channels to sell directly to consumers. Both online and retail sales will depend on retail outlet proximity through the rate of new customer arrivals and the shopping trip frequency of existing customers, as well the tradeoffs inherent in choosing between channel formats and the outside option. We find evidence of channel complementarity through increased arrivals of new customers and shopping frequency as the distance to retail outlets decreases. While the net effect of retail entry is demand-stimulative, we find these market expansion effects are partially offset by increased switching from online to retail formats in areas close to store locations. In a series of counterfactual experiments, we demonstrate how our estimates may be used to identify promising locations for new physical stores and to explore channel-based price discrimination policies.

*Shriver: Columbia Business School, Columbia University, ss4127@columbia.edu. Bollinger: The Fuqua School of Business, Duke University, bryan.bollinger@duke.edu. We thank the Wharton Customer Analytics Initiative (WCAI) for access to the data and seminar participants at the Marketing Science conference for their comments. We are also grateful to Avi Goldfarb, Kitty Wang, and Hema Yoganarasimhan for useful discussions regarding the paper. All errors are our own.
1 Introduction

Technological advances associated with the digital age have resulted in a proliferation of channels by which consumers interact with providers of goods and services. Increasingly, retailers manage a combination of geographically ubiquitous “always on” channels such as e-commerce websites and catalog-driven call centers, in addition to traditional “brick and mortar” outlets. A key question of interest to marketing practitioners and academics is how these channels interact with respect to consumer demand. Specifically, managers wish to know to what extent the various channels act as strategic complements that drive additional demand or as strategic substitutes that cannibalize one another’s sales while increasing operating costs. The quantification of these effects is essential for identifying optimal channel strategy.

As the operation of e-commerce channels has become standard business practice, managers optimizing total profits must be equipped with tools to identify the set of physical store locations that profitably complement their digital counterparts. In this paper we provide such a tool by developing an empirical framework to estimate the demand implications of retail expansion (or contraction) when operating both online and retail channels. We take a structural approach to facilitate counterfactual experiments that allow us to quantify the demand response to retail entry in specific locations and to channel-specific pricing policies. Our model captures separate demand effects relating to purchase incidence rates and total expenditure levels, as well as channel and product choices. To identify these effects, we estimate the model using individual transaction-level data that contains rich between-subject and within-subject variation in customer proximity to the brand’s retail locations, where the latter arises due to the brand’s rapid expansion of its retail footprint during our window of observation.

The multi-category model we develop employs a discrete-continuous framework that has been previously applied in marketing studies of consumer demand for variety (Wales and Woodland, 1983; Hanemann, 1984; Kim et al., 2002; Bhat, 2008). Our model incorporates a number of methodological innovations relative to this literature. First, similar to Chintagunta (1993), ours is a unified utility-based framework that jointly predicts purchase incidence, category choices and purchase quantities; in addition, our model explains channel format choices. We depart from Chintagunta (1993) in that we pursue a direct (rather than indirect) utility formulation for the product choice model, which leads to a second methodological contribution. Using insights from Pinjari and Bhat (2010), we devise a highly efficient (polynomial time) algorithm to solve the direct utility demand system for optimal quantities given model parameters and draws of the un-
observables, which in turn enables us to jointly estimate our multi-stage demand system via full information maximum simulated likelihood.\(^1\) The tractability of this approach demonstrates that direct utility specifications can be extended to richer modeling contexts than previously encountered in the literature. Third, our sequential choice, expected utility formulation illustrates a computationally attractive mechanism to endogenize trip expenditures with respect to both product and channel utility primitives.\(^2\) Fourth, our framework serves as a prototype for modeling demand for seasonal goods, where the product categories offered remain stable over time but the set of SKUs available within the categories vary over time.

Substantively, we find that the retail channel does act as a demand complement in both acquiring new customers and in increasing brand consideration (hence purchase incidence) by existing customers. This complementarity is partly offset by cannibalization of sales in the online channel, such that retail revenue gains due to entry are almost double the revenues lost in the online channel. While this revenue contribution ratio is relatively constant as a function of retail store distance, the magnitude of these effects and thus the absolute revenue gains from entry decrease rapidly as store distances increase. We find that the channels contribute equally to total revenues at a distance of approximately 25 miles from the retail store, and that the share of online sales quickly dominates at greater distances. We also find that product category price elasticities can vary considerably by channel and that the relative ordering of price elasticities also changes depending on the channel, highlighting the additional effect entry will have on the composition of consumer shopping baskets. Using a model specification with two latent classes, we find that a small segment of consumers (estimated to be 6.2% of the population) generates nearly ten times the expected (per capita) annual revenue of customers in the dominant segment, thus generating approximately 40% of the firm’s revenue. Exploiting differential channel and product preferences across these customer populations is one potential mechanism for the firm to implement a price discrimination policy that we explore in our counterfactual simulations.

The rest of the paper is organized as follows: in Section 1.1 we briefly discuss the related empirical literature. In Section 2 we describe the data used for the study and explore variation in key relationships. In Section 3 we present our structural model of product and channel utility. Section 4 describes our estimation

\(^1\)As explained in Section 3.2 and Appendix B, traditional approaches to forecasting demand from a direct utility system have required the use of constrained optimization procedures, for every draw of the model unobservables. Such an approach would be computationally infeasible (for any dataset of appreciable size) where the estimation procedure embeds computation of optimal demand.

\(^2\)As will be seen, differentiating the inclusive value of the channel decision stage with respect to expenditure provides a first order condition that pins down the optimal trip budget.
method and presents the model estimates. We further explore the implications of our estimates and perform counterfactual experiments in Section 5. We summarize our findings and propose further avenues of research in Section 6.

### 1.1 Related literature

Researchers have studied the effect of new channel introduction in multiple industries including print (Deleersnyder et al., 2002; Gentzkow, 2007) and music (Biyalogorsky and Naik, 2003). However, the majority of existing studies have investigated the impact of adding an online (website) channel to existing brick and mortar or catalog channels (Ansari et al., 2008; Biyalogorsky and Naik, 2003; Deleersnyder et al., 2002; Geyskens et al., 2002; Van Nierop et al., 2011). Generally, these papers find that addition of the online format does not cannibalize existing channels, with the size of demand complementarities varying by product category. To our knowledge, only three studies have analyzed the impact of a firm adding physical stores to its existing online and catalog channels. Table 1 summarizes findings across these studies.  

<table>
<thead>
<tr>
<th>Study</th>
<th>Data</th>
<th>Analysis</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avery et al. (2012)</td>
<td>Market panel</td>
<td>Diff-in-diff</td>
<td>Short-run: Catalog cannibalized</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Long-run: Channels complementary</td>
</tr>
<tr>
<td>Pauwels and Neslin</td>
<td>Aggregate time series</td>
<td>VAR</td>
<td>Catalog cannibalized</td>
</tr>
<tr>
<td>(2011)</td>
<td></td>
<td></td>
<td>Online unaffected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Net revenue gain</td>
</tr>
<tr>
<td>Wang and Goldfarb</td>
<td>Market panel</td>
<td>Diff-in-diff</td>
<td>Weak complementarity of online/retail channels. Cannibalization in areas with</td>
</tr>
<tr>
<td>(2014)</td>
<td></td>
<td></td>
<td>existing brand presence.</td>
</tr>
</tbody>
</table>

Table 1: Empirical studies of adding retail to existing channels

Avery et al. (2012) use market-level panel data from an apparel/home furnishings retailer to measure the impact of opening a new physical store on net catalog sales, net online sales, sales from new customers and sales from existing customers. They use a differences-in-differences methodology, where control markets are identified using a propensity scoring algorithm, and find that both catalog and online sales are cannibalized in the short run but tend to recover over time. Data limitations prevent Avery et al. (2012) from

---

Forman et al. (2009) also study the impact of retail stores on online sales but not in the context of a single firm managing multiple channels. They use a differences-in-differences approach and find significant sales declines on Amazon.com when local bookstores open. It is useful to consider why our application may be different. These authors focus on the sales of top-selling books only, which limits the potential to find an increase in product consideration by new customers – presumably the top titles are already known. There is almost no role of increased consideration by existing customers since it is not the decision to visit the channel that is modeled but the ranking of a particular book, and consumers seldom purchase the same book twice and almost always buy a single copy of the book, whereas in our case, the quantity of products purchased may also be affected by retail entry.
controlling for customer heterogeneity, firm marketing activities, and competitive conditions.

Pauwels and Neslin (2011) use vector autoregression to analyze similar data (also apparel categories) in an aggregate time series format. The VAR approach permits Pauwels and Neslin (2011) to simultaneously model multiple endogenous variables, including the frequency and size (in $) of orders, returns and exchanges by channel (online, catalog, store) as well as the total number of customers in the market. Pauwels and Neslin (2011) control for firm marketing activity but do not accommodate customer heterogeneity or competitive controls. The results are generally consistent with Avery et al. (2012) in that the authors find physical store introduction: (a) cannibalizes catalog sales but not online sales, (b) increases the rate of new customer acquisition, and (c) decreases catalog returns but increases overall returns.

The third paper is Wang and Goldfarb (2014), which uses some of the same data we analyze. Similar to Avery et al. (2012), Wang and Goldfarb (2014) use a differences-in-differences methodology to investigate the impact of retail entry on online and retail sales in markets defined by Census tracts. Wang and Goldfarb (2014) find that the offline channel can serve as a marketing communications device for the online channel, thereby increasing sales. In areas with a weak brand presence, defined as whether the Census tract has no sales in the three months prior to entry, this effect dominates (as expected, since there are no pre-existing online sales to cannibalize). In contrast, areas with a prior brand presence also exhibit channel cannibalization effects. Our empirical findings are broadly consistent with those of Wang and Goldfarb (2014), in that both studies find evidence of channel complementarity and substitution, with complementary effects dominating substitution effects. Differences in methods and data (they include two additional brands in their analysis and analyze market panel rather than individual panel data) make an exact comparison of results difficult. However, it appears our estimates point to a stronger (economic and statistical) net effect of channel complementarity in areas close to retail outlets. We conjecture that our operationalization of channel effects using individual customer distances to retail locations rather than counts of store locations within a Census tract provides a richer source of identifying variation for the effects of interest, but warrant that by using linear panel models Wang and Goldfarb (2014) are able to employ a richer set of controls than our estimation algorithm will permit. Another point of differentiation is that our structural model can speak to the specific mechanisms (e.g., brand consideration, channel preferences, product preferences, etc.) that lead to observed differences in channel revenues. The structural approach also allows us to perform counterfactual experiments such as assessing the desirability of new entry locations or channel-based pricing policies.

In addition to the literature on multichannel customer management, our paper is related to the literature
on store choice. Bell and Lattin (1998) study the choice of retail store format in an environment with different levels of price uncertainty since some retailers use EDLP and some use HILO. The choice of channels in our context has a similar character, in that channel formats differ with respect to their ability to resolve uncertainty about product attributes. This ability to assess product fit has been shown to be an important factor in channel format choices by Bell et al. (2013) and Soysal and Zentner (2014). The store choice literature has also emphasized the importance of planned expenditure levels (or “basket size” in the language of Bell and Lattin (1998)) and shopping trip fixed costs (e.g. Bell et al., 1998) as determinants of shopping format choices. We similarly model channel visit utility as a function of the trip budget and, in the case of the retail channel, a function of the transportation cost incurred to reach the store. Whereas the standard practice in the store choice literature has been to treat expenditure levels as exogenously determined (e.g. Bell and Lattin (1998) use a pre-estimation calibration to obtain the probability that consumers are either “large basket” or “small basket” shoppers, Bell et al. (1998) assume budgets arise implicitly from an unobserved shopping list, etc.), our approach is to endogenize the trip budget decision as a function of expected optimal channel and product selection. A primary benefit of this approach is that we are able to predict how expenditure levels change in response to changes in prices or retail channel accessability.

2 Data

Data for the study comes from a North American specialty retailer that sells to customers through both online and retail channels. The focal brand sells a variety of seasonal goods (e.g., apparel and housewares) that are available for a limited period of time. Given our interest in predicting channel-specific demand both in and out of sample, the perpetual introduction and retirement of goods makes individual SKUs both conceptually and computationally impractical as the unit of analysis (in our case, a SKU-level analysis would involve more than 70,000 products). Therefore, using methods that we describe in Section 2.1 below, for our empirical work we aggregate demand to the top level of the brand’s product hierarchy, which is comprised of six distinct categories. Our analytical framework will thus characterize consumer category demand and the extent to which channel formats influence that demand.

A consumer’s propensity to access the retail channel is of course dependent upon her proximity to the

---

4 See Neslin et al. (2006) for a nice survey of the literature on multi-channel customer management.
5 Confidentiality agreements preclude us from disclosing the identity of the brand or revealing precise descriptions of the products in their portfolio.
retail outlets operated by the brand. The brand exhibited a large expansion of its retail footprint over our two-year period of study (July 2010 - June 2012), growing from 37 retail outlets to 75. The distribution of stores and customers at the end of the observation window is shown in Figure 1. The firm provided us customer home locations by Census block as well as the coordinates (latitude, longitude) and opening dates of all its retail outlets. These data allow us to calculate a customer’s proximity to a retail store at any point in time with extremely high precision (roughly 1/3 of a mile accuracy). The brand’s extensive retail market entry therefore provides a rich source of both within and between subject variation that can be used to identify the different mechanisms (e.g., impact on purchase frequency, total expenditure and channel choices) by which demand responds to changes in the availability of channel formats. We explore these mechanisms through a series of descriptive regressions in Section 2.2.

