

Structural Analysis of Multi-Channel Demand

Scott Shriver and Bryan Bollinger *

April 1, 2015

Abstract

In this paper, we propose a structural framework to study multi-channel demand. Our model explains a comprehensive set of demand outcomes as a function of prices and retail store proximity, including the frequency with which consumers shop, how much they spend per purchase occasion, whether they buy from the online (web) or retail channel, and how they allocate expenditures among multiple product categories. We allow channels to convey different amounts of information about product categories, which in turn affects a consumer's expected utility from purchasing in a particular channel. For example, physical inspection of goods in the retail channel can provide information about product fit and feel that is difficult to assess in the online channel. Another distinguishing feature of our model is that shopping trip expenditures are endogenously determined in the first phase of a multi-stage budgeting process, where consumers allocate their income by trading off utility for the outside option and the expected utility from optimal channel and category expenditure choices with the focal brand. A key methodological contribution of the paper is to advance a highly efficient algorithm to compute optimal expenditures, which facilitates joint estimation of the model parameters by maximum simulated likelihood.

We estimate the model using the purchase histories of approximately 10,000 randomly selected customers from a firm that uses both online and retail channels to sell directly to consumers. The firm doubled its retail footprint over our two year observation window, providing a rich source of customer-specific variation in retail store proximity that we leverage to identify the demand effects of interest. We find evidence of channel complementarity through increased overall shopping frequency as the distance to retail outlets decreases, accompanied by increased substitution from online to retail formats. Our estimates imply a 10% reduction in retail store distance increases existing customer annual revenues 0.53%, by increasing retail revenues 1.96% and decreasing online revenues by 1.43%. In a series of counterfactual experiments, we demonstrate how our model can be used as a decision tool for managers 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, Duke University, the University of California at San Diego, the University of Southern California, and the University of Rochester for their comments. We are also grateful to Avi Goldfarb and Kitty Wang for useful discussions regarding the paper. All errors are our own.

1 Introduction

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 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. Furthermore, managers need tools to optimize their channel strategy, including methods to select promising physical store locations and to potentially set channel-specific prices.

In this paper, we provide an empirical framework to study multi-channel demand and demonstrate how it can inform key aspects of channel strategy. We take a structural approach to facilitate counterfactual experiments that allow us to quantify the demand response to channel-specific pricing policies and retail entry in locations that are not directly observed in data. Our model explains a comprehensive set of demand outcomes as a function of prices and retail store proximity, including the frequency with which consumers shop, how much they spend per purchase occasion, whether they buy from the online (web) or retail channel, and how they allocate expenditures among multiple product categories. Unlike analyses of aggregate channel revenues, this comprehensive micro-foundation approach can reveal the specific mechanisms by which channels influence demand, providing insights that may be used to fine-tune elements of the marketing mix. The framework is quite general and easily adaptable to any setting where historical customer transaction data are available.

The model contains a number of novel features. First, to our knowledge, ours is the first to jointly predict purchase incidence, shopping trip expenditures, channel choice, and multi-category purchase quantities in a unified utility-maximizing framework. A key aspect of this formulation is that shopping trip expenditures are endogenously determined in the first phase of a multi-stage budgeting process, where consumers allocate their income by trading off utility for the outside option and the expected utility from optimal channel and category expenditure choices with the focal brand. The 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 accessibility. A second distinguishing feature is that we allow channels to convey different amounts of information about product categories, which in turn affects a consumer’s expected utility from purchasing in a particular channel. For example, physical inspection of goods in the retail channel can provide information about product fit and feel that is difficult to assess in the online channel. We permit these information differences to vary across categories, recognizing for example that physical inspection of a suit is critical to fit assessment, while of lesser importance when evaluating products such as T-shirts (where sizing is fairly standardized).

Our estimation procedure also incorporates a noteworthy methodological contribution. In order to accommodate prediction of purchase quantities in multiple categories, we employ a direct utility specification as in the literature on multiple-discreteness (simultaneous choice of multiple alternatives) and marketing models of demand for variety (e.g., Wales and Woodland, 1983; Hanemann, 1984; Kim et al., 2002; Bhat, 2008).¹ While direct utility specifications are convenient for formulating a model of multiple-discreteness, they can be computationally burdensome for model prediction (as required for counterfactual evaluation or for joint estimation of multi-stage decision processes) because they require solving for optimal quantities, for every simulation draw of model unobservables. Historically, researchers have used computationally expensive constrained non-linear optimization techniques to solve for optimal quantities, limiting the appeal of direct utility demand systems. Using insights from Pinjari and Bhat (2010), we devise a highly efficient algorithm to solve for optimal quantities that replaces non-linear optimization with matrix operations that may be performed simultaneously over multiple simulation draws. In our context, this innovation enables us to jointly estimate our multi-stage demand system via maximum simulated likelihood. Further, the tractability of this approach demonstrates that direct utility specifications can be extended to richer modeling contexts than previously encountered in the literature.

Data for our empirical work comes from a speciality retailer that sells to directly customers through its e-commerce website and retail stores. We estimate the model using the purchase histories of 10,242 randomly selected customers over a two-year period, during which time we observe 29,130 unique purchase transactions. Two key aspects of the data are that we observe the entry of many new retail stores (the firm expanded from 37 to 75 stores during our study) and know customer home locations with great geographic accuracy (by Census block). Collectively, these features allow us to determine a customer's proximity to the nearest retail store at any point in time, a rich source of customer-specific variation in the mixture of available channels, which we leverage to identify the demand effects of interest.

We find evidence of channel complementarity through increased overall shopping frequency as the distance to retail outlets decreases, accompanied by increased substitution from online to retail formats. Our estimates imply channels function as net complements: a 10% reduction in retail store distance increases existing customer annual revenues 0.53%, by increasing retail revenues 1.96% and decreasing online revenues by 1.43%. The channels contribute equally to total revenues at a distance of approximately 30 miles from the retail store, and the share

¹Broadly, direct utility models are distinguished from indirect utility models by the treatment of purchase quantities. By definition, the indirect utility for an alternative is the utility obtained at the consumer's optimal consumption quantity for that alternative. An indirect utility model assumes the functional form of utility at optimal purchase quantity levels, whereas in a direct utility setup, optimal purchase quantities must be derived from the consumer's optimization problem. Models of pure discrete choice (i.e., single-discreteness) typically abstract from purchase quantities altogether and adopt an indirect utility approach. A direct utility specification is more natural in multiple-discreteness settings, where purchase quantities drive the utility trade-offs among different bundles of goods.

of retail sales increases rapidly at smaller distances. In supplemental analysis outside our structural model, we further document evidence that a retail presence plays a critical role in the acquisition of new customers: a 10% decrease in retail distance corresponds to a 8.6% increase in the number of new customers per quarter. We also find that product category price elasticities vary considerably across channels and as a function of store distance, highlighting that retail entry can influence the composition of consumer shopping baskets. A final insight of note is that our heterogeneity specification with two latent classes suggests 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. This finding lends empirical support for the conventional wisdom of the Pareto principle in sales (80% of revenues deriving from 20% of the customer base).

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 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

While broadly our paper is related to the literature on multi-channel customer management, the substantive question of to what extent channels function as substitutes or complements has primarily been explored within the literature on new channel introduction.² Researchers have studied the effect of channel introduction in multiple industries including newspapers (Deleersnyder et al., 2002; Gentzkow, 2007), music (Biyalogorsky and Naik, 2003) and apparel (Ansari et al., 2008). Most extant studies source data from the period of rapid expansion of e-commerce (late 1990s to early 2000s) and consequently investigate 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. Our view is that additional empirical work in this area is timely for three reasons. First, the adoption of e-commerce technologies is now mainstream among both consumers and major brands (at least in developed nations like the U.S.), suggesting shopping preferences may have shifted appreciably. Second, there is a lack of research that decomposes channel revenue effects in a manner that can reveal the underlying

²See Neslin et al. (2006) for a comprehensive survey of the literature on multi-channel customer management.

mechanisms and facilitate counterfactual evaluation. Finally, it has yet to be established that prior empirical generalizations are invariant to the order in which channels are introduced. With most firms now operating e-commerce channels, the dominant question for channel strategy is when and where to introduce retail outlets. Compared to the introduction of the e-commerce channel, which is a one-time event affecting the entire customer base, retail entry (or exit) has different implications for the firm (e.g., the ability to target specific geographic markets) and for the analyst seeking to quantify the effect of channel introduction (e.g., the ability to leverage within-subject variation in retail store proximity). To our knowledge, only three recent 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.³

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

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 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 demand outcomes, 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:

³Forman et al. (2009) also study the impact of retail stores on online sales, but in the context of a competitive environment. 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 results may be less pronounced. One possible explanation is that sales differences are driven by consumer brand preferences (Amazon vs. local store) rather than channel formats per se. Also, the results of our analysis are not directly comparable because Forman et al. (2009) focus only on top-selling books and measure sales by the book's sales ranking in the local market – thus, the relationship to total shopping trip expenditures (our measure of sales) is unclear.

(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, increasing sales through new customer acquisition. 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 slightly stronger (statistical and economic) net effect of channel complementarity. 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 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 setup of our model links it to the the literature on store choice, which similarly considers shopping format choices and the purchase of baskets of goods. 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 in several recent marketing studies (e.g., Bell et al., 2013; Soysal and Zentner, 2014; Dzyabura et al., 2015). 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, Ho, and Tang, 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. The standard practice in the store choice literature has been to treat expenditure

levels as exogenously determined – for example, Bell and Lattin (1998) use a pre-estimation calibration to obtain the probability that consumers are either “large basket” or “small basket” shoppers, and Bell, Ho, and Tang (1998) assume budgets arise implicitly from an unobserved shopping list. By contrast, we endogenize the trip budget decision as a function of expected optimal channel and product selection, which allows us to infer how expenditures respond to changes in prices or retail channel accessibility.