2.1 Preparation and summary

We observe individual-level purchase transactions over the two year period July 2010 to June 2012. We restrict our sample to customers residing in the continental United States, yielding a total of 29,130 transactions from 10,242 unique customers. Transaction records indicate the date and channel format of the shopping trip as well as the quantities and prices of individual SKUs purchased.

To facilitate our empirical analysis, we organize the data in a bi-level format. Specifically, our approach is to summarize demand outcomes as a sequence of trips (which includes the possibility of zero trips) within a time period (quarter) that are unique to a given customer. This tactic allows us to associate
continuously-varying price and store distance information with summary discrete-time measures. We begin our description of the data with the upper level, which summarizes customer/quarter observations of: a) the number of observed trips, b) the distribution of customer distances to the active set of retail stores, and c) the price indices for our six categories. Summary statistics for (a) and (b) are provided in Table 2, while the price indices, which are not specific to individuals, are summarized graphically in Figure 2. The data summarized in Table 2 is an unbalanced panel across the 8 quarters of study, reflecting the observed acquisition of new customers. The table indicates that on average consumers shop at the brand twice per year (1/2 trip per quarter) and live approximately 43 miles from nearest store, although there is considerable variation in retail proximity. The corresponding within-subject variation in store distance (induced by retail entry events) is substantial, with a standard deviation of 28.38 miles.

<table>
<thead>
<tr>
<th>observations</th>
<th>mean</th>
<th>std dev</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>quarter</td>
<td>57,577</td>
<td>5.13</td>
<td>2.21</td>
<td>1 8</td>
</tr>
<tr>
<td>trips</td>
<td>57,577</td>
<td>0.51</td>
<td>1.22</td>
<td>0 36</td>
</tr>
<tr>
<td>store distance</td>
<td>57,577</td>
<td>42.65</td>
<td>81.44</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 2: Upper level (customer/quarter panel) summary statistics

In the spirit of Stone (1954), we compute category \((k)\) and channel \((c)\) specific price indices for a quarter \((t)\) as geometric expenditure-weighted means of SKU-level prices:

\[
    p_{tck} = \exp \left( \frac{\sum_{j\in k} w_{cj} \log (p_{tcj})}{\sum_{j\in k} w_{cj}} \right) \quad \text{where} \quad w_{cj} = \frac{e_{cjt}}{\sum_{j,t} e_{cjt}} \tag{1}
\]

In (1) above, \(p_{cjt}\) is the average (across consumers) price paid for SKU \(j\) in channel \(c\) and quarter \(t\), while \(e_{cjt}\) is the total expenditure (across consumers) for the same SKU. The applied weights, which are the fraction of total channel expenditures attributable to a given SKU, naturally reflect more popular products in the resulting index but are time-stationary to minimize correlation with the quantity indices we construct to measure category demand in the lower (trip-level) data. As is readily apparent from Figure 2, for most categories, prices are slightly lower in the retail channel. A SKU-level analysis of prices reveals that these differences are largely due to more aggressive discounting in the retail channel, such as end of season clearance sales.
Next we turn to the lower-level data, which captures trip-level outcomes. We summarize observed trips in terms of the total expenditure, the selected channel (where we code online as 1 and retail as 2), and quantity indices for each of the purchased categories. To compute the latter, we simply sum SKU-level expenditures by category and divide by the corresponding price index. That is, the quantity index $q_{itlk}$ for consumer $i$ in quarter $t$ on trip $l$ in category $k$ is:

$$q_{itlk} = \frac{\sum_{j \in k} e_{itlj}}{P_{tck}}$$  \hspace{1cm} (2)

Table 3 indicates that the average per-trip expenditure is approximately $141 and that the 57% of observed trips are to the retail channel. Unit sales tend to be highest for category 1 and lowest for category 6, which is the highest priced category.

Table 3: Lower level (trip) summary statistics

<table>
<thead>
<tr>
<th></th>
<th>observations</th>
<th>mean</th>
<th>std dev</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>quarter</td>
<td>29,130</td>
<td>4.73</td>
<td>2.25</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>total expenditure</td>
<td>29,130</td>
<td>140.84</td>
<td>153.70</td>
<td>0.13</td>
<td>3215.96</td>
</tr>
<tr>
<td>channel</td>
<td>29,130</td>
<td>1.57</td>
<td>0.50</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>quantity 1</td>
<td>29,130</td>
<td>0.63</td>
<td>1.09</td>
<td>0</td>
<td>18.44</td>
</tr>
<tr>
<td>quantity 2</td>
<td>29,130</td>
<td>0.22</td>
<td>0.60</td>
<td>0</td>
<td>11.47</td>
</tr>
<tr>
<td>quantity 3</td>
<td>29,130</td>
<td>0.21</td>
<td>0.56</td>
<td>0</td>
<td>10.26</td>
</tr>
<tr>
<td>quantity 4</td>
<td>29,130</td>
<td>0.49</td>
<td>1.00</td>
<td>0</td>
<td>20.73</td>
</tr>
<tr>
<td>quantity 5</td>
<td>29,130</td>
<td>0.46</td>
<td>1.21</td>
<td>0</td>
<td>41.87</td>
</tr>
<tr>
<td>quantity 6</td>
<td>29,130</td>
<td>0.06</td>
<td>0.28</td>
<td>0</td>
<td>7.16</td>
</tr>
</tbody>
</table>

Figure 2: Channel-specific category price indices
We examine the relationship between expenditures and channel formats in Figure 3, which plots kernel density estimates of trip expenditures by channel. It is clear that customers spend more per trip in the online channel (average values are $168.85 for online and $119.42 for retail), and that there are a large number of small expenditure transactions in the retail channel. In our sample, 52% of total revenue comes from the online channel and 48% from retail. We explore variation in category choices by channel in Figure 4, which shows the average basket composition by channel as a proportion of trip expenditure. Expenditures on categories 1, 2 and 5 are higher in the retail channel whereas categories 3, 4 and 6 dominate in the online channel. The bulk of the expenditure share variation across channels is linked to categories 1, 5 and 6, where there is a clear preference for category 6 in the online channel and for categories 1 and 5 in the retail channel.

![Figure 3: Kernel density estimates of trip expenditure by channel](image1)

![Figure 4: Average category expenditure shares by channel](image2)

### 2.2 Descriptive analyses

In this section, through descriptive regressions we explore how distance to the retail outlet affects demand along four dimensions: the acquisition of new customers, the frequency of shopping trips, the total expen-
diture per trip, and the channel chosen for the shopping trip. Our findings here guide the formulation of our structural model in Section 3.

2.2.1 New customer acquisition

To assess how the presence of a retail outlet may influence new customer acquisition, we organize the data into a market (Census block group) and quarter panel format and model the arrival of new customers \(N_{mt}\) as a Poisson process:

\[
N_{mt} \sim \text{Poisson}(\rho_{mt}), \quad \log(\rho_{mt}) = \alpha \log(\text{store\_distance}_{mt}) + \delta_m + \mu_t
\]  

In (3), \(\text{store\_distance}_{mt}\) is the distance from Census block group \(m\)'s centroid to the nearest retail store in quarter \(t\). The specification includes bi-level (market, quarter) fixed effects to control for unobservables – \(\alpha\) is thus identified from deviations in the market average rates of new customer arrivals, after controlling for common (across market) time trends. The estimate of \(\alpha\) from this regression is \(\hat{\alpha} = -0.8522\) with a standard error of 0.0635, which is significant at well below the 1% level. The negative sign is expected, as it implies acquisition rates go up in areas closer to retail outlets. The large magnitude of this effect indicates that new customer acquisition is one of the primary benefits of operating a retail channel.

For the purpose of evaluating our counterfactual in Section 5.2, which uses model estimates to assess desirable retail entry locations, it is useful to decompose the market fixed effects \(\delta_m\) from equation (3) in terms of market level observables. To do this, we perform an auxiliary Poisson regression in which we constrain the distance effect to be as estimated above and include select market demographic information obtained from Census 2010 records, i.e.:

\[
N_{mt} \sim \text{Poisson}(\rho_{mt}), \quad \log(\rho_{mt}) = \alpha \log(\text{store\_distance}_{mt}) + \beta z_m + \mu_t
\]  

These parameter estimates are reported in Table 4 below. The results suggest strong positive correlation with market population, average age, median income, fraction of population white and fraction of population with college degrees. There is a negative correlation associated with rural markets.
<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>population</td>
<td>0.0002***</td>
</tr>
<tr>
<td>age</td>
<td>0.0115***</td>
</tr>
<tr>
<td>income</td>
<td>0.0035***</td>
</tr>
<tr>
<td>rural</td>
<td>-0.0742***</td>
</tr>
<tr>
<td>white</td>
<td>0.6323***</td>
</tr>
<tr>
<td>college</td>
<td>2.8100***</td>
</tr>
<tr>
<td>State FE</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 4: Decomposition of new customer acquisition market fixed effects

### 2.2.2 Shopping trip frequency

Here we return to our customer/quarter panel data format and model the number of shopping trips per quarter as a Poisson arrival process, again including bi-level fixed effects:

\[ L_{it} \sim \text{Poisson}(\rho_{it}), \log(\rho_{it}) = \alpha \log(\text{store\_distance}_{it}) + \delta_i + \mu_t \]  

(5)

The estimate of \( \alpha \) from this regression is \( \hat{\alpha} = -0.1040 \) with a standard error of 0.0164, again significant below the 1% level. We conclude that the presence of a retail outlet has a significant impact on the shopping trip incidence rate, presumably due to increased top of mind awareness of the brand.

### 2.2.3 Expenditure level

We use our full dataset and a linear model with bi-level (individual, quarter) fixed effects to investigate the effect of retail store distance on trip expenditures:

\[ e_{itl} = \alpha \log(\text{store\_distance}_{it}) + \delta_i + \mu_t + e_{itl} \]  

(6)

The estimate of \( \alpha \) from this regression is \( \hat{\alpha} = -0.0123 \) with a standard error of 0.0148, which is insignificant at any conventional level. Similar results hold if we condition upon only online or only retail format trips. We thus conclude there is no definitive descriptive evidence that retail distance directly influences trip expenditure levels.
2.2.4 Channel choice

Finally, we analyze the effect of retail distance on channel format choices. We use a binary logit model, again including bi-level fixed effects:

$$Pr(c_{lt} = 2) = \frac{\exp(\alpha \log(\text{store\_distance}_{lt}) + \delta_l + \mu_t)}{1+\exp(\alpha \log(\text{store\_distance}_{lt}) + \delta_l + \mu_t)}$$  

(7)

The estimated $\alpha$ is $\hat{\alpha} = -0.5253$ with a standard error of 0.0638, which is significant below the 1% level. We conclude, unsurprisingly, that proximity to a retail outlet has a strong effect on channel choices – the closer to the retail outlet, the higher the probability of a retail trip. The obvious interpretation of this result is that lower transportation costs drive higher retail shopping frequency.

3 Model

In this section, we develop our demand model for existing customers. When performing counterfactual experiments quantifying the benefits of retail entry, we also incorporate the effect of retail outlet proximity on new customer acquisition using our results in Section 2.2.1. Our proposed model for existing customer demand, outlined in Figure 5 below, incorporates a brand consideration arrival process followed by three subsequent consumer decision stages in which the shopping trip budget, channel choice and product choices are sequentially determined. At each decision stage, consumers are assumed to condition upon previously determined outcomes and to take actions that maximize their current stage utility, which incorporates expectations of optimal decision making in future stages. Consumers are further assumed to have rational expectations with respect to the distribution of future stage unobservables, channel-specific product prices, and their own trip budgets (i.e., the trip budget determined in the first stage binds with equality in subsequent product selection stage).

To illustrate the model flow, consider one existing customer in a single period (recall that our data is organized as a collection of trips for each customer/quarter combination, so that prices and store distances are assumed constant for all trips within the period). During the period, the customer is assumed to consider

---

6Our choice to model decisions in this particular sequence is motivated by the store choice literature (e.g., Bell and Lattin, 1998), where “basket size” is found to be an important determinant of retail format choices. However, in contrast to Bell and Lattin (1998), who classify consumers into large/small basket types during a model initialization phase, our approach is to endogenize the shopping trip budget as an optimal expenditure decision that incorporates beliefs about channel offerings. As a robustness check, we explored an alternative model formulation in which the trip budget is co-determined with product choices, conditional upon a channel already being chosen (i.e., the outside option enters the product selection model rather than in an independent decision stage). While this model generates qualitatively similar results, we prefer the current specification on theoretical grounds as well as substantially superior in-sample fit – results are available from the authors upon request.
a purchase with the focal brand $R$ times, where $R$ is a Poisson distributed variable with rate parameter $r$. The actual number of brand consideration events is unobserved, but $R$ must be at least as great as the observed number of purchase occasions (which we denote by $L$). The rate parameter $r$ will be a function of the distance to the nearest retail outlet, under the assumption that proximity to the retail store generates top of mind awareness of the brand. $r$ therefore reflects the (exogenous) frequency at which a customer needs products offered by the firm as well as the conditional probability that the consumer considers the focal brand. For each consideration “arrival,” the consumer first determines the optimal allocation of her income ($w$) to the brand’s products. Upon realization of the outside option utility shock ($\epsilon_0$), the consumer sets the shopping trip budget ($b$) to equate the marginal utility of allocating $w - b$ dollars to the outside option and the marginal utility she expects from allocating $b$ dollars to brand purchases, assuming she selects the optimal channel and product bundle upon realization of their respective utility shocks ($\eta$ in the case of channels, and $\{\epsilon, \nu\}$ in the case of products). In the event that the marginal utility of the outside option exceeds the expected marginal utility from allocating any income towards brand purchases, the consumer’s optimal decision is to set the trip budget equal to zero, resulting in a no-purchase event. Assuming a non-zero budget, in the next decision stage channel utility shocks are realized and the consumer chooses between the online and retail channels based upon her relative preference for the retail format (an intercept in the model), her distance from the retail store, and her expected utility from spending $b$ dollars on products in either channel. Finally, upon selecting a channel, channel-specific product utility shocks are realized and the consumer allocates her budget among the categories based upon her product preferences and current period prices.