2 Data

Data for the study comes from a North American speciality retailer that sells to customers exclusively through its e-commerce website (the online channel) and network of retail stores (the retail channel).⁴ The brand sells a variety of apparel products including clothing, footwear and accessories. As is typical in the apparel industry, product assortments vary over time, reflecting fashion trends and seasonal patterns of demand (e.g., warmer clothes in winter). However, there is limited variation in the product assortment across channels – the firm’s product database indicates that fewer than 0.1% of products are offered exclusively via one channel.⁵ Since our objective is to predict 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 – such an analysis would involve more than 70,000 unique SKUs, many of which would not enter future assortments. Therefore, using methods that we describe in Section 2.1 below, for our empirical work we aggregate demand into six distinct product categories, which correspond to the top level of the brand’s product hierarchy. These category definitions are sufficiently broad so as not to be season-specific (e.g., a footwear category would encompass winter boots as well as summer sandals) but narrow enough to share common requirements regarding fit assessment (e.g., the need to try on shoes does not vary greatly across specific shoe styles but is presumably different from the need to try on pants). Given this aggregation scheme, our analytical framework will characterize consumer demand for product categories and explore how the mixture of available channels influences that demand.

With approximately 87% of the US population now using the Internet,⁶ the mixture of channels available to a consumer is effectively determined by the consumer’s proximity to the retail outlets operated by the brand. Observing variation in retail outlet proximity is therefore critical to identifying the different mechanisms (e.g.,

⁴Confidentiality agreements preclude us from disclosing the identity of the brand or revealing precise descriptions of the products in their portfolio.

⁵Operational issues such as product delivery schedules and stock-outs may result in temporary assortment variation across the channels, at least with respect to some retail stores. Lacking data on store and online channel inventories, we cannot characterize such assortment variation precisely.

⁶Source: <http://www.internetlivestats.com/internet-users>.

impact on purchase frequency and channel choices) by which channels influence demand. Our dataset provides a rich source of such variation. 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 high precision (roughly 1/3 of a mile accuracy). In addition to the cross-sectional variation in retail proximity derived from the initial spatial distribution of customers and stores, we observe significant within-subject variation due to the entry of new stores – during our two-year period of study, the brand expanded its retail footprint from 37 retail outlets to 75. We explore the effect of retail distance on various demand outcomes through a series of descriptive regressions in Section 2.2.

2.1 Preparation and summary

We observe individual-level transaction data for a sample of 10,242 customers residing in the continental United States, which is randomly drawn from the firm's database. The window of observation spans July 2010 through June 2012. Over this period, we observe a total of 29,130 transactions, each corresponding to a separate purchase occasion. Transaction records indicate the date and channel format, the quantities and prices of individual SKUs purchased, as well as a return indicator for each SKU. Reflecting the one-to-one correspondence between a purchase occasion and channel format, we conceptualize a transaction as the outcome of a shopping trip, where the "trip" can involve physical travel to a retail store or simply a visit to the e-commerce website. As our principal interest lies in the net economic contribution of channels to demand, we omit returned items from our computation of trip expenditures and aggregation of category demand.

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 (a 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 1. 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. While much of the variation in store distance is cross-sectional (the between-subject standard

deviation is 74.30 miles), the within-subject variation induced by observed retail entry events is also substantial, with a standard deviation of 28.38 miles.

	observations	mean	std dev	min	max
quarter	57,577	5.13	2.21	1	8
trips	57,577	0.51	1.22	0	36
store distance	57,577	42.65	81.44	0.10	499.74

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

Similar to previous treatments in the marketing and economics literatures (e.g., Gordon et al., 2013; Chevalier et al., 2003), we compute category (k) and channel (c) specific price indices for a quarter (t) as geometric expenditure-weighted means of SKU-level prices:

$$p_{tck} = \exp \left(\frac{\sum_{j:J_t \in k} w_j \log(p_{tcj})}{\sum_{j:J_t \in k} w_j} \right) \text{ where : } w_j = \frac{\sum_c \sum_t e_{cjt}}{\sum_c \sum_j \sum_t e_{cjt}} \quad (1)$$

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 (summed over consumers) for the same SKU. Similarly, J_t represents the set of SKUs offered in quarter t . The applied weights, which are the fraction of total observed expenditures attributable to a given SKU, naturally reflect more popular products in the resulting index. We use fixed (as opposed to time-varying) weights to avoid inducing endogeneity with the quantity indices we construct to measure customer category demand on purchase occasions.⁷ As is readily apparent from Figure 1, for most categories, prices are slightly lower in the retail channel. A SKU-level analysis of prices reveals that these differences stem from more aggressive discounting in the retail channel, primarily end of season markdowns.

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 for consumer i in quarter t on trip l in category k is:

$$q_{itlk} = \frac{\sum_{j \in k} e_{itlj}}{p_{tck}} \quad (2)$$

The shopping trip data summarized in Table 3 indicates that the average per-trip expenditure is approximately \$141 and that the 57% of observed trips are in the retail channel. Unit sales tend to be highest for category 1 and

⁷We note that despite using fixed weights, our price indices can vary over time even when SKU prices do not change because of the changing SKU composition of the categories. We have explored a number of alternative indexing schemes, including restricting attention to SKUs available throughout the observation window, as well as using quantity (rather than expenditure) based weights. These choices have little influence on our qualitative results.

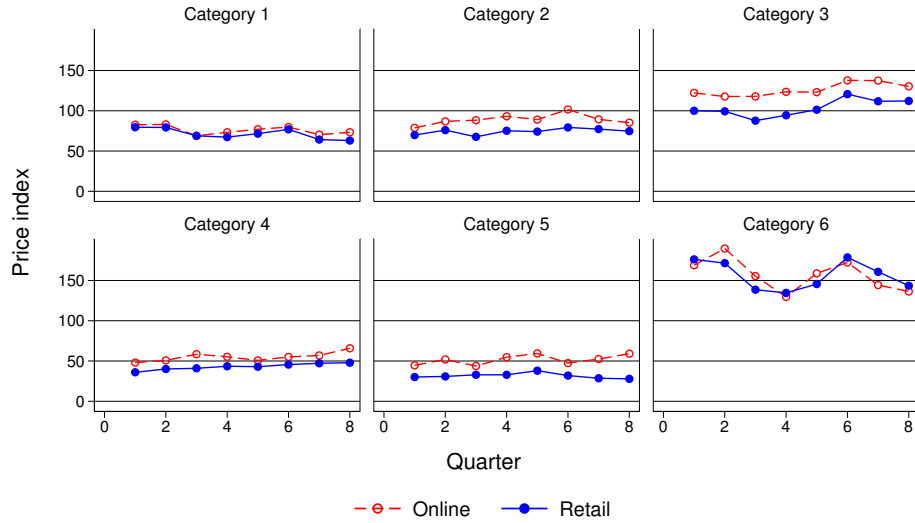


Figure 1: Channel-specific category price indices

	observations	mean	std dev	min	max
total expenditure	29,130	140.84	153.70	0.13	3215.96
channel	29,130	1.57	0.50	1	2
quantity 1	29,130	0.63	1.09	0	18.44
quantity 2	29,130	0.22	0.60	0	11.47
quantity 3	29,130	0.21	0.56	0	10.26
quantity 4	29,130	0.49	1.00	0	20.73
quantity 5	29,130	0.46	1.21	0	41.87
quantity 6	29,130	0.06	0.28	0	7.16

Table 3: Lower level (trip) summary statistics

lowest for category 6, which is the highest priced category.

We examine the relationship between expenditures and channel formats in Figure 2, 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. Combining these per-trip expenditure figures with channel choice frequencies, 52% of total sample revenues come from the online channel and 48% from retail. We explore variation in category choices by channel in Figure 3, 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.

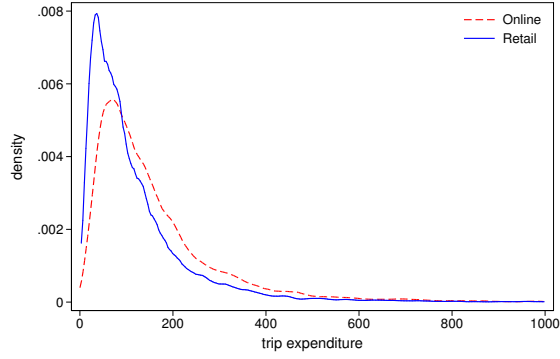


Figure 2: Kernel density estimates of trip expenditure by channel

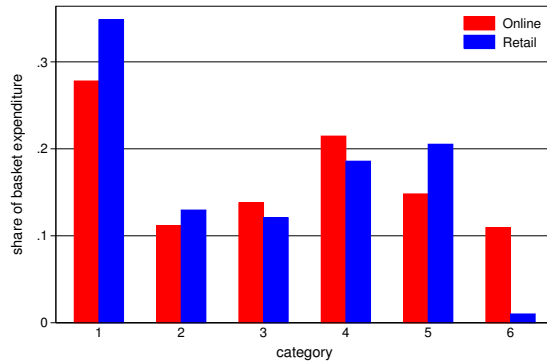


Figure 3: Average category expenditure shares by channel

2.2 Descriptive analyses

In this section, through descriptive regressions we explore how distance to the retail outlet affects demand among existing customers along three dimensions: purchase frequency, the total expenditure per shopping trip, and the channel chosen for the shopping trip. Our findings here guide the formulation of our structural model in Section 3.

Throughout this section and in our model development, we define the variable d as the log distance to the nearest retail outlet.