In the following subsections, we develop the model starting with the final (product choice) decision stage.
and build upwards. Throughout the model development, we account for persistent unobserved heterogeneity in consumer preferences via latent classes. To avoid parameter proliferation and to ensure the numerical stability of our estimation routine, we restrict the specification of heterogeneous effects to intercepts in the various model stages.

3.1 Product choice model

As noted in our discussion of the data in Section 2, a primary challenge we face is how to model demand in a seasonal goods setting, where specific products are only available for a short period of time and the cumulative product catalog is extremely large (well in excess of 10,000 unique products). Recognizing that the set of product categories offered by the brand is stable over time, our approach to this problem is to pursue a category level demand model. That is, we assume consumers form preferences for the categories offered by the brand, but have uncertainty regarding the match value of products within the category for their needs on any given shopping occasion. We conjecture that match value uncertainty may be both category and channel-specific, since channel formats may differ with respect to the amount of product information they convey (e.g., retail formats may be a more efficient means to assess product “fit” or “feel”) as well as the SKU-level availability of product assortments (e.g., websites commonly stock a wider range of product sizes and colors than may be found in retail outlets). In the context of the model, match value uncertainty is captured by product/channel specific utility shocks that are revealed to the consumer upon visiting the chosen channel.

A second consideration is that the model must be able to rationalize observations of purchases in multiple categories and, given our interest in forecasting revenues, must be able to deliver predictions of the quantities purchased. Accordingly, we develop a multiple discrete/continuous model of product category demand that extends the specification of Bhat (2008) with latent classes. Conditional upon channel choice \( c \) and trip budget \( b \), consumer \( i \) (a member of market segment \( s \)) obtains the following utility from purchasing product bundle \( \bar{q} \) on the \( l \)th shopping trip in period \( t \):

\[
\begin{align*}
    u(q_{itl}|c_{itl}, b_{itl}, i \in s) &= \sum_{k=1}^{K} \Psi_{ltlk} \gamma_k \log \left( \frac{q_{itlk}}{\gamma_k} + 1 \right) \\
    \Psi_{ltlk} &= \exp(\beta_{sk} + \nu_{ltlk} + \varepsilon_{ltlk}) \\
    \varepsilon_{ltlk} &\sim iid \ EV(0, \sigma_{\varepsilon}) \\
    \nu_{ltlk} &\sim iid \ N(0, \sigma_{\nu_{ltlk}})
\end{align*}
\]
Equation (8a) implies the utility of the bundle is additively separable in sub-utilities for each of the $K$ product categories. The presence of the $\log(\cdot)$ expression ensures that the consumer has decreasing marginal consumption utility in each category. The role of the $\gamma$ parameters are to alter the rates at which the consumer’s utility satiates in category consumption – as illustrated in Figure 6, a higher $\gamma$ implies a lower rate of satiation. The $\Psi$ terms are the so-called “baseline” marginal utilities, because $\Psi_k$ is the marginal utility obtained in the limit of consuming zero quantity of good $k$ ($\lim_{q_k \to 0} \frac{\partial u}{\partial q_k}$). The parameterization of the $\Psi$ terms is provided in equation (8b). We allow for heterogenous category preferences through the intercepts in the baseline marginal utilities, $\beta_{sk}$. Further, we introduce two stochastic terms ($\nu$ and $\epsilon$) into the baseline marginal utilities, the distributions of which will induce a likelihood for the observed product data. The $\nu$ terms are normally distributed channel/category specific shocks that capture product fit and assortment information which is observed by the consumer but unobserved by the analyst. We additionally include extreme value shocks ($\epsilon$) which are iid across all categories and channels – this compound shock specification is taken to reduce the computational burden of computing the likelihood function, much in the same way a mixed logit model offers computational advantages to the multivariate probit model in a purely discrete choice setting. We discuss the implications of this formulation for parameter identification further in Section 4.2.

The reader will note the absence of prices from the model in equation (8), which is customary when specifying a direct utility function. Under the direct utility approach, the consumer is assumed to maximize her utility subject to a budget constraint, which incorporates product prices. Prices affect demand through the (binding) budget constraint and are subsequently reflected in the optimal consumption quantities ob-
tained from solving the constrained optimization problem. In the current context, the consumer’s problem is (henceforth we suppress $i, t$ and $l$ subscripts where possible for expositional clarity): 

$$\max_{q_1, \ldots, q_K} u(\vec{q} | c, b) \text{ subject to: } \sum_{k=1}^{K} q_k p_{ck} = b, \; q_k \geq 0$$

(9)

When deriving the likelihood function for this model, it proves convenient to work with an alternative but entirely equivalent specification of equation (9). Specifically, we formulate the consumer’s problem in terms of selecting category expenditures rather than product quantities, where the category expenditure is defined as $e_k = q_k p_{ck}$:

$$\max_{e_1, \ldots, e_K} u(\vec{e} | c, b) \text{ subject to: } \sum_{k=1}^{K} e_k = b, \; e_k \geq 0$$

(10)

To proceed, we first form the Lagrangian for the problem: $\mathcal{L} = u(\vec{e} | c, b) - \lambda \left( \sum_{k=1}^{K} e_k - b \right)$. Optimal quantities will then satisfy the Kuhn-Tucker (KT) conditions: $\frac{\partial \mathcal{L}}{\partial e_k} \leq 0$, $e_k \geq 0$, $\frac{\partial \mathcal{L}}{\partial c_k} e_k = 0$ for all $k = 1, \ldots, K$. Without loss of generality, assume that the first good is chosen.\footnote{With a non-zero budget, at least one alternative must be chosen. Moreover, designating a chosen category as category 1 amounts to a simple re-labeling of the alternatives. As will be seen, the choice of the which category is first has no bearing on the form of the likelihood function when the consumer’s problem is formulated in terms of selecting expenditure levels. If the problem is cast in terms of selecting product quantities, the resulting likelihood is scaled by the price of the category that is selected as the reference category. While maximization of either likelihood function will result in the same set of parameters, using the expenditure formulation facilitates the fit comparison of different empirical model specifications.}

With this convention, we show in Appendix A.1 that the KT conditions for the consumer’s problem reduce to:

$$V_{ck} + e_k \begin{cases} = V_{c1} + \varepsilon_1 & \text{if } e^*_k > 0 \\ < V_{c1} + \varepsilon_1 & \text{if } e^*_k = 0 \end{cases}$$

(11)

where: $V_{ck} \equiv \beta_{ck} + V_{ck} - \log \left( \frac{e^*_k}{p_{ck} \lambda_k} + 1 \right) - \log (p_{ck})$

In (11), star superscripts denote optimal (as opposed to conjectured) expenditures. Note that for purchased goods, we have $\varepsilon_k = V_{c1} - V_{ck} + \varepsilon_1$ and for non-purchased goods we have $\varepsilon_k < V_{c1} - V_{ck} + \varepsilon_1$. These conditions, which define the portion of the $\varepsilon$ space that can rationalize the observed expenditure pattern $\vec{e}^*$, are sufficient to derive the likelihood function. Again without loss of generality, assume that $M$ (where $M \geq 1$) of the $K$ product categories are chosen, and that they are ordered such that categories 1 to $M$ are the chosen categories. With this convention, the probability of observing expenditure pattern $\vec{e}^*$ may be written as

$$Pr(\vec{e}^* | c, b) = \int_{\mathcal{I}_{\varepsilon_1, \ldots, \varepsilon_M}} \int_{e_1}^{\infty} \int_{e_2}^{\infty} \cdots \int_{e_M}^{\infty} dF(\vec{e})dF(\vec{V})$$

(12)
This likelihood function may be conceptually partitioned in two parts. The probability of observing the set of chosen categories is generated through the equality KT conditions in (11) above. These conditions map the probability density of the $e$ error terms to the pattern of non-zero expenditures, which generates the Jacobian term by a change-of-variables calculus. The probability of observing the set of non-chosen categories requires integrating over the portion of the $e$ space that is consistent with a “corner” solution of no category purchase, which generates the integrals over $e_{M+1}$ to $e_K$. The additional integration over $e_1$ appears because the KT conditions are conditioned upon $e_1$ – integrating over $e_1$ thus generates the unconditional probability of the expenditure pattern. A similar rationale motivates integration over the distribution of channel/category shocks ($v$).

Under the iid extreme value assumption, the integration with respect to the $e$ terms may be computed analytically, as may the Jacobian determinant. The resulting likelihood reduces to a remarkably simple expression (see Appendix A.2 for a full derivation):

$$Pr(e^\ast|c,b) = \frac{(M - 1)!}{\sigma_e^{M-1}} \left( \prod_{j=1}^{M} e_j + \gamma_j p_{cj} \right) \left( \sum_{j=1}^{M} e_j + \gamma_j p_{cj} \right) \int_v \left( \prod_{j=1}^{M} e_j^{V_{cj}/\sigma_e} \right) \left( \sum_{j=1}^{K} e_j^{V_{cj}/\sigma_e} \right)^{-M} dF(\widetilde{V})(13)$$

As pointed out by Bhat (2008), equation (13) is the multiple discrete/continuous analog of the mixed logit model for discrete choices (the mixed logit model is recovered if $M=1$). As with the mixed logit model, the integration over the mixing distribution ($v$) must be performed numerically – we employ simulation methods to compute this integral in our estimation procedure.

### 3.2 Channel choice model

Conditional upon a positive budget amount $b$, consumers are assumed to choose between the online (1) and retail (2) formats based on their preferences for the formats and their expectations of how much utility they will acquire by spending $b$ on products in the respective channels. When making their choices, we assume consumers anticipate channel-specific product prices and know the distribution of channel-specific unobservables. Presuming that, all else equal, consumers may prefer a channel format where their expected product utility is less variable (i.e., they may be risk averse), we allow channel utility to be a function of the
product utility variance as well as its expectation. We thus define risk-adjusted channel utilities as follows:

\[ U_{itl} = \mathbb{E}[u(\tilde{q}^*)|b_{itl}, c = 1] + \phi_3 \text{Var}[u(\tilde{q}^*)|b_{itl}, c = 1] + \eta_{itl} \equiv \bar{U}_{itl} + \eta_{itl} \quad (14a) \]

\[ U_{itl} = \mathbb{E}[u(\tilde{q}^*)|b_{itl}, c = 2] + \phi_3 \text{Var}[u(\tilde{q}^*)|b_{itl}, c = 2] + \phi_2 \text{Distance}_u + \phi_{itl} + \eta_{itl} \equiv \bar{U}_{itl} + \eta_{itl} \quad (14b) \]

\[ \eta_{itl} \sim \text{iid EV}(0, \sigma_\eta) \quad (14c) \]

In (14), \( \phi_{itl} \) captures the consumer’s relative preference for the retail format (assuming she is a member of segment \( s \)), while \( \phi_2 \) captures the (dis)utility incurred from traveling to the retail outlet, and \( \phi_3 \) captures risk preferences related to expected product utility. Note that there is no coefficient on the expected product utility, which is instead captured through the estimated variance of the channel utility shocks (\( \sigma_\eta \)).

The expectations and variances in (14) are taken with respect to the distributions of \( \varepsilon \) and \( \nu \). A complication that arises when computing these terms is that no closed form expressions are available – they must be simulated by taking a large number of draws of the unobservables, solving for the optimal quantities associated with each draw, and then computing the first two moments of the resulting utility distribution. We develop a highly efficient algorithm to solve for optimal product quantities and nest it within the maximization of the full model likelihood, which enables us to achieve full econometric efficiency by jointly estimating all model parameters.

Consumers are assumed to choose the channel that gives the highest risk-adjusted utility. Given the iid extreme value assumption on \( \eta \), the probability of observing channel choice \( c^* \) is given by the standard logit formula (\( itl \) subscripts are omitted but implied, as is conditioning on the individual’s segment assignment):

\[
Pr(c^* | b) = \int_{\eta} (U_c = \max[U_1, U_2])dF(\tilde{\eta}) = \frac{\exp \left( \frac{U_1}{\sigma_\eta} \right)}{\exp \left( \frac{U_1}{\sigma_\eta} \right) + \exp \left( \frac{U_2}{\sigma_\eta} \right)} \quad (15)
\]

\(^8\)As noted by Train (2009) and Swait and Louviere (1993), this model is equivalent to one in which the extreme value variance is normalized to 1 and the coefficients are rescaled by a factor of \( \frac{1}{\sigma_\eta} \). In this sense, the coefficient on the expected product utility is \( \frac{1}{\sigma_\eta} \).

\(^9\)Historically, direct utility (Kuhn-Tucker) demand systems have required the use of constrained optimization procedures to solve for optimal quantity choices, for each draw of model unobservables. Such an approach would be computationally infeasible for any large scale estimation problem that embeds computation of the expected utility from optimal product choices. Rather, our algorithm solves for optimal quantities exactly via matrix operations in polynomial time for the number of observations, categories and simulation draws.
3.3 Trip budget model

Conditional upon a consideration event, consumers choose how much of their income to allocate to purchasing products at the focal brand by setting a shopping trip budget. Similar to the product model, this is a discrete/continuous decision, such that a budget allocation of zero dollars corresponds to a no purchase event. In making this decision, consumers apportion their income ($w$) between the outside good ($q_0$) and the shopping budget for the focal brand ($b$). The direct utility governing this decision is given by:

$$U_0(b_{itl}, q_{itl0}) = \Psi_{itl0} + E \left[ \max \left\{ \overline{U}_{itl1}(b_{itl}) + \eta_{itl1}, \overline{U}_{itl2}(b_{itl}) + \eta_{itl2} \right\} \right]$$

(16a)

$$= \Psi_{itl0} + \sigma_\eta \log \left( \exp \left( \frac{\overline{U}_{itl1}(b_{itl})}{\sigma_\eta} \right) + \exp \left( \frac{\overline{U}_{itl2}(b_{itl})}{\sigma_\eta} \right) \right)$$

$$= \Psi_{itl0} + g(b_{itl})$$

(16b)

$$\Psi_{itl0} = \exp (\beta_\alpha + \epsilon_{itl0})$$

(16c)

The first term on the right side of equation (16a) is the utility gleaned from the outside good and the second is the expected utility of choosing the optimal product bundle from the focal brand in the optimal channel. The second equality follows from the iid extreme value assumption on $\eta$ (the channel choice shocks); the log-sum term is defined as $g(b)$ for notational convenience. As in the other model stages, we allow heterogeneity in preferences for the outside option by permitting $\beta_\alpha$ to be segment specific.

The consumer solves the problem:

$$\max_{q_0, b} U_0(q_0, b) \text{ subject to } q_0 + b = w, b \geq 0$$

(17)

Some allocation of income to the outside good is considered essential. Under this assumption, it is shown in Appendix C that the Kuhn-Tucker conditions corresponding to (17) are:

$$V_0(b) \equiv \log \left( \frac{\partial g}{\partial b} \right) - \beta_\alpha \begin{cases} = \epsilon_0 & \text{if } b^* > 0 \\ < \epsilon_0 & \text{if } b^* = 0 \end{cases}$$

(18)

The conditions in (18) may be used to derive the probability of observing a trip budget (total trip expenditure) of $b^*$. However, it must first be recognized that we never directly observe a zero dollar budget (only trips resulting in purchase are observed). Therefore, the relevant probability is the conditional probability of a budget given that it is strictly positive, i.e., $Pr(b^* | b^* > 0)$. With this aspect in mind, the probability of
observing a trip budget of $b^*$ (itl subscripts are omitted) is given by:

$$ Pr(b^* \mid b^* > 0) = |J_{e_0 - e}| \frac{\lambda \left( \frac{V_0(b^*)}{\sigma_{e_0}} \right)}{\Lambda \left( \frac{V_0(0)(0)}{\sigma_{e_0}} \right)} = \frac{\partial^2 g}{\partial b^2} \left| \frac{\partial g}{\partial b} \right| \Lambda \left( \frac{V_0(0)}{\sigma_{e_0}} \right) $$ \hfill (19)

where $\lambda (\cdot)$ and $\Lambda (\cdot)$ are the standard extreme value pdf and cdf, respectively. The expression $\Lambda (V_0(0))$ enters (19) to account for the conditional probability that $b^* > 0$. A full derivation of (19) with the Jacobian expressed in terms of the simulation draws is provided in Appendix C.

### 3.4 Consideration arrival process

A consumer’s consideration of the focal brand follows a Poisson arrival process with rate parameter $r_{it} = \exp(\mu_z z_{it})$, where $z_{it}$ will include an intercept, the consumer’s distance to the retail store, and controls for seasonality (quarter dummies). Let $R_{it}$ be the number of consideration events per period (quarter), so that $R_{it} \sim \text{Poisson}(r_{it})$. Given the iid assumption on the outside good shocks ($e_0$), the probability of purchase given consideration is independent across consideration events and is equal to $\Lambda \left( \frac{V_0(0)(0)}{\sigma_{e_0}} \right)$.

To close the model, we must derive a likelihood function to recover the arrival rate parameters $\mu$. To do this, we relate the unobserved consideration process to the observed number of shopping trips in the quarter, $L$. The key insight is that conditional upon a realization of $R$, the number of observed purchase occasions $L$ follows a binomial distribution ($R$ Bernoulli trials, each with probability of “success” $Pr(b^* > 0) = \Lambda \left( \frac{V_0(0)}{\sigma_{e_0}} \right)$). To form the likelihood, we integrate over the distribution of unobserved consideration events that can rationalize an observation of $L$ purchase occasions, as follows (it subscripts are suppressed but implied):

$$ Pr[L] = \sum_{R=L}^{\infty} Pr[L \mid R] Pr[R] = \sum_{R=L}^{\infty} \binom{R}{L} \left[ \Lambda \left( \frac{V_0(0)}{\sigma_{e_0}} \right) \right]^{L} \left[ 1 - \Lambda \left( \frac{V_0(0)}{\sigma_{e_0}} \right) \right]^{R-L} \frac{e^{-r} r^R}{R!} \hfill (20) $$

Thus, the observed number of purchase occasions is also Poisson distributed, with rate parameter $\Lambda \left( V_0(0) \right) r$. This compound rate parameter provides a linkage between the purchase incidence rate and the structural parameters governing channel and product utility.
4 Estimation and results

Before presenting our results in Section 4.3, we first describe our maximum likelihood estimation procedure in Section 4.1 and then briefly discuss identification issues in Section 4.2.

4.1 Joint likelihood

We begin construction of the full model likelihood function with the probability of observed outcomes for a single trip. Using equations (13), (15) and (19), this likelihood may be written:

\[ L_{itl} | i \in s = \text{Pr}(b_{itl}^s, c_{itl}^s, \epsilon_{itl}^s | i \in s) = \text{Pr}(b_{itl}^s | b_{itl} > 0, i \in s) \text{Pr}(c_{itl}^s | b_{itl}, i \in s) \text{Pr}(\epsilon_{itl}^s | b_{itl}, c_{itl}, i \in s) \]  

(21)

Given the independence across trips and using equation (20), the likelihood of the collection of trips for customer \( i \) in period \( t \) is:

\[ L_{it} | i \in s = \text{Pr}(L_{it} | i \in s) \left( \prod_{l=1}^{L_{it}} L_{it} | i \in s \right)^{(L_{it} > 0)} \]  

(22)

Under our latent class heterogeneity specification, the joint unconditional probability of the \( T \) period observations for customer \( i \) is then:

\[ L_{i} = \sum_{s=1}^{S} \left( \prod_{t=1}^{T} L_{it} | i \in s \right) \text{Pr}(i \in s) \]  

(23)

Where we specify the probability of segment assignment as:

\[ \text{Pr}(i \in s) = \frac{\text{exp}(\xi_s)}{\sum_{j=1}^{S} \text{exp}(\xi_j)} \]  

(24)

Finally, the joint likelihood of the model parameters is given by the product of the customer-level likelihood expressions:

\[ L(\theta | Z) = \prod_{i=1}^{I} L_{i} \]  

(25)

where \( I \) is the number of unique customers, \( \theta \) is the collection of all model parameters and \( Z \) represents all observed data.

The parameters are estimated via full information maximum simulated likelihood: \( \hat{\theta} = \text{argmax}_{\theta} \left( \log(L(\theta)) \right) \).

\(^{10}\)With the inclusion of latent classes, the joint likelihood function can be difficult to maximize using standard gradient-based optimization packages. Our experience has been that the KNITRO active set (sequential quadratic programming) algorithm using the quasi-Newton SR1 Hessian approximation substantially outperforms other alternatives. To ensure convergence to a global maximum, we leverage the multi-start feature of the KNITRO solver and subsequently attempt to tighten solutions using a Nelder-Mead simplex algorithm.
To speed convergence of the algorithm, we obtain starting values for the joint estimation procedure by first sequentially maximizing the likelihood with respect to the product, channel, trip budget and consideration arrival parameters, generating inclusive value terms as necessary. We compute the Monte Carlo integration over $v$ in (13) and the simulate the expected product utilities in (15) using 250 draws generated from a Halton sequence. Given our “large $N$, small $T$” panel data format, to obtain standard errors, we appeal to asymptotics based on the individual customer likelihoods. Specifically, we compute the outer product of the gradients (OPG) estimator of the information matrix (evaluated at $\hat{\theta}$) by averaging the gradient of $L_i$ over the individuals in our sample.

4.2 Identification

In a non-linear multi-stage model such as the one presented here, functional form and distributional assumptions inevitably play an important role in parameter identification. Nevertheless, key patterns of variation in the data link unambiguously to certain parameters. Here we discuss those patterns as well as any parameter normalizations required for identification. As with our model exposition, we begin at the lowest level (the product model) and move upwards.

The product model involves four types of parameters: the baseline utility intercepts ($b_k$), the satiation parameters ($g_k$), the channel/category shock variances ($\sigma_{\eta_{ck}}$) and the variance of the extreme values shocks common to all categories ($\sigma_e$). The $b_k$ values are primarily pinned down by variation in category purchase incidence rates, while the $g_k$ values relate most directly to the quantity purchased conditional on category incidence. The shock variances $\sigma_{\eta_{ck}}$ are linked to the demand response to price variation – the larger the magnitude of $\sigma_{\eta_{ck}}$, the smaller the corresponding demand response to changes in price.11 As in the purely discrete choice setting, neither the level nor scale of utility is separately identified. Normalizing the level of utility requires normalizing one of the $b_k$ parameters – for this, we set $b_1 = 0$ in all specifications. For the scale of utility, at most $K$ separate variance terms may be estimated per channel, meaning that either $\sigma_e$ or one of the $\sigma_{\nu_{ck}}$ values must be normalized. We choose to normalize $\sigma_e$ and estimate the full set of $\sigma_{\nu_{ck}}$ values. As explained by Bhat (2008), $\sigma_e$ must be normalized to a small enough (but still positive) level such that $\frac{1}{\sigma_e}$ is larger than the actual demand response for all categories.12 We found that normalizing $\sigma_e = 0.25$

---

11 Note that in this model $\frac{1}{\sqrt{\sigma_{\eta_{ck}} + \sigma_e}}$ is effectively the price coefficient for category $k$ in channel $c$. Absent price variation, $\gamma$ and $\beta$ would still be identified while the $\sigma$ terms would not.

12 As a practical matter, this translates to setting $\sigma_e$ small enough to get statistically significant estimates of the $\sigma_{\nu_{ck}}$. 

---
was sufficiently small in our setting.

In the channel choice model, the channel variance $\sigma_1$ and the $\phi_3$ terms are identified by the explanatory power of the first two moments of the product utility distribution in determining channel choices – these terms are separately identified provided the product model is identified. The $\phi_2$ parameter is identified by both across and within subject variation in channel choice probabilities as a function of distance to the retail outlet. The retail intercept is identified by average channel choice probabilities, conditional upon the other factors in the model. For utility level identification, there is no intercept for the online channel. In the trip budget model, $\sigma_{e0}$ is identified by the explanatory power of the channel choice inclusive value in determining trip expenditures while $\beta_0$ is identified by average trip expenditures, conditional upon the inclusive value of the channel choice. Finally, the consideration arrival rate parameters are identified by the average shopping trip incidence rate conditional on distance, season of the year, and the no-purchase probability implied by the trip budget model.

4.3 Results

We estimate a homogeneous (1 segment) model as well as a specification with two latent classes. Parameter estimates are presented in Table 5. In the case of the two segment model, only the heterogenous effects (intercepts) are reported for the second segment (a “-” thus implies the parameter is constrained to be the same as the first segment). With the exception of $\mu_4$, the parameter capturing 1st quarter seasonality in the brand consideration rate, all estimates are significant at the 1% level.

We begin our discussion of the results with the homogeneous specification. Given the normalization of the category 1 intercept to 0, the ordering of categories in terms of baseline marginal utility ($\beta$ parameters) is: 1, 3, 2, 6, 4 and 5. The categories are ordered in terms of decreasing satiation rates (increasing $\gamma$) as: 6, 3, 2, 1, 4 and 5. For all but the first category, the online shock variance estimates ($\sigma_{v1k}$) exceed the retail shock variance estimates ($\sigma_{v2k}$), indicating that conditional upon a fixed budget, demand is generally more responsive to changes in price in the retail channel. This result is consistent with the notion that the retail channel provides greater product information (such as fit assessment), i.e., demand will be more sensitive to prices when, all else equal, unobservable product factors are smaller. To facilitate interpretation of how the model estimates collectively influence demand, we simulate outcomes from the model and plot the average category expenditure as a function of the budget amount by channel in Figure 7. We plot the corresponding utility distributions in Figure 8. These plots reflect, for example, the relative preference for category 6 in the
online channel and slightly larger total expected utility from purchases in the online channel (given a fixed budget level).