2.2.1 Purchase frequency

To investigate the impact of retail outlet distance on overall shopping incidence rates, we define purchase frequency in terms of the number of purchase occasions per quarter, where a purchase occasion may correspond to either the online or retail channel. We model the number of purchase occasions for customer i in quarter t , L_{it} , as a Poisson arrival process and use the customer/quarter panel data to estimate the model. The specification includes

bi-level (individual, quarter) fixed effects (δ_i, μ_t) to control for unobservables – the effect of retail store distance (α) is thus identified by deviations from individual-specific average purchase incidence rates, after controlling for time trends common to all individuals. Formally, the specification is:

$$L_{it} \sim \text{Poisson}(\rho_{it}), \log(\rho_{it}) = \alpha d_{it} + \delta_i + \mu_t \quad (3)$$

The estimate of α from this regression is $\hat{\alpha} = -0.1040$ with a standard error of 0.0164, which is significant below the 1% level. We conclude that the increased proximity to a retail outlet (smaller store distance) has a significant positive impact on shopping incidence rates, presumably due to increased top of mind awareness of the brand.

2.2.2 Expenditure level

Although we have demonstrated that, *ceteris paribus*, retail outlet proximity increases customer purchase frequency, to infer the net effect on total expenditure levels with the brand, we must also consider the potential impact of retail distance on the average expenditure per trip. To explore this issue, we use the trip-level data and model the expenditure for the l 'th trip by customer i in quarter t , b_{itl} , using a log-linear model with bi-level (individual, quarter) fixed effects (δ_i, μ_t) :

$$b_{itl} = \alpha d_{it} + \delta_i + \mu_t + \varepsilon_{itl} \quad (4)$$

The estimate of α from this regression is $\hat{\alpha} = -0.0123$ with a standard error of 0.0148. The sign of $\hat{\alpha}$ is consistent with expenditures increasing at closer retail distances; however, the effect 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 material evidence that retail distance directly influences trip expenditure levels.

2.2.3 Channel choice

Finally, we analyze the effect of retail distance on channel format choices. We denote customer i 's channel choice for the l 'th trip in quarter t as c_{itl} , and following our previous convention, associate a channel choice of 2 with the retail channel (whereas a choice of 1 represents the online channel). We proceed by modeling the probability of a retail channel choice using a binary logit model, again including bi-level fixed effects:

$$\Pr(c_{itl} = 2) = \frac{\exp(\alpha d_{it} + \delta_i + \mu_t)}{1 + \exp(\alpha d_{it} + \delta_i + \mu_t)} \quad (5)$$

The estimated α 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.

2.2.4 Descriptive analysis summary

The descriptive analysis suggests two key effects of retail proximity on demand: an effect on the overall shopping frequency and an effect on the preferred channel for shopping. We therefore include direct dependencies on store distance in the utility functions governing these decisions when specifying in our structural model in Section 3 below. In contrast to the descriptive regressions, the structural model provides a unified utility framework that links decisions relating to shopping incidence, trip expenditure levels, channel choices and category expenditure allocations. The structural framework can therefore capture inter-relationships among these decisions and their mutual dependence on category prices, the informational content of channels, and retail store proximity.

3 Model

In this section, we develop our demand model for existing customers. When performing counterfactual experiments quantifying the benefits of retail entry in Section 5.2, we also incorporate the effect of retail outlet proximity on new customer acquisition. To streamline the model development, we defer further discussion of customer acquisition to Section 5.2.

Our proposed model for existing customer demand, outlined in Figure 4 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.⁸ Our 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. 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

⁸We warrant that in some instances consumers may shop in alternative ways than the sequence we have described. For example, consumers may first determine the set of product to be purchased in the form of a shopping list as in Bell, Ho, and Tang (1998). Since in general the consumer’s exact decision sequence cannot be identified from observable data, any complete model of multi-channel shopping behavior must employ comparable assumptions that will not apply in all circumstances. In this sense, our model serves as a standard against which other approaches may be compared. In this spirit, 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 due to substantially superior in-sample fit – results are available from the authors upon request.

future stages. Consumers are further assumed to know the distribution of future stage unobservables and to have rational expectations with respect to product prices in both channels.

To illustrate the model flow, consider an existing customer in a given quarter. During the quarter, the customer is assumed to consider a purchase with the focal brand R times, where R is a Poisson distributed variable with rate parameter r and each consideration event is associated with an income allotment of w .⁹ 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). For each consideration “arrival,” the consumer first determines the optimal allocation of her income allotment (w) to the brand’s products. Upon realization of the outside option utility shock (ε_0), the consumer sets the shopping trip budget with the focal brand (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 (η in the case of channels, and $\{\varepsilon, \nu\}$ 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 trip budget, in the next decision stage channel utility shocks are realized and the consumer chooses between the online and retail channels based upon her native preferences for the channel formats, 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 category 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.

⁹We conceptualize a consideration event as arising from two inter-related processes. First, a consumer’s need for apparel arrives according to a Poisson process, where every “need arrival” is associated with an income allotment of w . The rate parameter for the need arrival process may be heterogeneous and can depend on seasonal factors. Second, for each need arrival, the consumer considers the brand with some probability that depends on the distance to the nearest retail store (the assumption being that proximity to the retail store generates top of mind awareness of the brand). Following the logic developed formally in Section 3.4, the number of consideration events per quarter will be Poisson distributed, with a rate parameter equal to the rate parameter of the need arrival process times the conditional probability of consideration given need. Since the arrival of need for apparel and the conditional consideration of the focal brand given need are not directly observable (and hence cannot be separately identified), we capture both effects directly in the specification of the consideration arrival rate parameter r .

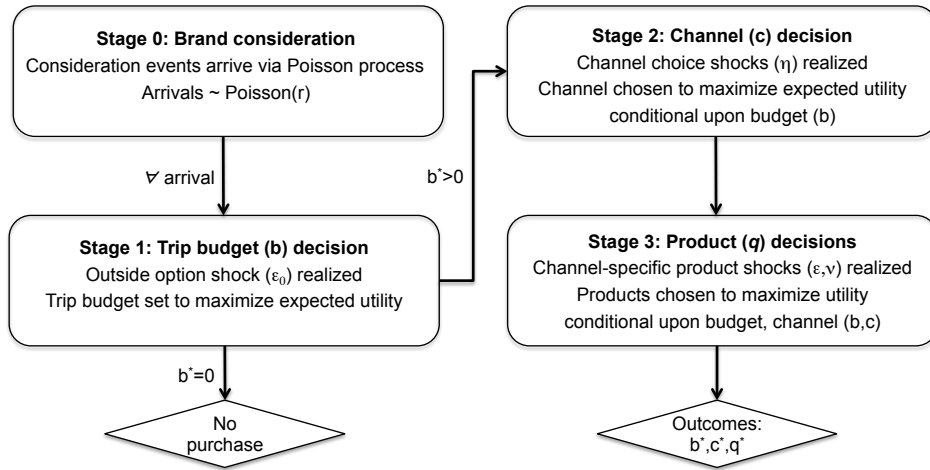


Figure 4: Model overview

3.1 Product choice model

As noted in our discussion of the data in Section 2, we model product purchases at the category level, recognizing that the set of product categories offered by the brand is stable over time. We assume consumers form preferences for the categories offered by the brand, but have uncertainty regarding the channel-specific match value of products within the category for their needs on any given shopping occasion. We use a multiple discrete/continuous model since it can predict what subset of imperfectly-substitutable alternatives are simultaneously chosen (the discrete component) and how much of each alternative is chosen (the continuous component), which is necessary in order to forecast revenues. Moreover, multiple discrete/continuous models are econometrically parsimonious (in terms of the number of estimated parameters) and derived from the economic theory of consumer utility maximization.

Accordingly, our approach is to develop a multiple discrete/continuous model of demand for product categories that we operationalize through the use of category price and quantity indices. Our specification extends the model of Bhat (2008) by incorporating latent classes for persistent unobserved heterogeneity.

3.1.1 Utility specification

We represent the dependent variable for the product model by \vec{q} , a vector containing the $K = 6$ category quantity indices associated with a consumer's purchase occasion. 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 \vec{q} on

the l 'th shopping trip in period t :

$$u(\vec{q}_{il}|c_{il}, b_{il}, i \in s) = \sum_{k=1}^K \Psi_{itlk} \gamma_k \log \left(\frac{q_{itlk}}{\gamma_k} + 1 \right) \quad (6a)$$

$$\Psi_{itlk} = \exp(\beta_{sk} + v_{itlk} + \varepsilon_{itlk}) \quad (6b)$$

$$\varepsilon_{itlk} \sim iid EV(0, \sigma_\varepsilon) \quad (6c)$$

$$v_{itlk} \sim iid N(0, \sigma_{vck}) \quad (6d)$$

Equation (6a) 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 utility in the quantity of each category. The role of the γ parameters are to alter the rates at which the consumer's utility satiates in category consumption – a higher γ implies a lower rate of satiation. The Ψ terms are the so-called “baseline” marginal utilities, because Ψ_k is the marginal utility obtained in the limit of consuming zero quantity of good k ($\lim_{q_k \rightarrow 0} \frac{\partial u}{\partial q_k}$). The parameterization of the Ψ terms is provided in equation (6b). We allow for heterogenous category preferences through the intercepts in the baseline marginal utilities, β_{sk} . Further, we introduce two stochastic terms (v and ε) into the baseline marginal utilities, the distributions of which will induce a likelihood for the observed product data. The v terms are normally distributed channel/category specific shocks that capture product fit and assortment information which is observed by the consumer upon visiting the channel but known only in distribution in prior decision stages. By assumption, the extreme value (EV) shocks (ε) are iid across all categories and channels – this compound shock specification is taken to reduce the computational burden of calculating 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.

3.1.2 Utility maximization and Kuhn-Tucker conditions

The reader will note the absence of prices from the model in equation (6), 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 obtained from solving the constrained optimization problem. In the current context, the consumer's problem is as follows (we suppress i , t and l subscripts where possible for expositional clarity):

$$\max_{q_1, \dots, q_K} u(\vec{q}|c, b) \text{ subject to: } \sum_{k=1}^K q_k p_{ck} = b, q_k \geq 0 \quad (7)$$

When deriving the likelihood function for this model, it proves convenient to work with an alternative but equivalent specification of equation (7). Specifically, we formulate the consumer's problem in terms of selecting category expenditures rather than product quantities, where the expenditure for category k is defined as $e_k = q_k p_{ck}$:

$$\max_{e_1, \dots, e_K} u(\vec{e} | c, b) \text{ subject to: } \sum_{k=1}^K e_k = b, e_k \geq 0 \quad (8)$$

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 expenditures will then satisfy the Kuhn-Tucker (KT) conditions: $\frac{\partial \mathcal{L}}{\partial e_k} \leq 0$, $e_k \geq 0$, $\frac{\partial \mathcal{L}}{\partial e_k} e_k = 0$ for all $k = 1, \dots, K$. Without loss of generality, assume that the first good is chosen.¹⁰ With this convention, we show in Online Appendix B.1 that the KT conditions for the consumer's problem reduce to:

$$V_{ck} + \varepsilon_k \begin{cases} = V_{c1} + \varepsilon_1 & \text{if } e_k^* > 0 \\ < V_{c1} + \varepsilon_1 & \text{if } e_k^* = 0 \end{cases} \quad (9)$$

where: $V_{ck} \equiv \beta_{sk} + v_{ck} - \log \left(\frac{e_k^*}{p_{ck} \gamma_k} + 1 \right) - \log(p_{ck})$

In equation (9) above, star superscripts are used to denote the fact that expenditures satisfying these relations are optimal.

3.1.3 Likelihood function

Note from equation (9) 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 ε space that can rationalize the observed expenditure pattern \vec{e}^* , are sufficient to derive the likelihood function. 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. We partition the $K \times 1$ vector of extreme value shocks ($\vec{\varepsilon}$) in two parts, with the $M \times 1$ vector $\vec{\varepsilon}_M$ corresponding to chosen alternatives and the $(K - M) \times 1$ vector $\vec{\varepsilon}_{\bar{M}}$ corresponding to non-chosen alternatives. With these conventions, the probability of observing expenditure pattern \vec{e}^* may be written (where $f(\cdot)$ represents a probability density function):

$$Pr(\vec{e}^* | c, b) = |J_{\vec{\varepsilon}_M \rightarrow \vec{e}_M}| \int_{\mathbf{v}} \int_{\varepsilon_1 = -\infty}^{\infty} \int_{\varepsilon_{M+1} = -\infty}^{V_{c1} - V_{c, M+1} + \varepsilon_1} \cdots \int_{\varepsilon_K = -\infty}^{V_{c1} - V_{cK} + \varepsilon_1} f(\vec{\varepsilon}) f(\vec{v}) d\vec{\varepsilon}_{\bar{M}} d\vec{v} \quad (10)$$

¹⁰With 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.

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 (9) above. These conditions map the probability density of the ε 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 ε space that is consistent with a “corner” solution of no category purchase, which generates the integrals over ε_{M+1} to ε_K . The additional integration over ε_1 appears because the KT conditions are conditioned upon ε_1 – integrating over ε_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 ε terms may be computed analytically, as may the Jacobian determinant. The resulting likelihood reduces to a remarkably simple expression (see Online Appendix B.2 for a full derivation):

$$Pr(\vec{e}^*|c, b) = \frac{(M-1)!}{\sigma_\varepsilon^{M-1}} \left(\prod_{j=1}^M \frac{1}{e_j^* + \gamma_j p_{cj}} \right) \left(\sum_{j=1}^M e_j^* + \gamma_j p_{cj} \right) \int_{\mathbf{v}} \left(\prod_{j=1}^M e^{v_{cj}/\sigma_\varepsilon} \right) \left(\sum_{j=1}^K e^{v_{cj}/\sigma_\varepsilon} \right)^{-M} f(\vec{v}) d\vec{v} \quad (11)$$

As pointed out by Bhat (2008), equation (11) 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: a) their native preferences for the channel formats, b) transportation costs to the retail channel (operationalized as the log distance to the nearest retail outlet, d_{it}), and c) their expectations of how much utility they will acquire by spending b on products in the respective channels. When forming these expectations, 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:¹¹

$$U_{itl1} = E[u(\vec{q}^*)|b_{itl}, c = 1] + \phi_3 \text{Var}[u(\vec{q}^*)|b_{itl}, c = 1] + \eta_{itl1} \equiv \bar{U}_{itl1} + \eta_{itl1} \quad (12a)$$

$$U_{itl2} = E[u(\vec{q}^*)|b_{itl}, c = 2] + \phi_3 \text{Var}[u(\vec{q}^*)|b_{itl}, c = 2] + \phi_2 d_{it} + \phi_{s1} + \eta_{itl2} \equiv \bar{U}_{itl2} + \eta_{itl2} \quad (12b)$$

$$\eta_{itlc} \sim \text{iid } EV(0, \sigma_\eta) \quad (12c)$$

In equation (12b), ϕ_{s1} captures the consumer's relative preference for the retail format, assuming she is a member of segment s . No corresponding parameter appears in equation (12a) because the level of online utility is normalized to zero for identification purposes. The parameter ϕ_2 captures the (dis)utility incurred from traveling to the retail outlet, which therefore only enters the utility of the retail channel choice in (12b). Finally, ϕ_3 , which captures risk preferences related to expected product utility, enters the utility for both channels since the variance of product utility for a given budget varies across these alternatives. Note that there is no coefficient on the expected product utility terms, which is instead captured through the estimated variance of the channel utility shocks (σ_η).¹²

The expectations and variances in (12) are taken with respect to the distributions of ε and v . 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. 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. We develop a highly efficient algorithm to solve for optimal product quantities and nest it within the maximization of the full model likelihood (in a manner similar to nested fixed point algorithms), which enables us to achieve full econometric efficiency by jointly estimating all model parameters. Our algorithm solves for optimal quantities exactly via matrix operations in polynomial time for the number of observations, categories and simulation draws. We provide details of the algorithm, which follows Pinjari and Bhat (2010), in Appendix A.

Consumers are assumed to choose the channel that gives the highest risk-adjusted utility. Given the iid

¹¹Note that this is an indirect utility specification, which is convenient given the purely discrete nature of channel choice. The specification can be interpreted as corresponding to some direct utility function that is maximized subject to budget constraints that capture channel-specific transaction costs, which could include shopping time costs as well as transportation costs for the retail channel. Implicitly, we assume the income allotment reserved for expenditures on apparel (w) is distinct from any income allotment related to transportation costs.

¹²As 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 $\frac{1}{\sigma_\eta}$.

extreme value assumption on η , 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):

$$Pr(c^* | b) = \int_{\eta} (U_c = \max[U_1, U_2]) dF(\vec{\eta}) = \frac{\exp\left(\frac{\bar{U}_c}{\sigma_{\eta}}\right)}{\exp\left(\frac{\bar{U}_1}{\sigma_{\eta}}\right) + \exp\left(\frac{\bar{U}_2}{\sigma_{\eta}}\right)} \quad (13)$$

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:

$$\begin{aligned} U_0(b_{itl}, q_{itl0}) &= \Psi_{itl0} q_{itl0} + E \left[\max \left\{ \bar{U}_{itl1}(b_{itl}) + \eta_{itl1}, \bar{U}_{itl2}(b_{itl}) + \eta_{itl2} \right\} \right] & (14a) \\ &= \Psi_{itl0} q_{itl0} + \sigma_{\eta} \log \left(\exp \left(\frac{\bar{U}_{itl1}(b_{itl})}{\sigma_{\eta}} \right) + \exp \left(\frac{\bar{U}_{itl2}(b_{itl})}{\sigma_{\eta}} \right) \right) \\ &\equiv \Psi_{itl0} q_{itl0} + g(b_{itl}) \end{aligned}$$

$$\Psi_{itl0} = \exp(\beta_{s0} + \varepsilon_{itl0}) \quad (14b)$$

$$\varepsilon_{itl0} \sim iid EV(0, \sigma_{\varepsilon0}) \quad (14c)$$

The first term on the right side of equation (14a) 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 η (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 β_0 to be segment specific.

The consumer solves the problem:

$$\max_{q_0, b} U_0(q_0, b) \text{ subject to : } q_0 + b = w, b \geq 0 \quad (15)$$

Some allocation of income to the outside good is considered essential. Under this assumption, it is shown in Online Appendix C that the Kuhn-Tucker conditions corresponding to (15) are:

$$V_0(b) \equiv \log \left(\frac{\partial g}{\partial b} \right) - \beta_{s0} \begin{cases} = \varepsilon_0 & \text{if } b^* > 0 \\ < \varepsilon_0 & \text{if } b^* = 0 \end{cases} \quad (16)$$

The conditions in (16) 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:

$$Pr(b^* | b^* > 0) = |J_{\varepsilon_0 \rightarrow e}| \frac{\lambda\left(\frac{V_0(b^*)}{\sigma_{\varepsilon_0}}\right)}{\Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right)} = \left| \frac{\partial^2 g}{\partial b^2} \right| \frac{\lambda\left(\frac{V_0(b^*)}{\sigma_{\varepsilon_0}}\right)}{\Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right)} \quad (17)$$

where $\lambda(\cdot)$ and $\Lambda(\cdot)$ are the standard extreme value pdf and cdf, respectively. The expression $\Lambda(V_0(0))$ enters (17) to account for the conditional probability that $b^* > 0$. A full derivation of (17) with the Jacobian expressed in terms of the simulation draws is provided in Online Appendix C.