Consistent with our descriptive results in Section 2.2.4, the channel utility estimates indicate a relative preference for the retail channel ($\phi_1$ is positive) that decreases rapidly in the distance to the retail outlet ($\phi_2$ is negative). The negative coefficient on $\phi_3$ suggests that consumers exhibit mild risk aversion when choosing channels: the higher the variance of the product utility, the lower the risk-adjusted utility. Further, the estimated variance of the channel utility shocks ($\sigma_\eta$) is sufficiently small to imply that expected product utility has significant predictive power in determining channel choices (e.g., at sample average trip expenditures of $140, expected basket utilities are in the 3 to 4 range, yielding a utility contribution on par with the retail intercept value). Similar comments apply to the estimate of the outside good shock variance in the expenditure utility parameters. The brand consideration rate parameters also parallel the descriptive results in Section 2.2.2: consideration events (hence shopping trips) become less frequent as the distance to the retail store increases ($\mu_5$ is negative), and we have a higher consideration rates in the 3rd and 4th quarters ($\mu_2$ and $\mu_3$ are positive), consistent with more shopping during holiday periods. At sample average values, the implied consideration rate is 15.9 events per quarter (approximately 1.2 consideration events per week).

Turning to the two segment model estimates, the latent class assignment parameter ($\xi_1$) implies 93.8% of consumers are members of segment 2 and 6.2% belong to segment 1. Among the category intercepts in the product model, the biggest differences across segments are in categories 3 and 6, with preferences for these categories being larger among members of segment 2. For the channel choice model, segment 2 has slightly higher relative preference for the retail channel (higher $\phi_1$). Compared to the one segment model, the estimated channel shocks ($\sigma_\eta$) have smaller variance, indicating expected product utility plays a larger role in determining channel choices. Segment 1 has a lower intercept for the outside good and a higher consideration rate intercept, implying that segment 1 members shop more frequently and spend more than their segment 2 counterparts. The fit of the two segment model is significantly higher as measured by the log likelihood value and the Bayesian Information Criterion (BIC).

To further assess the overall fit of the model to the data, we compare the first two moments of the key outcomes in the data (number of trips, expenditure levels, channel selection, and product choices) to outcomes generated by simulating from the model estimates. Fit statistics are provided in Table 6. We perform these simulations using a random sample of 1000 customers, where we retain customer acquisition dates, prices and store distances at their historical values. As may be seen, despite its relative parsimony
the model provides an excellent fit to the data. In moving from the homogeneous model to the two segment model, the primary improvements in fit correspond to the higher (2nd) moments of the number of trips per quarter and the expenditure levels.

Figure 7: Expected category expenditures by channel, homogeneous model

Figure 8: Category contributions to expected utility by channel, homogeneous model
<table>
<thead>
<tr>
<th>Product utility parameters</th>
<th>Homogeneous model</th>
<th>2-segment model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>std err</td>
</tr>
<tr>
<td>Category 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_1$</td>
<td>6.5724</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_1$</td>
<td>-</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v11}$</td>
<td>0.4364</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v21}$</td>
<td>0.4445</td>
</tr>
<tr>
<td>Category 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_2$</td>
<td>6.2796</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_2$</td>
<td>-0.3106</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v12}$</td>
<td>0.5595</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v22}$</td>
<td>0.4141</td>
</tr>
<tr>
<td>Category 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_3$</td>
<td>5.8365</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_3$</td>
<td>-0.2147</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v13}$</td>
<td>0.8708</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v23}$</td>
<td>0.6588</td>
</tr>
<tr>
<td>Category 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_4$</td>
<td>7.3432</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_4$</td>
<td>-0.6400</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v14}$</td>
<td>0.7187</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v24}$</td>
<td>0.4991</td>
</tr>
<tr>
<td>Category 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_5$</td>
<td>7.4008</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_5$</td>
<td>-0.9639</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v15}$</td>
<td>0.9622</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v25}$</td>
<td>0.4921</td>
</tr>
<tr>
<td>Category 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>satiation</td>
<td>$\gamma_6$</td>
<td>5.1571</td>
</tr>
<tr>
<td>baseline utility intercept</td>
<td>$\beta_6$</td>
<td>-0.3182</td>
</tr>
<tr>
<td>online shock variance</td>
<td>$\sigma_{v16}$</td>
<td>0.9985</td>
</tr>
<tr>
<td>retail shock variance</td>
<td>$\sigma_{v26}$</td>
<td>0.3732</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel utility parameters</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail intercept</td>
<td>$\phi_1$</td>
<td>3.3816</td>
<td>0.0765</td>
<td>2.6144</td>
<td>0.0826</td>
<td>2.8942</td>
</tr>
<tr>
<td>log distance to store</td>
<td>$\phi_2$</td>
<td>-0.7762</td>
<td>0.0214</td>
<td>-0.6186</td>
<td>0.0231</td>
<td>-</td>
</tr>
<tr>
<td>Common risk preference</td>
<td>$\phi_3$</td>
<td>-0.0114</td>
<td>0.0004</td>
<td>-0.0162</td>
<td>0.0004</td>
<td>-</td>
</tr>
<tr>
<td>channel shock variance</td>
<td>$\sigma_{\eta}$</td>
<td>2.2133</td>
<td>0.0635</td>
<td>1.8206</td>
<td>0.0693</td>
<td>-</td>
</tr>
</tbody>
</table>

| Expenditure utility parameters |          |         |          |         |          |         |
| Outside option intercept    | $\beta_0$ | -2.1582 | 0.0626  | -2.7870 | 0.0474  | -2.5599 | 0.0518 |
| shock variance              | $\sigma_{00}$ | 1.1046  | 0.0252  | 0.9690  | 0.0227  | -       | -       |

| Consideration arrival parameters |          |         |          |         |          |         |
| intercept                    | $\mu_1$  | 2.7143  | 0.1018  | 3.0277  | 0.0692  | 1.4639  | 0.0850 |
| (3rd quarter)                | $\mu_2$  | 0.3626  | 0.0146  | 0.3459  | 0.0151  | -       | -       |
| (4th quarter)                | $\mu_3$  | 0.5236  | 0.0143  | 0.5220  | 0.0146  | -       | -       |
| (1st quarter)                | $\mu_4$  | -0.0001 | 0.0159  | 0.0690  | 0.0161  | -       | -       |
| log distance to store        | $\mu_5$  | -0.1171 | 0.0022  | -0.0862 | 0.0024  | -       | -       |

| Segment assignment parameters |          |         |          |         |          |         |
| intercept                    | $\xi_1$  | -       | -       | -       | 2.7126  | 0.0494 |

Log likelihood              | -401,721.2 | -393,555.3 |
# customer/quarter observations | 57,577 | 57,577 |
BIC                          | 803,815.1 | 787,581.9 |

Table 5: Model estimates
### Table 6: Model fit statistics

<table>
<thead>
<tr>
<th></th>
<th>data mean</th>
<th>std dev</th>
<th>1 segment model mean</th>
<th>std dev</th>
<th>2 segment model mean</th>
<th>std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>trips/quarter</td>
<td>0.51</td>
<td>1.22</td>
<td>0.50</td>
<td>0.72</td>
<td>0.51</td>
<td>0.93</td>
</tr>
<tr>
<td>expenditure/trip (unconditional)</td>
<td>140.84</td>
<td>153.70</td>
<td>136.29</td>
<td>135.72</td>
<td>143.68</td>
<td>145.15</td>
</tr>
<tr>
<td>expenditure/trip (online)</td>
<td>168.85</td>
<td>179.00</td>
<td>155.20</td>
<td>140.75</td>
<td>164.84</td>
<td>145.95</td>
</tr>
<tr>
<td>expenditure/trip (retail)</td>
<td>119.42</td>
<td>127.00</td>
<td>121.73</td>
<td>129.90</td>
<td>127.94</td>
<td>142.57</td>
</tr>
<tr>
<td>channel choice</td>
<td>1.57</td>
<td>0.50</td>
<td>1.57</td>
<td>0.50</td>
<td>1.57</td>
<td>0.49</td>
</tr>
<tr>
<td>fraction category 1 expenditure</td>
<td>0.32</td>
<td>0.41</td>
<td>0.32</td>
<td>0.43</td>
<td>0.33</td>
<td>0.43</td>
</tr>
<tr>
<td>fraction category 2 expenditure</td>
<td>0.12</td>
<td>0.29</td>
<td>0.12</td>
<td>0.29</td>
<td>0.10</td>
<td>0.27</td>
</tr>
<tr>
<td>fraction category 3 expenditure</td>
<td>0.13</td>
<td>0.30</td>
<td>0.12</td>
<td>0.30</td>
<td>0.12</td>
<td>0.30</td>
</tr>
<tr>
<td>fraction category 4 expenditure</td>
<td>0.20</td>
<td>0.35</td>
<td>0.22</td>
<td>0.37</td>
<td>0.21</td>
<td>0.36</td>
</tr>
<tr>
<td>fraction category 5 expenditure</td>
<td>0.18</td>
<td>0.34</td>
<td>0.18</td>
<td>0.34</td>
<td>0.20</td>
<td>0.36</td>
</tr>
<tr>
<td>fraction category 6 expenditure</td>
<td>0.05</td>
<td>0.21</td>
<td>0.05</td>
<td>0.20</td>
<td>0.04</td>
<td>0.17</td>
</tr>
</tbody>
</table>

### 5 Applications and strategic implications

In this section, we demonstrate the application of our results and further explore their implications.

#### 5.1 Quantifying channel substitution/complementarity

The primary substantive issue motivating the paper was the extent to which channels function as strategic substitutes or complements. Our approach to investigate this issue is to use the model estimates to calculate expected demand under different assumptions about the availability of the retail channel, operationalized through the distance to the nearest retail outlet. Specifically, we decompose demand effects into shopping incidence rates, total expenditure per trip and channel selection, which we then combine to make predictions about total revenue by channel as a function of distance to the retail outlet.\(^\text{13}\) For the sake of brevity, we perform the full decomposition for the homogeneous model and report summary revenue implications for the two segment model. Throughout, we assume prices are those realized in the final period of observation.

We begin with the uppermost layer of the model and calculate the expected number of shopping trips per year. The expected number of trips for a given quarter may be computed unconditionally given the model parameters, since the influence of channel and product utility primitives only enter through expectations over subsequent decision stages.\(^\text{14}\) To obtain the expected number of trips per year, we simply aggregate

\(^{\text{13}}\) An alternate approach would focus on profits, but this requires either observations of costs (which we do not have) or assumptions about the firm’s pricing rules. To the extent that our interest here is positive rather than normative, analysis of revenues is sufficient to summarize channel demand interactions.

\(^{\text{14}}\) From Section 3.4, the expected number of trips per quarter is \(\exp(\mu z_o \cdot \Lambda(V_0(0)))\). The latter term may be computed via
over quarters. We summarize this calculation in the left panel of Figure 9. The number of expected trips per year declines monotonically from about 2.3 trips/year at a distance of 0.1 miles (-2 in log scale) to 1.5 trips/year at 400 miles (6 in log scale), primarily due to the influence of distance in the brand consideration arrival process. In the right panel of Figure 9, we plot the expected trip budget (expenditure) as a function of distance to the retail outlet. This quantity, which is unconditional with respect to channel choice, shows a modest decline (a maximal difference of roughly $4 over a 400+ mile range) in total expenditures as distances increase. The origin of this effect is the greater marginal utility associated with online purchases, which in turn come from the larger estimated product shock variances in the online channel – at larger distances, online purchases dominate and, all else equal, less expenditure is required to achieve a given level of utility in the online channel. In sum, shopping frequency and trip expenditures both increase as the distance retail decreases. Since these demand components are unconditional with respect to channel and sufficient to determine annual per capita expected revenues, we conclude that the channels serve as net complements, with the strength of the complementarity decreasing as retail distance increases.

Although Figure 9 is sufficient to establish that the net effect of increasing retail channel availability is demand-stimulative, we must incorporate channel choices in order to determine whether the channels function as strict complements. If the channels are strict complements, there should be no reduction in

\[ E[b^*(\text{distance})] = \int b^*(\varepsilon_0|\text{distance})dF(\varepsilon_0) \]

numerically using adaptive quadrature for the integration over \( \varepsilon_0 \).

\( b^*(\varepsilon_0) \) is calculated by solving the first order condition (18) using simulated expected product utility values as described in Appendix B.
the level of online revenues as retail availability increases. Mathematically, the distinction we are drawing is that the channels are net complements if \( \frac{\partial R_{\text{total}}}{\partial d} < 0 \) and strict complements when \( \frac{\partial R_{\text{online}}}{\partial d} \leq 0 \), where \( R \) represents revenues and \( d \) represents retail store distance. We begin exploration of this issue by plotting the probability of choosing the retail channel in Figure 10. Recall that we modeled the retail choice probability as conditional upon both a budget level and as a function of retail store distance. The left panel of Figure 10 shows this joint distribution, with the x axis representing budget amounts and the level curves representing distances to the retail outlet. As may be seen, retail probabilities are decreasing in both budget amounts and distance, which reflects the patterns in the data shown in Figure 3 and the descriptive regression of Section 2.2.3. In the right panel of Figure 10, we integrate over the budget distribution and plot the unconditional probability of choosing the retail channel as a function of distance. This plot clearly shows a rapid decline in the retail choice probability as a function of distance.

![Figure 10](image)

**Figure 10**: Probability of retail trip conditional on distance and expenditure (left) and distance alone (right), homogeneous model

In Figure 11, we summarize all the preceding demand effects by plotting expected annual revenues by channel as a function of distance to the retail outlet in log scale (left panel) and linear scale (right panel). These plots demonstrate that the channels do not function as strict complements because online revenues decline uniformly as the distance to the retail outlet declines. That is, switching from online to retail increases more rapidly than commensurate increases in the number of shopping trips and revenue per trip. Inspection

---

\[ \text{We compute expected annual revenues (as a function of distance) as the product of the expected number of trips per year and the expected expenditure per trip. We obtain annual trip rates by aggregating over quarterly trips rates, the calculation of which was previously described. To obtain channel-specific expected revenues, we multiply the expected number of trips per year by the expected expenditure given the channel choice, i.e., } E[b|c] = \int b(e_0)Pr[c|b]dF(e_0). \]
of the left panel of Figure 11 reveals the nature of the tradeoff: the slope of the expected retail revenue curve as a function of retail distance is approximately twice that of the expected online revenue curve (in magnitude), implying that decreasing retail store distance generates about twice as much incremental retail revenue as corresponding online revenue losses due to cannibalization. An alternate means of summarizing the net effect is the elasticity of annual expected revenue with respect to retail store distance, which we obtain by regressing a log-transformation of the total revenue curve (black uppermost line in the left panel of Figure 11) on log distance to the retail outlet. Our estimate of this elasticity is -0.055, implying a 10% reduction in retail store distance increases annual revenues by 0.55%. By extension, the effect of reducing store distances by a fixed amount will induce much larger revenue changes among customers close to the store than those at greater distances. This effect is clearly seen in the right panel of Figure 11, which uses a linear scale for retail store distance. The figure indicates that the “crossover” point at which channels contribute equally to total expected revenues occurs at approximately 25 miles from the retail store, and that the rate of change in channel revenue patterns diminishes rapidly in linear distance beyond this radius.

Before concluding this section, we present plots analogous to the left panel of Figure 11 for the two segment model. These plots are provided in Figure 12 below. We see that the pattern of net but not strict channel complementarity also holds for the heterogeneous specification. The plots also make clear that, while relatively uncommon, segment 1 customers generate far more revenue for the firm than their segment 2 counterparts.
5.2 Retail location selection

The choice of where to locate retail stores is a difficult problem that many firms face. In this section, we demonstrate how our model may inform such decisions. Our method is to generate a comprehensive set of potential entry locations and to calculate the expected incremental revenue generated from locating a store at each of these locations. To maintain high spatial resolution yet keep the exercise tractable, we discretize the set of potential entry locations as Census 2010 block group centroids in the continental United States, and thus evaluate approximately 216,000 candidate locations.

For each candidate location, we compute a revenue generation index (defined as $\Gamma$) using a two step process. In the first step, we determine the set of markets (other block groups) that would be impacted by entry at the candidate location. This set is generated by identifying block groups for which the distance to the nearest retail outlet would be reduced as a result of entry at the candidate location. In the second step, for each market in the impacted set, we forecast demand under two conditions: first assuming no entry were to occur, and then assuming it does occur. Taking the difference in these forecasts generates a prediction of the incremental contribution of entry.

Our revenue generation index incorporates revenue contributions from both new and existing customers. As shorthand notation, let $r$ be the expected annual revenue for the market, which is computed as described in the preceding section. Let $\rho$ be the expected number of new customer arrivals per year, which is calculated from our estimates in Section 2.2.1. Note that the values of $r$ and $\rho$ will be different under the “with entry”
and “no entry” conditions for each of the potentially affected markets. Further, let \(e\) represent the number of existing customers in the market, \(\delta\) the firm’s discount factor, and \(N\) the planning horizon (in years). Under this setup, the market will generate a discounted revenue stream of \(\delta r(e + \rho) + \delta^2 r(e + 2\rho) + \ldots + \delta^N r(e + N\rho)\). Thus, the revenue generation index (\(\Gamma\)) is given by:

\[
\Gamma = \sum_{n=1}^{N} \delta^n r(e + \rho n) = \delta \frac{(1 - \delta^N)}{1 - \delta} re + \frac{\delta (1 - ((1 - \delta)N + 1)\delta^N)}{(1 - \delta)^2} \rho r
\]

For the purpose of our application, we assume a discount rate of 10% (thus \(\delta = \frac{1}{1+0.1} = 0.91\)) and a planning horizon of \(N = 10\) years.

Since neighboring block groups within a metro region often yield similar revenue indices, it is difficult to discern the relative desirability of broad regions using an ungrouped ranking of potential entry locations. We thus summarize the results of the experiment at the MSA (Metropolitan Statistical Area) level. To do this, we assign a MSA-level index using the maximal revenue index among the block groups in that MSA. These indices are reported in Table 7.

<table>
<thead>
<tr>
<th>Rank</th>
<th>MSA</th>
<th>Revenue index ($MM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Minneapolis-St. Paul, MN-WI</td>
<td>2.09</td>
</tr>
<tr>
<td>2</td>
<td>San Francisco-Oakland-San Jose, CA</td>
<td>1.76</td>
</tr>
<tr>
<td>3</td>
<td>St. Louis, MO-IL</td>
<td>1.43</td>
</tr>
<tr>
<td>4</td>
<td>Miami-Fort Lauderdale, FL</td>
<td>1.30</td>
</tr>
<tr>
<td>5</td>
<td>Columbus, OH</td>
<td>1.21</td>
</tr>
<tr>
<td>6</td>
<td>Orlando, FL</td>
<td>1.20</td>
</tr>
<tr>
<td>7</td>
<td>Cincinnati-Hamilton, OH-KY-IN</td>
<td>1.03</td>
</tr>
<tr>
<td>8</td>
<td>West Palm Beach-Boca Raton, FL</td>
<td>1.02</td>
</tr>
<tr>
<td>9</td>
<td>Washington-Baltimore, DC-MD-VA-WV</td>
<td>1.02</td>
</tr>
<tr>
<td>10</td>
<td>New York, Northern New Jersey, Long Island, NY-NJ-CT-PA</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 7: Top 10 model-predicted entry locations (MSA level)

By comparing the set of cities in Table 7 to the store location map in Figure 1, it is clear the experiment suggests that desirable locations may be found in markets with an existing brand presence as well as in “virgin territory” (although greater weight is placed on the latter). Lending some face validity to our experiment, we note that since the end of the observation window, the brand has opened store locations in three of the seven “new territory” locations identified in Table 7. To demonstrate that the counterfactual results can help inform both local and wide-geography choices, we generate revenue heat maps for the top ranked MSA,
Minneapolis-St. Paul, which are provided in Figure 13. The map identifies the most desirable location by the darkest shade of red, as well as close substitutes in the event entry is not possible in that specific location (due, for example, to zoning restrictions or lack of available rental space).

![Revenue Generation Index](image)

**Figure 13: Revenue generation index by Census block group, Minneapolis-St. Paul MSA**

### 5.3 Channel-based price discrimination

Another important issue for multi-channel firms is whether or not to implement channel-specific pricing. In principle, consumers may be segmented according to their propensity to shop in either the online or retail channels, and prices may be tailored to these segments in order to extract more surplus than would be possible under a uniform pricing scheme. Fully optimizing prices in this manner is a formidable computational task, as it would involve determining the best price for each product category in the online channel as well as at each retail location. Rather than fully solve this high-dimension problem, our purpose in this section is to demonstrate how our model estimates may be applied towards this end and to explore the predicted demand response to variation in channel-specific pricing policies.

We begin our exploration of channel-based pricing by first reporting channel-specific price elasticity measures for each of the six categories in our study. To illustrate the implications of the different model stages for substitution patterns, we report three measures. When calculating each set of measures, the baseline prices for the categories are assumed to be those in the final period of observation. In Table 8, we report the price elasticity of expected category demand (i.e., the category quantity index), conditioned
upon a channel choice and an expenditure level of $141 (the sample average). That is, we compute $\epsilon_{ck} = \frac{\partial E[q^*_k|c,b=\$141]}{\partial p_k}$, where the expectation is taken with respect to channel-specific product shocks ($\epsilon$, $\nu$). These elasticities reflect the rate of substitution into other categories as the price of the focal category changes, assuming the budget amount and channel remain fixed. As may be seen by comparing Table 8 to the estimation results in Table 5, for a given category the pattern of price response across channels tracks the estimates of the corresponding channel/category shock variances ($\sigma_{\nu,ck}$) – a higher variance leading to a smaller (in magnitude) price response. With the exception of category 1, demand is more responsive to price changes in the retail channel, which is consistent with the firm’s decision to price lower in the retail channel.

<table>
<thead>
<tr>
<th>segment category</th>
<th>homogeneous 1</th>
<th>homogeneous retail</th>
<th>2 segment 1</th>
<th>2 segment retail</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-2.30</td>
<td>-2.21</td>
<td>-2.18</td>
<td>-2.11</td>
</tr>
<tr>
<td>2</td>
<td>-2.84</td>
<td>-3.44</td>
<td>-2.79</td>
<td>-3.33</td>
</tr>
<tr>
<td>3</td>
<td>-2.51</td>
<td>-2.88</td>
<td>-2.46</td>
<td>-2.98</td>
</tr>
<tr>
<td>4</td>
<td>-2.55</td>
<td>-2.83</td>
<td>-2.47</td>
<td>-2.80</td>
</tr>
<tr>
<td>5</td>
<td>-2.28</td>
<td>-2.53</td>
<td>-2.22</td>
<td>-2.39</td>
</tr>
<tr>
<td>6</td>
<td>-2.44</td>
<td>-5.43</td>
<td>-2.62</td>
<td>-6.04</td>
</tr>
</tbody>
</table>

Table 8: Price/quantity elasticities conditional upon channel and expenditure of $141 (sample mean)

While indicative of how consumers with a fixed budget may substitute among categories upon arrival at a store or the website, the elasticities in Table 8 do not account for potential substitution across channels or with the outside option. To obtain a measure of the unconditional demand response to price changes, we compute $\epsilon_{ck} = \frac{\partial E[q^*_k|c,b,p_k=0]}{\partial p_k}$, where the the expectation is taken with respect to the channel-specific product shocks ($\epsilon$, $\nu$), the channel choice shocks ($\eta$), the outside good shock ($\varepsilon_0$), and the probability of $L$ purchase occasions per year. The latter quantity is obtained by integrating the purchase incidence rate, $\Lambda(V_0(0)) \exp(\mu z)$, over the quarter dummies contained in $\mathbb{z}$. This measure incorporates the effect of price changes on shopping frequency and total expenditure, in addition to channel and product choices. Given our specification of channel utility and the consideration arrival rate, these elasticities will be a function of the distance to the nearest retail outlet. We therefore summarize these elasticity computations graphically as a function of retail store distance in Figure 14. For brevity, Figure 14 reports elasticities for the homogeneous model only (the 2 segment model elasticities follow a similar pattern). Several points are worth noting with respect to Figure 14. First, as might be expected, compared
to comparable estimates in Table 8, the magnitude of demand response is significantly lower when richer substitution patterns including the outside option are allowed. A second intuitive observation is that the magnitude of demand response is decreasing in retail store distance for products in the retail channel and increasing in store distance for products in the online channel. The rate of these changes varies by category, with category 1 displaying the greatest rate of change in distance.