3.4 Consideration arrival process

A consumer's consideration of the focal brand follows a Poisson arrival process with rate parameter $r_{it} = \exp(\mu_s z_{it})$, where z_{it} includes an intercept, the consumer's (log) distance to the retail store, and controls for seasonality (quarter dummies).¹³ Let R_{it} be the number of consideration events per period (quarter), so that $R_{it} \sim Poisson(r_{it})$. Given the iid assumption on the outside good shocks (ε_0), the probability of purchase given consideration is independent across consideration events and is equal to $\Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right)$.

To close the model, we must derive a likelihood function to recover the arrival rate parameters μ . 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" $Pr(b^* > 0) = \Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right)$). 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):

$$\begin{aligned} Pr[L] &= \sum_{R=L}^{\infty} Pr[L|R]Pr[R] = \sum_{R=L}^{\infty} \binom{R}{L} \left[\Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right) \right]^L \left[1 - \Lambda\left(\frac{V_0(0)}{\sigma_{\varepsilon_0}}\right) \right]^{R-L} \frac{e^{-R} r^R}{R!} \\ &= \frac{e^{-\Lambda(V_0(0))r} [\Lambda(V_0(0))r]^L}{L!} \end{aligned} \quad (18)$$

Thus, the observed number of purchase occasions is also Poisson distributed, with rate parameter $\Lambda(V_0(0))r$. This compound rate parameter provides a linkage between the purchase incidence rate and the structural parameters

¹³Poisson (and related negative binomial) models have been previously applied in a wide variety of settings to describe the timing of consumer/firm interactions, from early models of brand purchase incidence (e.g., Ehrenberg, 1959; Morrison and Schmittlein, 1988) to more recent application in auction settings (e.g., Yoganarasimhan, 2013, 2015). Whereas the standard approach has been to model event arrivals from the perspective of the firm, we model the arrival rate from the perspective of the consumer.

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, assuming consumer i is a member of segment s . Using equations (11), (13) and (17), this likelihood may be written:

$$\mathbb{L}_{it}|i \in s = Pr(b_{it}^*, c_{it}^*, \vec{e}_{it}^* | i \in s) = Pr(b_{it}^* | b_{it}^* > 0, i \in s) Pr(c_{it}^* | b_{it}^*, i \in s) Pr(\vec{e}_{it}^* | b_{it}^*, c_{it}^*, i \in s) \quad (19)$$

Given the independence across trips and using equation (18), the likelihood of the collection of trips for customer i in period t is:

$$\mathbb{L}_{it}|i \in s = Pr(L_{it}|i \in s) \left(\prod_{l=1}^{L_{it}} \mathbb{L}_{itl}|i \in s \right)^{(L_{it}>0)} \quad (20)$$

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

$$\mathbb{L}_i = \sum_{s=1}^S \left(\prod_{t=1}^{T_i} \mathbb{L}_{it}|i \in s \right) Pr(i \in s) \quad (21)$$

Where we specify the probability of segment assignment as:

$$Pr(i \in s) = \frac{\exp(\xi_s)}{\sum_{j=1}^S \exp(\xi_j)} \quad (22)$$

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

$$\mathbb{L}(\theta | Z) = \prod_{i=1}^I \mathbb{L}_i \quad (23)$$

where I is the number of unique customers, θ is the collection of all model parameters and Z represents all observed data.

The parameters are estimated via maximum simulated likelihood: $\hat{\theta} = \underset{\theta}{\operatorname{argmax}}(\log(\mathbb{L}))$.¹⁴ 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 \mathbf{v} in (11) and simulate the expected product utilities in (13) 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 \mathbb{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 (β_k), the satiation parameters (γ_k), the channel/category shock variances ($\sigma_{\eta ck}$) and the variance of the extreme values shocks common to all categories (σ_ϵ). The β_k values are primarily pinned down by variation in category purchase incidence rates, while the γ_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.¹⁵ 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 β_k parameters – for this, we set $\beta_1 = 0$ in all specifications. For the scale of utility, at most K separate variance terms may be estimated per channel, meaning that either σ_ϵ or one of the σ_{vck} values must be normalized. We choose to normalize σ_ϵ and estimate the full set of σ_{vck} values. As explained by Bhat (2008), σ_ϵ must be normalized to a small enough (but still positive) level such that $\frac{1}{\sigma_\epsilon}$ is larger than the actual demand response for all

¹⁴With 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.

¹⁵Note that in this model $\frac{1}{\sqrt{\sigma_{vck}^2 + \sigma_\epsilon^2}}$ is effectively the price coefficient for category k in channel c . Absent price variation, γ and β would still be identified while the σ terms would not.

categories.¹⁶ We found that normalizing $\sigma_\varepsilon = 0.25$ was sufficiently small in our setting.

In the channel choice model, the channel variance σ_η and the ϕ_3 terms are identified by variation in channel choices in response to the first two moments of the expected product utility distribution – these terms are separately identified provided the product model is identified. The ϕ_2 parameter is identified by both between and within subject variation in channel choice frequencies as a function of distance to the nearest retail outlet. The retail intercept is identified by average channel choice frequencies, 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, $\sigma_{\varepsilon 0}$ is identified by the response of trip expenditures to variation in the channel choice inclusive value (the log sum term in equation 14), while β_0 is identified by average trip expenditure levels 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 4. In the case of the two segment model, only the heterogeneous 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 μ_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 (β parameters) is: 1, 3, 2, 6, 4 and 5. The categories are ordered in terms of decreasing satiation rates (increasing γ) as: 6, 3, 2, 1, 4 and 5. For all but the first category, the online shock variance estimates (σ_{v1k}) exceed the retail shock variance estimates (σ_{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 5. We plot the corresponding utility distributions in Figure 6. These plots reflect, for example, the relative preference for category 6 in the online channel and slightly larger total

¹⁶As a practical matter, this translates to setting σ_ε small enough to get statistically significant estimates of the σ_{vck} .

expected utility from purchases in the online channel (given a fixed budget level).

Consistent with our descriptive results in Section 2.2.3, the channel utility estimates indicate a relative preference for the retail channel (ϕ_1 is positive) that decreases rapidly in the distance to the retail outlet (ϕ_2 is negative). The negative coefficient on ϕ_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 (σ_η) 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 2.2.1: consideration events (hence shopping trips) become less frequent as the distance to the retail store increases (μ_5 is negative), and we have a higher consideration rates in the 3rd and 4th quarters (μ_2 and μ_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).

We now turn to the two segment model estimates. As measured by the log likelihood value and the Bayesian Information Criterion (BIC), the fit of the two segment model is significantly higher and is thus our preferred specification. Segment sizes may be inferred from the latent class assignment parameter (ξ_1), which implies 93.8% of consumers are members of segment 2 and 6.2% belong to segment 1. 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. For this reason, we refer to segment 1 members as “high types” (and segment 2 as “low types”) – as will be shown in our analysis of expected revenue contributions to the firm (Section 5.1), segment 1 generates approximately 40% of firm revenues. 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 ϕ_1). Compared to the one segment model, the estimated channel shocks (σ_η) have smaller variance, indicating expected product utility plays a larger role in determining channel choices.

To further assess the overall fit of the model to the data, we compare the first two moments of 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 5. 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.

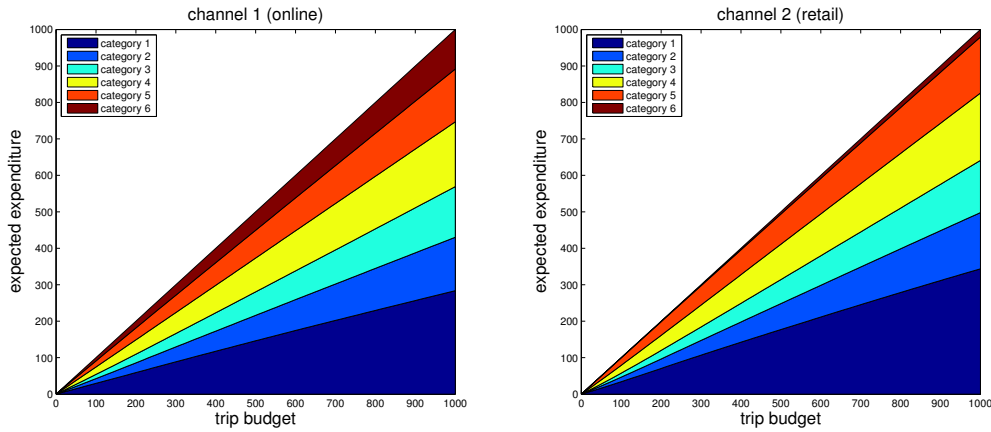


Figure 5: Expected category expenditures by channel, homogeneous model

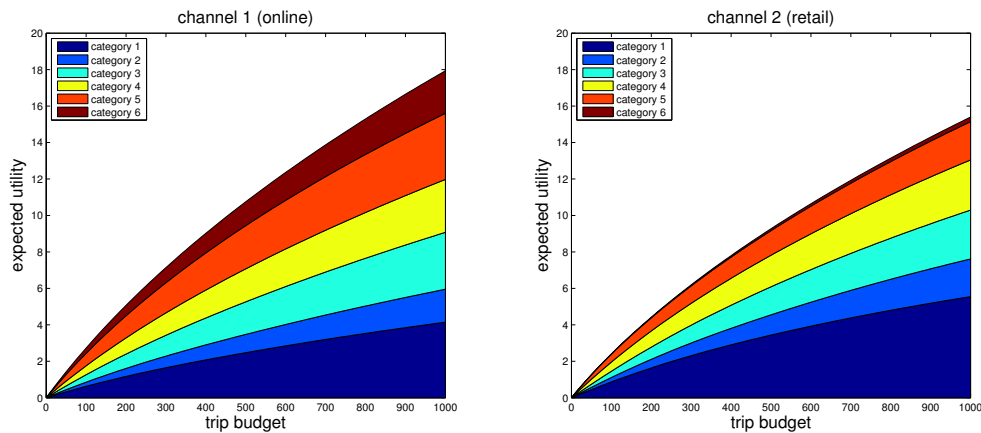


Figure 6: Category contributions to expected utility by channel, homogeneous model

			Homogeneous model		2-segment model			
					segment 1		segment 2	
			estimate	std err	estimate	std err	estimate	std err
<i>Product utility parameters</i>								
Category 1	satiation	γ_1	6.5724	0.0932	6.0052	0.0913	-	-
	online shock variance	σ_{v11}	0.4364	0.0090	0.5023	0.0094	-	-
	retail shock variance	σ_{v21}	0.4445	0.0057	0.4934	0.0069	-	-
Category 2	satiation	γ_2	6.2796	0.1256	5.7059	0.1082	-	-
	baseline utility intercept	β_2	-0.3106	0.0066	-0.3077	0.0081	-0.3079	0.0085
	online shock variance	σ_{v12}	0.5595	0.0108	0.5691	0.0108	-	-
Category 3	retail shock variance	σ_{v22}	0.4141	0.0088	0.4199	0.0090	-	-
	satiation	γ_3	5.8365	0.1285	5.2237	0.1138	-	-
	baseline utility intercept	β_3	-0.2147	0.0103	-0.3277	0.0120	-0.1718	0.0107
Category 4	online shock variance	σ_{v13}	0.8708	0.0112	0.8827	0.0116	-	-
	retail shock variance	σ_{v23}	0.6588	0.0113	0.6794	0.0115	-	-
	satiation	γ_4	7.3432	0.1078	6.6566	0.1038	-	-
Category 5	baseline utility intercept	β_4	-0.6400	0.0048	-0.6478	0.0076	-0.6159	0.0070
	online shock variance	σ_{v14}	0.7187	0.0088	0.7480	0.0098	-	-
	retail shock variance	σ_{v24}	0.4991	0.0062	0.5049	0.0078	-	-
Category 6	satiation	γ_5	7.4908	0.0887	6.7439	0.0844	-	-
	baseline utility intercept	β_5	-0.9639	0.0055	-0.9733	0.0076	-0.9518	0.0072
	online shock variance	σ_{v15}	0.9622	0.0093	0.9988	0.0100	-	-
Category 6	retail shock variance	σ_{v25}	0.4921	0.0075	0.5500	0.0078	-	-
	satiation	γ_6	5.1571	0.1076	4.9842	0.1068	-	-
	baseline utility intercept	β_6	-0.3182	0.0213	-0.2524	0.0171	-0.1146	0.0189
Category 6	online shock variance	σ_{v16}	0.9985	0.0184	0.8605	0.0153	-	-
	retail shock variance	σ_{v26}	0.3732	0.0205	0.2341	0.0218	-	-
	<i>Channel utility parameters</i>							
Retail	intercept	ϕ_1	3.3816	0.0765	2.6144	0.0826	2.8942	0.0930
	log distance to store	ϕ_2	-0.7762	0.0214	-0.6186	0.0231	-	-
Common	risk preference	ϕ_3	-0.0114	0.0004	-0.0162	0.0004	-	-
	channel shock variance	σ_η	2.2133	0.0635	1.8206	0.0693	-	-
<i>Expenditure utility parameters</i>								
Outside option	intercept	β_0	-2.1582	0.0626	-2.7870	0.0474	-2.5599	0.0518
	shock variance	σ_{E0}	1.1046	0.0252	0.9690	0.0227	-	-
<i>Consideration arrival parameters</i>								
	intercept	μ_1	2.7143	0.1018	3.0277	0.0692	1.4639	0.0850
	I(3rd quarter)	μ_2	0.3626	0.0146	0.3459	0.0151	-	-
	I(4th quarter)	μ_3	0.5236	0.0143	0.5220	0.0146	-	-
	I(1st quarter)	μ_4	-0.0001	0.0159	0.0690	0.0161	-	-
	log distance to store	μ_5	-0.1171	0.0022	-0.0862	0.0024	-	-
<i>Segment assignment parameters</i>								
	intercept	ξ_1			-	-	2.7126	0.0494
Implied segment size			100%		6.2%		93.8%	
Log likelihood			-401,721.2		-393,555.3			
# customer/quarter observations			57,577		57,577			
BIC			803,815.1		787,581.9			

Table 4: Model estimates

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

Table 5: Model fit statistics

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

A primary substantive question motivating the paper was to what extent channels function as demand substitutes or complements. Our approach to investigate this issue is to use the model estimates to calculate expected demand by channel as a function of distance to the nearest retail outlet, and to summarize how channel demand patterns change in response to variation in retail store distance. Specifically, we calculate expected annual revenues by channel as a function of retail store distance and use retail distance elasticities of annual revenues to summarize changes in demand patterns.¹⁷ For this exercise, we assume prices are those realized in the final period of observation – in Section 5.3, we relax this assumption and consider how demand responds to price changes.

For our analysis, we use our preferred specification that employs two latent classes for unobserved heterogeneity. We denote the annual revenue in channel c from a consumer in market segment s by \mathbb{R}_{sc} . Conceptually, the expected value of \mathbb{R}_{sc} may be calculated by simulating shopping trip outcomes for a large sample of individuals (each belonging to the segment of interest) in each of the four quarters comprising a calendar year, summing trip expenditures by channel, and averaging the result. To minimize simulation noise and computation time, in practice we take a hybrid approach that involves computing expectations analytically where possible and using

¹⁷An 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.

adaptive quadrature or simulation techniques where numerical methods are required.¹⁸ To illustrate (in a slight abuse of notation), we express the expected annual revenue as a function of log distance to the nearest retail store, d , as follows:

$$E[\mathbb{R}_{sc}(d)] = E\left[\sum_{t=1}^4 \sum_{l=1}^{L_{st}(d)} b_{stl}^*(d) \cdot I(c_{stl}^*(d) = c)\right] = \sum_{t=1}^4 E[L_{st}(d)] E[b_{stl}^*(d) \cdot I(c_{stl}^*(d) = c)] \quad (24a)$$

$$= \left[\sum_{t=1}^4 \exp(\mu_{1s} + \mu_t I(t > 1) + \mu_5 d) \Lambda(V_0(0))\right] \left[\int_{\varepsilon_0} b^*(\varepsilon_0|d, s) Pr(c|b^*, d, s) dF(\varepsilon_0)\right] \quad (24b)$$

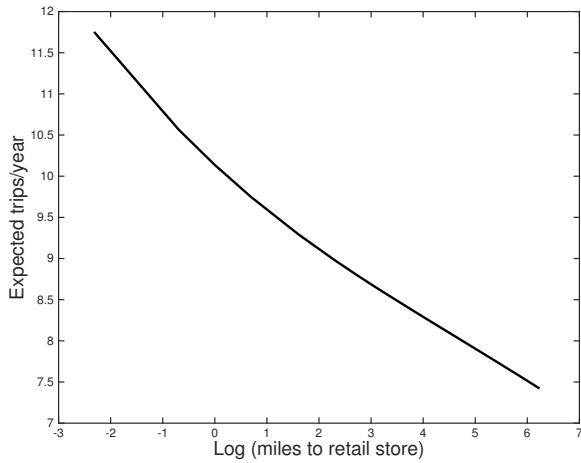
The first expression in equation (24a) recognizes that model-predicted annual revenues are given by summing the optimal budgets of shopping trips in the channel of interest, over every quarter of the year. The second equality in (24a) follows from the stochastic independence of the number of trips per quarter (L) and trip-specific outcomes (b, c). The first bracketed term in (24b), which represents the expected number of trips (purchase occasions) per year, follows from the fact that L is Poisson-distributed (see equation 18) and the functional form assumptions on the rate parameter.¹⁹ The second bracketed term in (24b) is the expected product of the optimal trip budget and the probability of the channel being chosen given the budget.²⁰

Before presenting the results of the expected revenue calculations, we first summarize two components of equation (24b), the expected number of trips per year (first bracketed term) and the conditional probability of channel choice ($Pr(c|b^*, d, s)$, which appears in the second bracketed term). Our intent here is to give some intuition for the relative contributions of these factors to expected revenues. We begin with expected number of trips per year, which we plot by segment in Figure 7. For Segment 1, the expected number of trips per year declines monotonically from about 11.5 at a distance of 0.1 miles (-2 in log scale) to 7.5 at 400 miles (6 in log scale). We refer to segment 1 as “high types” because their purchase frequency is approximately 7.5 times higher than the majority low-type consumers. In Figure 8, we plot the retail choice probabilities for the two segments, conditional upon a budget level and retail store distance – the x axis represents budget amounts and the level curves correspond to retail store distances. As may be seen, retail probabilities are decreasing in both budget amounts and distance, which reflects the patterns in the data shown in Figure 2 and the descriptive regression of Section 2.2.2. Inspection of these plots reveals that distance effects dominate expected product utility differences across the channels, which operate through the budget dependence. Further, comparing retail probabilities across the panels (holding distance and budget fixed) makes clear segment 2’s higher preference for retail.