![Figure 14: Unconditional quantity index price elasticity measures, homogeneous model](image)

The effect of retail store distance is even more readily apparent in our final set of elasticity measures, which consider the impact of price changes on total expected (annual) revenues. These elasticities, which are summarized in Figure 15, are computed analogously to those in Figure 14, with the optimal quantity \( q^*_k \) in the expectation replaced by total revenues \( E \left[ \sum_{k=1}^{K} q^*_k p_{ek(c,b)} \right] \). An important difference between Figures 14 and 15 is that the latter in some sense controls for variation in the baseline purchase rate of categories. Note, for example, that in the retail channel, category 6 has a relatively large quantity elasticity (Figure 14) but a small total revenue elasticity (Figure 15), reflecting that large percentage changes in quantity need not translate into large revenue changes when the average purchase quantities are small.
Figures 14 and 15 demonstrate that the demand response to price changes differs by category over the range of retail store distances. This pattern of demand response, coupled with knowledge of firm costs by category and the distribution of customer proximity to retail outlets, can be used to construct pricing policies that optimize firm profits. Since we do not directly observe firm costs, we must we employ additional assumptions in order to proceed with the analysis. For the purpose of this exercise, we assume that: a) the firm sets online product category prices using the “keystone” markup rule of 100% over wholesale price, and b) category costs are constant across channels. These assumptions allow us to back into category costs and compute contribution margins under alternative pricing schemes.\footnote{An alternative approach might assume we observe optimal pricing by the firm, which would allow us to back out marginal costs from our demand estimates and the supply model. Apart from the fact that our discussions with the firm suggest that a markup rule is a closer approximation of their price setting behavior, assuming optimality here would leave little scope for the counterfactual of interest. The firm of course could employ more accurate estimates of category-specific costs and thereby obtain more informative predictions of counterfactual pricing policy.}

We further assume that the firm wishes to maintain its current online pricing policy and seeks a simple markup or markdown rule to generate retail prices from online prices. Thus, our experiment will attempt to maximize profits by varying the ratio of retail to online prices, holding online prices fixed at their values in the final period of observation. Recognizing that price demand response and hence profits depend upon retail store distance, our method is to compute total expected profits by integrating expected profits conditional upon retail distance over the distribution of customer retail distances. That is, our objective function is:

\[ \pi = E\left[ \sum_{k=1}^{K} q_{ck}^* (p_{ck} - \omega_k) | c, b, s, d \right \} Pr[c|b, s, d] Pr[b > 0, s] Pr[L|s, z, d] Pr[s] Pr[d] \]  

\[ (26) \]
In (26), \( w \) represents category costs computed as described above, \( d \) denotes the retail store distance (in log scale), and \( s \) represents the segment membership (we use our two segment results to evaluate this counterfactual). The expectation in (26) is taken over all econometric shocks \( (\varepsilon, \nu, \eta, \varepsilon_0) \), seasonality dummies entering the purchase arrival rate \( (z) \) and the distribution of retail store distances \( (d) \). To facilitate computation of the expectation with respect to retail distance, we fit the empirical distribution of customer retail proximity (in log scale) to a normal distribution, as shown in Figure 16. As may be seen, a lognormal distribution provides a reasonable fit to the empirical distribution of customer retail distances.

![Figure 16: Distribution of customer retail proximity in the final period](image)

We summarize our experiment graphically in Figure 17 below. In addition to searching for the optimal retail/online price ratio for the fitted distribution of customer retail distances, we explore the sensitivity of the results to assumptions about the distribution of customer proximity to retail locations. Specifically, we scale the mean of the distance distribution by a factor of 1/2 and by 2 and recompute expected profits – the three level curves in Figure 17 correspond to these different scenarios. For each scenario, the optimal price ratio is indicated by a red ‘x’ on the corresponding curve. As may be seen, total expected profits are highest when customers are located closer to retail store, ceteris paribus. However, it is also notable that while profits go up, the firm’s ability to price discriminate across channels goes down, since retail discounting under optimal pricing is larger with a higher average distance to retail outlets. Optimal retail prices range between 67% and 70% of online prices, which is somewhat smaller than the average retail/online price ratio observed in the data (approximately 83%). While this result suggests potential sub-optimal pricing by the firm, we caution that a more rigorous treatment of product costs (as would be available to the firm) is
required to firmly establish this finding.

Figure 17: Pricing counterfactual

6 Conclusion

Our intended contributions from this research are threefold. Substantively, the paper adds to a small but growing literature that examines the demand implications of operating a mixture of online and retail channels. We document evidence that these channel formats function as net complements for existing customer demand. Our estimates imply increasing retail channel availability through market entry provides revenue gains that outweigh losses from cannibalized online sales by approximately a 2:1 ratio. While this ratio remains relatively constant as function of distance to the retail outlet, the magnitude of incremental contributions to demand decreases in distance such that material gains are found within approximately a 25 mile radius of the entering location. We further document evidence that a retail presence plays a critical role in the acquisition of new customers, providing another important benefit to operating a retail channel.

From a methodological standpoint, we develop an integrated, utility-based model that jointly predicts purchase incidence and channel choices in addition to purchase quantities in multiple categories. In the context of our application, the structural formulation allows us to draw inference on the mechanisms by which channels interactions contribute to observed patterns of demand. We further demonstrate how a direct utility formulation can be extended to a multi-stage decision process wherein expenditures are endogenized as a function of channel and product preferences. Finally, we develop a computationally and econometrically efficient algorithm to jointly estimate the multi-stage model.
We also contribute to managerial practice by providing a set of tools to address challenging problems such as where to locate new retail stores and how to optimize product prices in a multi-channel environment. Our retail entry experiment demonstrates the computational feasibility of exhaustively exploring potential entry locations at high spatial resolution, a process that could be applied iteratively to obtain even more precise predictions of optimal locations. While our experiment focused on revenue predictions due to data limitations, the firm could easily extend the application to incorporate knowledge of product marginal costs and even recurring store fixed costs to rank entry locations on the basis of profit potential.

Our analysis is of course limited in certain respects. One consideration is that we do not observe product inventories and thus cannot control for cross-channel variation in assortments or stockout rates, which the model will rationalize via the channel-specific product variances. While this limits our ability to interpret the category variance terms purely as “fit uncertainty”, it is not problematic for our applications under the assumption that this aspect of the operating environment remains unchanged. In addition, we abstract from dynamic considerations such as state dependence in channel choices and learning about product categories over time. While we believe that our focus on seasonal goods to some extent limits the scope for dynamics (e.g., the ability to learn is restricted by the fact that products within the observed categories are perpetually changing), accounting for these aspects would be an interesting and challenging extension to the current work.
References


Appendices

A Details of the product model

A.1 Kuhn-Tucker conditions

As stated in the text, the Kuhn-Tucker conditions are: a) \( \frac{\partial \mathcal{L}}{\partial \epsilon_k} \leq 0 \) (stationarity), b) \( \epsilon_k \geq 0 \) (primal feasibility), and c) \( \frac{\partial \mathcal{L}}{\partial \epsilon_k} \epsilon_k = 0 \) (complementary slackness). Formulating the Lagrangian in terms of expenditures yields:

\[
\mathcal{L} = \sum_{k=1}^{K} \Psi_{ck} \gamma_k \log \left( \frac{e_k}{p_{ck} Y_k} + 1 \right) - \lambda \left( \sum_{k=1}^{K} e_k - b \right)
\]

Recall that without loss of generality, we assume the first category is chosen (categories may be re-labeled to assure this is the case) and thus \( \epsilon_1^* > 0 \). By condition (c) \( \frac{\partial \mathcal{L}}{\partial \epsilon_1} = 0 \), which implies the Lagrange multiplier is given by \( \lambda = \frac{\Psi_1}{p_{c1} \left( \frac{e_1^*}{p_{c1} Y_1} + 1 \right)} \). For other chosen categories (where \( \epsilon_k^* > 0 \)), we must have \( \frac{\partial \mathcal{L}}{\partial \epsilon_k} = \frac{\Psi_{ek}}{p_{ek} \left( \frac{e_k^*}{p_{ek} Y_k} + 1 \right)} - \lambda = \frac{\Psi_{ek}}{p_{ek} \left( \frac{e_k^*}{p_{ek} Y_k} + 1 \right)} - \frac{\Psi_1}{p_{c1} \left( \frac{e_1^*}{p_{c1} Y_1} + 1 \right)} = 0 \). Rearranging, taking logarithms, and substituting in from equation (8b) yields:

\[
\beta_{ek} + v_{ek} - \log \left( \frac{e_k^*}{p_{ek} Y_k} + 1 \right) - \log (p_{ek}) \epsilon_k = \beta_{c1} + v_{c1} - \log \left( \frac{e_1^*}{p_{c1} Y_1} + 1 \right) - \log (p_{c1}) + \epsilon_1
\]

\[
V_{ek} + \epsilon_k = V_{c1} + \epsilon_1
\]

Where \( V_{ek} \equiv \beta_{ek} + v_{ek} - \log \left( \frac{e_k^*}{p_{ek} Y_k} + 1 \right) - \log (p_{ek}) \). For non-chosen categories, a similar derivation yields \( V_{ek} + \epsilon_k < V_{c1} + \epsilon_1 \). These conditions correspond to those stated in equation (11) in the text.

A.2 Likelihood function

This derivation follows Bhat (2008). We begin by evaluating the integrals with respect to the \( \epsilon_k \) terms in equation (12):

\[
Pr(\bar{c}^*|c, b) = \left| J_{E_M \rightarrow E_M} \right| \int_{V} \int_{\epsilon_1 = -\infty}^{+\infty} \int_{\epsilon_{M+1} = -\infty}^{+\infty} \cdots \int_{\epsilon_K = -\infty}^{+\infty} \prod_{j=1}^{M} \frac{1}{\sigma_e} \left( \frac{V_{c1} - V_{ej} + \epsilon_1}{\sigma_e} \right) \prod_{j=M+1}^{K} A \left( \frac{V_{c1} - V_{ej} + \epsilon_1}{\sigma_e} \right) \frac{1}{\sigma_e} \lambda \left( \frac{\epsilon_1}{\sigma_e} \right) d\epsilon_1 d\bar{F}(\bar{v})
\]

\[
= \left| J_{E_M \rightarrow E_M} \right| \int_{V} \int_{\epsilon_1 = -\infty}^{+\infty} \left\{ \prod_{j=M+1}^{K} \frac{1}{\sigma_e} \lambda \left( \frac{V_{c1} - V_{ej} + \epsilon_1}{\sigma_e} \right) \right\} d\epsilon_1 d\bar{F}(\bar{v})
\]
where $\lambda(x) = e^{-(x+e^{-x})}$ and $\Lambda(x) = e^{-e^{-x}}$ are the standard extreme value pdf and cdf, respectively.

Substituting and rearranging yields:

$$
Pr(\hat{e}^* | c, b) = \left|J_{eM \to e_M}\right| \frac{1}{\sigma_e^{M-1}} \int \prod_{j=2}^{M} \exp\left(\frac{V_{c1} - V_{cj}}{\sigma_e}\right) \times \int_{-\infty}^{+\infty} e^{-\sum_{j=1}^{K} \exp\left(-\frac{(V_{c1} - V_{cj} + \varepsilon_1)}{\sigma_e}\right) + \frac{1}{\sigma_e} e^{-\frac{\varepsilon_1}{\sigma_e}} d\varepsilon_1 dF(\tilde{V})
$$

Let $u = e^{-\frac{\varepsilon_1}{\sigma_e}}$. Then, the inner integral in the second line above may be written:

$$
\int_{-\infty}^{+\infty} e^{-\sum_{j=1}^{K} \exp\left(-\frac{(V_{c1} - V_{cj} + \varepsilon_1)}{\sigma_e}\right) + \frac{1}{\sigma_e} e^{-\frac{\varepsilon_1}{\sigma_e}} d\varepsilon_1 = \int_{u=0}^{+\infty} e^{-u\sum_{j=1}^{K} \exp\left(-\frac{(V_{c1} - V_{cj})}{\sigma_e}\right) + \frac{1}{\sigma_e} e^{-\frac{\varepsilon_1}{\sigma_e}} u^{M-1} du
$$

Substituting back in and simplifying yields:

$$
Pr(\hat{e}^* | c, b) = \left|J_{eM \to e_M}\right| \frac{(M-1)!}{\sigma_e^{M-1}} \int \prod_{j=2}^{M} \exp\left(\frac{V_{c1} - V_{cj}}{\sigma_e}\right) \left(\sum_{j=1}^{K} \exp\left(-\frac{(V_{c1} - V_{cj})}{\sigma_e}\right)\right)^{-M} dF(\tilde{V})
$$

Next, consider the elements of the Jacobian:

$$
J_{ij} = \frac{\partial e_{i+1}}{\partial e_{j+1}} (V_{c1} - V_{c,i+1} + \varepsilon_1) \quad \forall i, j = 1, ..., M - 1
$$

$$
= \frac{\partial}{\partial e_{j+1}} \left(\beta_{c1} + v_{c1} - \log\left(\frac{e_{j+1}^*}{p_{c1} Y_1} + 1\right) - \log(p_{c1}) - \left(\beta_j, v_{c,j+1} + v_{c,j+1} - \log\left(\frac{e_{j+1}^*}{p_{c,j+1} Y_j} + 1\right) - \log(p_{c,j+1})\right) + \varepsilon_1\right)
$$

$$
= \frac{1}{b - \sum_{k=1}^{K} e_{k+1}^*} + \frac{1}{p_{c1} Y_1} + \frac{1}{p_{c,i+1} Y_{i+1}} + \frac{1}{p_{c,j+1} Y_{j+1}} I(i = j)
$$

$$
= \frac{1}{e_{i+1}^* + p_{c1} Y_1} + \frac{1}{e_{j+1}^* + p_{c,j+1} Y_{j+1}} I(i = j)
$$

The Jacobian determinant is then:

$$
|J_{eM \to e_M}| = \left|\begin{array}{cccc}
\frac{1}{e_{1}^* + p_{c1} Y_1} + \frac{1}{e_{1}^* + p_{c2} Y_2} & \frac{1}{e_{1}^* + p_{c1} Y_1} & \cdots & \frac{1}{e_{1}^* + p_{cM} Y_M} \\
\frac{1}{e_{1}^* + p_{c2} Y_2} & \frac{1}{e_{1}^* + p_{c1} Y_1} + \frac{1}{e_{1}^* + p_{c3} Y_3} & \cdots & \frac{1}{e_{1}^* + p_{cM} Y_M} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{e_{1}^* + p_{cM} Y_M} & \frac{1}{e_{1}^* + p_{cM} Y_M} & \cdots & \frac{1}{e_{1}^* + p_{cM} Y_M} + \frac{1}{e_{M}^* + p_{cM} Y_M}
\end{array}\right| = \left(\prod_{j=1}^{M} \frac{1}{e_{j}^* + p_{cj} Y_j}\right) \left(\sum_{j=1}^{M} e_{j}^* + p_{cj} Y_j\right)
$$