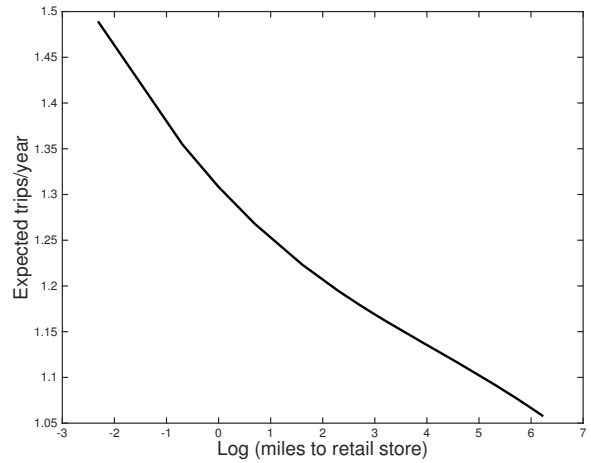
¹⁸In particular, optimal category quantities are computed using simulation as described in Appendix A, integration over the outside good shock distribution is computed via adaptive quadrature, and the remaining computations are done analytically.

¹⁹Referring to the results in Table 4, note that quarterly fixed effects correspond to parameters μ_2 to μ_4 .

²⁰In this term, $b^*(\varepsilon_0)$ is calculated by solving the first order condition (16) using simulated expected product utility values as described in Appendix A; $Pr(c|b^*, d, s)$ is calculated directly from equation (13).

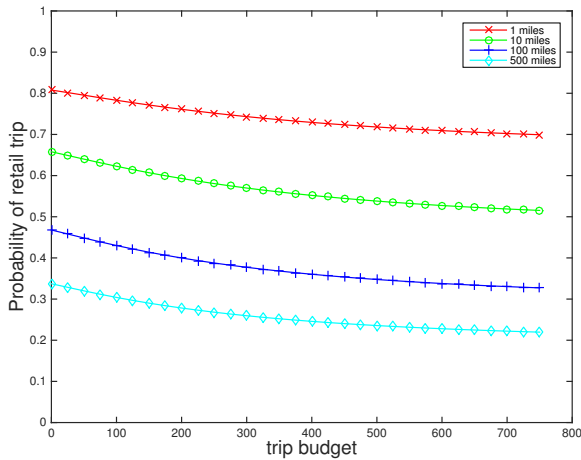


(a) Segment 1

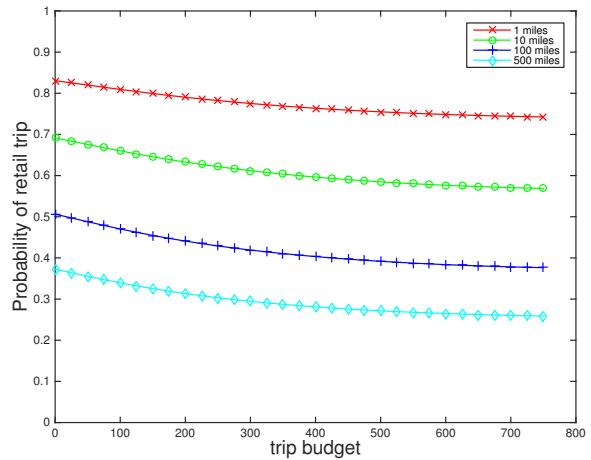


(b) Segment 2

Figure 7: Expected number of trips (purchase occasions) per year conditional on retail store distance, by segment



(a) Segment 1



(b) Segment 2

Figure 8: Probability of retail channel choice, conditional on budget and retail store distance, by segment

We now present the expected annual revenue calculations by channel for the two segments in Figure 9. For each segment, we plot the expected online (dotted red line), retail (dashed blue line) and total (solid black line) per-capita annual revenues as a function of log retail store distance. Several points are worth noting about these plots. First, the “crossover” point at which channels contribute equally to total expected revenues occurs approximately between 25 and 30 miles (3.0 to 3.4 in log scale) from the retail store, depending on the segment. Second, total per-capita expected annual revenues from segment 1 members are approximately ten times that of segment 2 members, at any distance. Coupled with our previous insight that segment 1 consumers purchase about 7.5 times more frequently than segment 2, we infer that roughly 75% of the total revenue differential across the segments is attributable to purchase frequency, while 25% is from larger per-trip budgets for segment

1. Further, consideration of the estimated segment assignment probabilities implies that approximately 40% of total firm revenues come from segment 1.²¹ Third, for both segments, as the distance to the retail store decreases, total expected revenues increase while online revenues decrease – i.e., retail revenues increase faster than online revenues decline. We characterize this relationship as the channels being *net* but not *strict* demand complements, where the former condition requires $\frac{\partial R_{s1}}{\partial d} + \frac{\partial R_{s2}}{\partial d} < 0$ but the latter requires both $\frac{\partial R_{s1}}{\partial d} < 0$ and $\frac{\partial R_{s2}}{\partial d} < 0$.

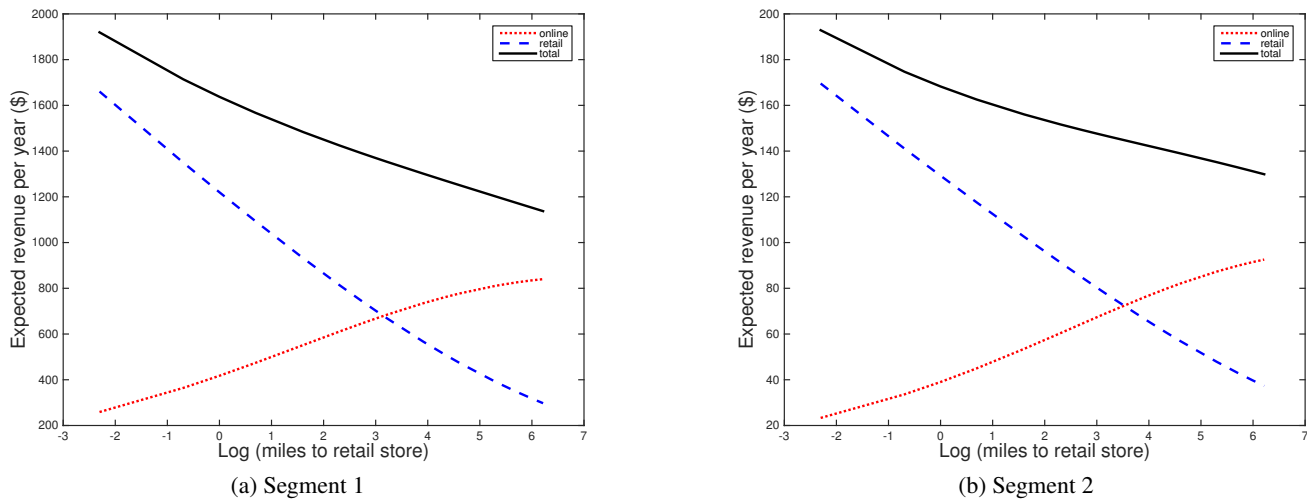


Figure 9: Expected annual revenue by channel as a function of retail store distance, by segment

To derive a summary view of the revenue/distance relationships, in the left hand panel of Figure 10 we generate a segment probability-weighted average of the plots in Figure 9. In the right hand panel of Figure 9, we plot the same information using a log scale for expected revenues (the y axis), which has the attractive feature that the slope at any point on each curve is the demand elasticity at that distance. Using distances falling within the 10th and 90th percentiles of the empirical retail distance distribution for a linear approximation to these curves, the average retail distance elasticity of online revenue is 0.143, while the retail distance elasticity of retail revenue is -0.196. The corresponding elasticity for total revenue is -0.053. Thus, in expectation a 10% reduction in retail distance increases total per-capita revenues 0.53%, by increasing retail revenues 1.96% and decreasing online revenues by 1.43%.

Before concluding this section, we emphasize that the preceding results reflect changes in individual demand among *existing* customers, which we view as the preferred mode of analysis to assess issues of channel complementarity and substitution. However, inasmuch as changing store distance implies store entry (or exit), we have not yet considered all the demand implications of changing retail store distance – namely, the role of customer

²¹To see this, let x be the expected annual revenue from a member of segment 2 (at arbitrary distance) and note that $\frac{.062 \cdot 10 \cdot x}{.062 \cdot 10 \cdot x + .938 \cdot x} \approx 0.4$.

acquisition. In the next application, we build upon our analysis of existing customer demand and expand it to include customer acquisition considerations.

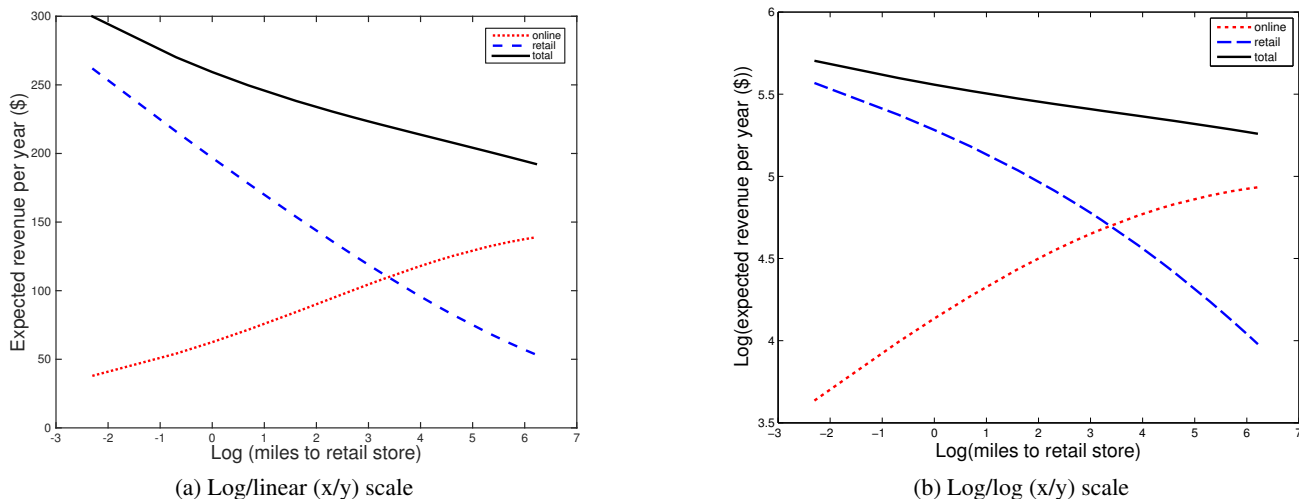


Figure 10: Expected annual revenue by channel as a function of retail store distance, weighted by segment probabilities