43
Substituting this expression for the Jacobian into (27) above gives the fully simplified likelihood:

\[
Pr(\tilde{e}^*|c, b) = \frac{(M - 1)!}{\sigma_c^{M-1}} \left( \prod_{j=1}^{M} \frac{1}{e_j^* + p_c y_j} \right) \left( \sum_{j=1}^{M} e_j^* + p_c y_j \right)^{-M} \int_{v} \left( \prod_{j=1}^{M} e_{v_j/\sigma_c} \right) \left( \sum_{j=1}^{K} e_{v_j/\sigma_c} \right)^{-M} dF(\tilde{v})
\]

B Details of the channel choice model

B.1 Computation of optimal quantities and expected utilities

We first demonstrate the calculation of optimal quantities when the set of chosen categories is known and then describe the procedure to determine the set of chosen categories. For a given draw of the product shocks \((e, n)\), without loss of generality assume the first \(M\) categories are chosen. For chosen categories, the Kuhn-Tucker condition \(\frac{\partial L}{\partial e_k} = 0\) implies

\[
y_{ck} p_{ck} \left( \frac{\Psi_{ck}}{p_{ck} e_k^* + 1} \right) = \lambda_k \Rightarrow e_k^* = \kappa_k p_{ck} \left( \frac{\Psi_{ck}}{p_{ck} e_k^* + 1} \right). 
\]

Next, we use the budget constraint equation to eliminate the Lagrange multiplier \(\lambda\):

\[
b = \sum_{k=1}^{M} e_k^* = \sum_{k=1}^{M} \kappa_k p_{ck} \left( \frac{\Psi_{ck}}{p_{ck} e_k^* + 1} \right). 
\]

Solving for \(\lambda\) gives

\[
\lambda = \frac{\sum_{k=1}^{M} \kappa_k y_{ck}}{b + \sum_{k=1}^{M} \kappa_k y_{ck}}. 
\]

Then, substituting this expression for \(\lambda\) back into the optimal expenditure equation above gives:

\[
e_k^* = \kappa_k p_{ck} \left( \frac{\Psi_{ck} \left( b + \sum_{k=1}^{M} \kappa_k y_{ck} \right)}{p_{ck} \left( \sum_{k=1}^{M} \kappa_k y_{ck} \right)} - 1 \right). 
\]

Equivalently, the optimal quantities are:

\[
q_k^* = \frac{e_k^*}{p_{ck}} = \kappa_k \left( \frac{\Psi_{ck} \left( b + \sum_{k=1}^{M} \kappa_k y_{ck} \right)}{p_{ck} \left( \sum_{k=1}^{M} \kappa_k y_{ck} \right)} - 1 \right) \quad (28)
\]

To find the optimal set of chosen categories, we use the “enumerative” algorithm of Pinjari and Bhat (2010). The method is based on the insight that the price normalized baseline utilities \(\frac{\Psi_{ck}}{p_{ck}}\) of chosen (non-chosen) goods are greater than or equal to (less than) the Lagrange multiplier, which is an implicit function of the set of chosen categories. The algorithm to predict quantity choices thus proceeds as follows:

1. Take a draw of the product shocks \((e, n)\)

2. Compute the price normalized baseline utilities for all \(K\) categories: \(y_k = \frac{\Psi_{ck}}{p_{ck}}\)

3. Sort the categories from highest to lowest \(Y\); denote quantities sorted in this order with a tilde, e.g. \(\tilde{Y}\)

4. Iteratively compute the Lagrange multiplier, assuming the first \(m\) categories are chosen: \(\lambda^m = \frac{\sum_{j=1}^{m} \gamma_j y_{cj}}{b + \sum_{j=1}^{m} \gamma_j p_{cj}}\)

5. Determine the number of chosen categories by the relation: \(M = \sum_{m=1}^{K} I(\tilde{Y}_m \geq \lambda^m)\)
6. Compute the optimal quantities for the chosen categories as
\[ \tilde{q}_m^* = \tilde{q}_m \left( \Psi_{ak} \left( \frac{b + \sum_{k=1}^{M} \tilde{g}_m \tilde{p}_m}{\tilde{p}_m \left( \sum_{n=1}^{M} \tilde{g}_n \Psi_{cn} \right)} \right) - 1 \right) \]
for \( m \leq M \) and \( \tilde{q}_m^* = 0 \) for \( m > M \)

7. Invert the sort order to restore the original category ordering, yielding \( q_k^* \)

This procedure may be vectorized so that solutions may be sought for simultaneously for all the draws (across all observations), yielding a highly efficient polynomial time algorithm (proportional to \( N \cdot D \cdot K^2 \) operations, where \( N, D \) and \( K \) are respectively the number of observations, draws and categories).

Once optimal quantities are in hand, it is a simple matter to evaluate the utility distribution by substituting optimal quantities into equation (8a).

C Details of the trip budget model

C.1 Kuhn-Tucker conditions

From section 3.3, the Lagrangian for the trip budget problem is
\[ \mathcal{L} = \Psi_0 q_0 + g(b) - \lambda (q_0 + b - w), \]
where \( \Psi_0 = \exp (\beta_0 + \varepsilon_0) \) and \( g(b) = \sigma_{\eta} \log \left( \exp \left( \frac{V_1(b)}{\sigma_{\eta}} \right) + \exp \left( \frac{V_2(b)}{\sigma_{\eta}} \right) \right) \). Since some allocation to the outside good is considered essential, we have \( \frac{\partial \mathcal{L}}{\partial q_0} = 0 \), which implies \( \lambda = \Psi_0 \). Therefore, for the trip budget we have:
\[
\frac{\partial g}{\partial b} - \Psi_0 = 0 \quad \text{if } b^* > 0 \\
< 0 \quad \text{if } b^* = 0
\]

Rearranging, substituting for \( \Psi_0 \) and taking logs gives equation (18) in the text.

C.2 Likelihood function

From Section 3.3, the objective is to compute \( Pr(b^* \mid b^* > 0) = \frac{Pr(b^* \cap b^* > 0)}{Pr(b^* > 0)} = \frac{Pr(b^*)}{Pr(b^* > 0)} \). Given the one to one mapping of the outside good shock (\( \varepsilon_0 \), distributed \( EV(0, \sigma_{\varepsilon_0}) \)) to the budget \( b^* \), the numerator of the preceding equation is simply \( |J_{\varepsilon_0 \rightarrow b} \lambda \left( \frac{V_0(b^*)}{\sigma_{\varepsilon_0}} \right) \), where \( \lambda (\cdot) \) is the extreme value pdf. Similarly, from the KT conditions (18), \( Pr(b^* > 0) = 1 - Pr(b^* = 0) = 1 - Pr(\varepsilon_0 > V_0(0)) = Pr(\varepsilon_0 \leq V_0(0)) = \Lambda \left( \frac{V_0(0)}{\sigma_{\varepsilon_0}} \right) \), where \( \Lambda (\cdot) \) is the extreme value cdf. Finally, for positive budgets, the Jacobian term is computed from (18) as
\[
J = \frac{\partial \varepsilon_0}{\partial b} = \frac{\partial}{\partial b} \left( \log \left( \frac{\partial g}{\partial b} \right) - \beta_0 \right) = \left( \frac{\partial g}{\partial b} \right)^{-1} \left( \frac{\partial^2 g}{\partial b^2} \right), \]
where the derivatives of \( g \) are evaluated at \( b^* \).

Putting these sub-calculations together gives equation (19) in the text.
C.3 Computation of $\frac{\partial g}{\partial b}$ and $\frac{\partial^2 g}{\partial b^2}$

Here we provide an explanation of how to compute the $\frac{\partial g}{\partial b}$ and $\frac{\partial^2 g}{\partial b^2}$ terms using the draws of $\varepsilon$ and $\nu$, which are generated when computing the expected product utility in the channel choice model. Recall that simulation of the expected product utilities requires that optimal quantities for each draw to be computed, and that categories can be ordered for each draw such that the first $M_d$ categories are the chosen categories.

We introduce several shorthand notations to describe the computation. First, let $u_{cdk}(b)$ represent the utility contribution from category $k$ for draw $d$ assuming the chosen channel is $c$. The optimal quantity computed in via equation (28), may be substituted into (8a) to express the utility contribution in terms of the draw values (via the $Y$ terms), the model parameters, and the budget value $b$. This utility contribution and its first two derivatives with respect to the budget amount may then be written:

$$u_{cdk}(b) = \Psi_{cdk} \gamma_k \log \left( \frac{q_{dk} + 1}{\gamma_k} \right) = \Psi_{cdk} \gamma_k \log \left( \frac{\Psi_{cdk} \left( b + \sum_{j=1}^{M_d} \gamma_j p_{cj} \right)}{p_{ck} \left( \sum_{j=1}^{M_d} \gamma_j \Psi_{cdj} \right)} \right)$$

$$u'_{cdk}(b) = \frac{\partial u_{cdk}(b)}{\partial b} = \gamma_k p_{ck} \left( \sum_{j=1}^{M_d} \gamma_j \Psi_{cdj} \right) \left( b + \sum_{j=1}^{M_d} \gamma_j p_{cj} \right)^{-1}$$

$$u''_{cdk}(b) = \frac{\partial^2 u_{cdk}(b)}{\partial b^2} = -\gamma_k p_{ck} \left( \sum_{j=1}^{M_d} \gamma_j \Psi_{cdj} \right) \left( b + \sum_{j=1}^{M_d} \gamma_j p_{cj} \right)^{-2}$$

Next, we introduce $Eu_c$ as shorthand notation for the expected product utility from spending $b$ in channel $c$. $Eu_c$ and its first two derivatives with respect to the budget amount are:

$$Eu_c = E[u_q^+ | b, c] = \frac{1}{D} \sum_{d=1}^{D} \sum_{k=1}^{M_d} u_{cdk}$$

$$Eu'_c = \frac{\partial Eu_c}{\partial b} = \frac{1}{D} \sum_{d=1}^{D} \sum_{k=1}^{M_d} u'_{cdk}$$

$$Eu''_c = \frac{\partial^2 Eu_c}{\partial b^2} = \frac{1}{D} \sum_{d=1}^{D} \sum_{k=1}^{M_d} u''_{cdk}$$

Finally, let $Vu_c$ be the product utility variance from spending $b$ in channel $c$. $Vu_c$ and its first two
derivatives with respect to the budget amount are:

\[
Vu_c = \text{Var}[u(q^*|b, c)] = \frac{1}{D} \sum_{d=1}^{D} \left( \sum_{k=1}^{M_d} u_{cdk} \right)^2 - \frac{1}{D^2} \left( \sum_{d=1}^{D} \sum_{k=1}^{M_d} u_{cdk} \right)^2
\]

\[
Vu'_{c} = \frac{\partial Vu_c}{\partial b} = \frac{2}{D} \sum_{d=1}^{D} \left( \sum_{k=1}^{M_d} u'_{cdk} \right) \left( \sum_{k=1}^{M_d} u_{cdk} \right) - \frac{2}{D^2} \left( \sum_{d=1}^{D} \sum_{k=1}^{M_d} u_{cdk} \right) \left( \sum_{d=1}^{D} \sum_{k=1}^{M_d} u'_{cdk} \right)
\]

\[
Vu''_{c} = \frac{\partial^2 Vu_c}{\partial b^2} = \frac{2}{D} \sum_{d=1}^{D} \left( \sum_{k=1}^{M_d} u''_{cdk} \right) + \frac{2}{D^2} \left( \sum_{d=1}^{D} \sum_{k=1}^{M_d} u'_{cdk} \right) - \frac{2}{D^2} \left( \sum_{d=1}^{D} \sum_{k=1}^{M_d} u'_{cdk} \right)^2
\]

Using these notations, it is straightforward (albeit tedious) to show that \( \frac{\partial g}{\partial b} \) and \( \frac{\partial^2 g}{\partial b^2} \) may be written:

\[
\frac{\partial g}{\partial b} = \text{Pr}(c = 1 \mid b) \left( Eu'_1 + \phi_3Vu'_1 \right) + \text{Pr}(c = 2 \mid b) \left( Eu'_2 + \phi_3Vu'_2 \right)
\]

\[
\frac{\partial^2 g}{\partial b^2} = \frac{1}{\sigma_\eta} \text{Pr}(c = 1 \mid b)\text{Pr}(c = 2 \mid b) \left( Eu''_1 - Eu'_2 + \phi_3Vu'_1 - \phi_3Vu'_2 \right)^2
\]

\[
+ \text{Pr}(c = 1 \mid b) \left( Eu''_1 + \phi_3Vu''_1 \right) + \text{Pr}(c = 2 \mid b) \left( Eu''_2 + \phi_3Vu''_2 \right)
\]