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 Γ) 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 $r(d)$ be the expected per-capita total annual revenue for a market located at a log distance d from the entry location of interest. Computation of $r(d)$ is as described in the preceding section and pictured

in Figure 9. Let $\rho(d)$ be the corresponding expected number of new customer arrivals per year for the market; we explain how $\rho(d)$ is calculated from supplementary analysis in Section 5.2.1. Note that the values of r and ρ will be different under the “with entry” and “no entry” conditions for each of the potentially affected markets because the value of d changes across the conditions. Further, let e represent the number of existing customers in the market, δ the firm’s discount factor, and N the planning horizon (in years). Under this setup, the market will generate a discounted revenue stream of $\delta r(e + \rho) + \delta^2 r(e + 2\rho) + \dots + \delta^N r(e + N\rho)$. Thus, the revenue generation index (Γ) is given by:

$$\begin{aligned}\Gamma &= \sum_{n=1}^N \delta^n r(e + \rho n) \\ &= \frac{\delta(1 - \delta^N)}{1 - \delta} r e + \frac{\delta(1 - ((1 - \delta)N + 1)\delta^N)}{(1 - \delta)^2} r \rho\end{aligned}$$

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.

Before presenting the experiment results in Section 5.2.2, we first discuss how we calculate the expected number of new customer arrivals, $\rho(d)$.

5.2.1 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:

$$N_{mt} \sim \text{Poisson}(\rho_{mt}), \log(\rho_{mt}) = \alpha d_{mt} + \delta_m + \mu_t + \log(\text{exposure}_{mt}) \quad (25)$$

In (25), d_{mt} is the log distance from Census block group m 's centroid to the nearest retail store in quarter t . As in Section 2.2, the specification includes bi-level (market, quarter) fixed effects (δ_m, μ_t) to control for unobservables. The variable exposure_{mt} is defined as the market population minus the number of existing customers at that time; including this factor with a coefficient constrained to 1 controls for the number of potential new customers in the market (i.e., the model considers the rate of new customer acquisition among the potential new customer population). The estimate of α from this regression is $\hat{\alpha} = -0.8570$ with a standard error of 0.0636, 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. Since d is in log scale, we may interpret $\hat{\alpha}$ as an elasticity, so that a 10% decrease in retail store distance corresponds to a 8.6% increase in the number of new customers per quarter. The relatively large magnitude of this effect indicates that new customer acquisition is a primary benefit of operating a retail channel.

For the purpose of evaluating our counterfactual, it is useful to ascertain the influence of market demographics on new customer acquisition because estimates of δ_m are not available from the fixed effects Poisson regression above (they are concentrated out of the objective function rather than estimated). To obtain these estimates without weakening the controls identifying the distance effect (α), we perform an auxiliary Poisson regression in which we constrain α to be as estimated above and include select market demographic information obtained from Census 2010 records, i.e.:

$$N_{mt} \sim \text{Poisson}(\rho_{mt}), \log(\rho_{mt}) = \alpha d_{mt} + \beta z_m + \mu_t + \log(\text{exposure}_{mt}) \quad (26)$$

These parameter estimates are reported in Table 6 below – all effects are significant at the 1% level. The results suggest strong positive correlation with market average age, median income, fraction of population white and fraction of population with college degrees. There is a negative correlation associated with rural markets.

regressor	estimate
age	0.0198
income	0.0026
rural	-0.1163
white	0.6954
college	2.9361
state fixed effects	Y
quarter fixed effects	Y

Table 6: Estimates of demographic effects on new customer acquisition

With these estimates in hand, it is a simple matter to compute the expected number of new customers in a market at log retail distance d with a demographic profile summarized by the vector z_m : $\rho(d_m|z_m, t) = \text{exposure}_{mt} * \exp(\hat{\alpha}d + \hat{\beta}z_m + \hat{\mu}_t)$.

5.2.2 Experiment results

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.

Rank	MSA	Revenue index (\$MM)
1	Minneapolis-St. Paul, MN-WI	2.07
2	San Francisco-Oakland-San Jose, CA (C)	1.73
3	St. Louis, MO-IL	1.42
4	Miami-Fort Lauderdale, FL (C)	1.29
5	Columbus, OH	1.20
6	Orlando, FL	1.20
7	Cincinnati-Hamilton, OH-KY-IN (C)	1.02
8	West Palm Beach-Boca Raton, FL	1.02
9	Washington-Baltimore, DC-MD-VA-WV (C)	1.00
10	New York, Northern New Jersey, Long Island, NY-NJ-CT-PA (C)	0.97

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

Comparing the recommended entry MSAs in Table 7 to the brand’s store locations operative during our study reveals that three of these MSAs have existing retail stores. The experiment thus 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 11. 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).

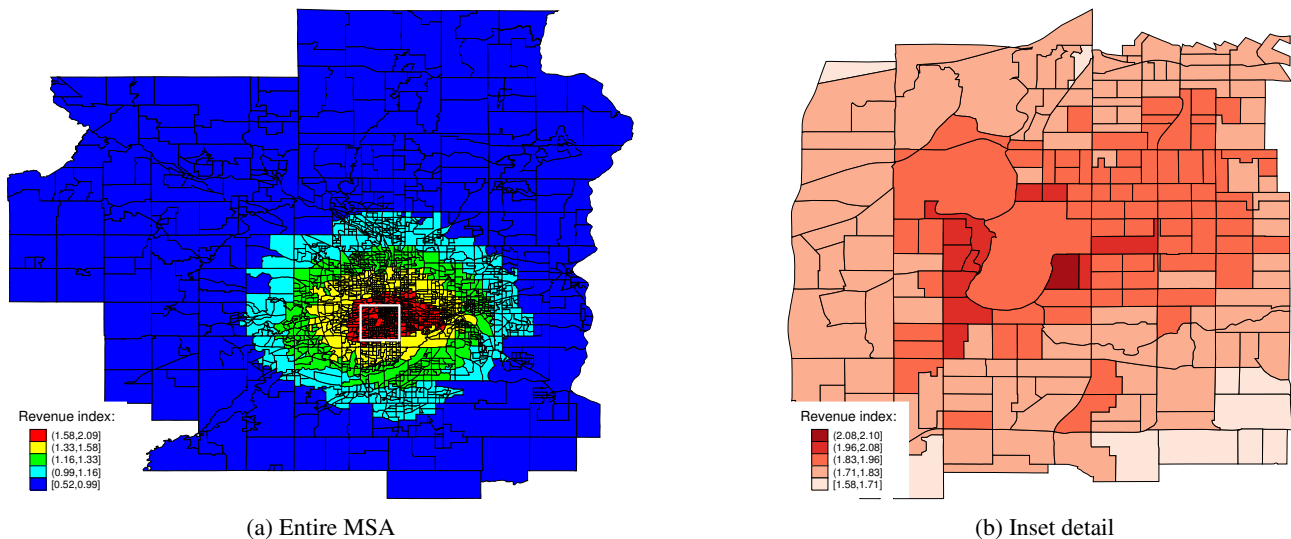


Figure 11: Revenue generation index by Census block group, Minneapolis-St. Paul MSA

5.3 Channel-based pricing

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 twofold: a) summarize the demand response to channel-specific price changes using a series of elasticity measures, and b) demonstrate how the model estimates can be used to solve a simplified price discrimination problem, where the firm holds online category prices fixed and optimizes retail prices by uniformly scaling online category prices.

5.3.1 Demand price elasticities

To illustrate the implications of the different model stages for substitution patterns, we report three different price elasticity measures. The first measure considers substitution patterns across categories conditional upon a budget and channel choice, which thus reflects how consumers with a fixed budget would substitute among categories upon arrival at a store or the website. The second measure allows for substitution across channels and with the outside option, and thus captures the unconditional (channel-specific) category demand response to price changes. The final measure considers the unconditional demand response of total expenditures (revenue) to price changes. 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 our first measure, 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 $\epsilon_{ck} = \frac{\partial E[q_k^* | c, b=\$141]}{\partial p_k} \frac{p_k}{E[q_k^* | c, b=\$141]}$, where the expectation is taken with respect to channel-specific product shocks (ϵ , ν). 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 4, for a given category the pattern of price response across channels tracks the estimates of the corresponding channel/category shock variances (σ_{vck}) – 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.

segment category	homogeneous		2 segment			
	1		1		2	
	online	retail	online	retail	online	retail
1	-2.30	-2.21	-2.18	-2.11	-2.26	-2.15
2	-2.84	-3.44	-2.79	-3.33	-2.83	-3.45
3	-2.51	-2.88	-2.46	-2.98	-2.42	-2.80
4	-2.55	-2.83	-2.47	-2.80	-2.48	-2.75
5	-2.28	-2.53	-2.22	-2.39	-2.22	-2.39
6	-2.44	-5.43	-2.62	-6.04	-2.46	-6.36

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

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 $\epsilon_{ck} = \frac{\partial E[q_{ck}^*]}{\partial p_{ck}} \frac{p_{ck}}{E[q_{ck}^*]}$, where the expectation is taken with respect to the channel-specific product shocks (ε, ν), the channel choice shocks (η), the outside good shock (ε_0), and the probability of L purchase occasions per year. We evaluate these expectations using the methods described in Section 5.1. This elasticity 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 12. For brevity, Figure 12 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 12. 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.

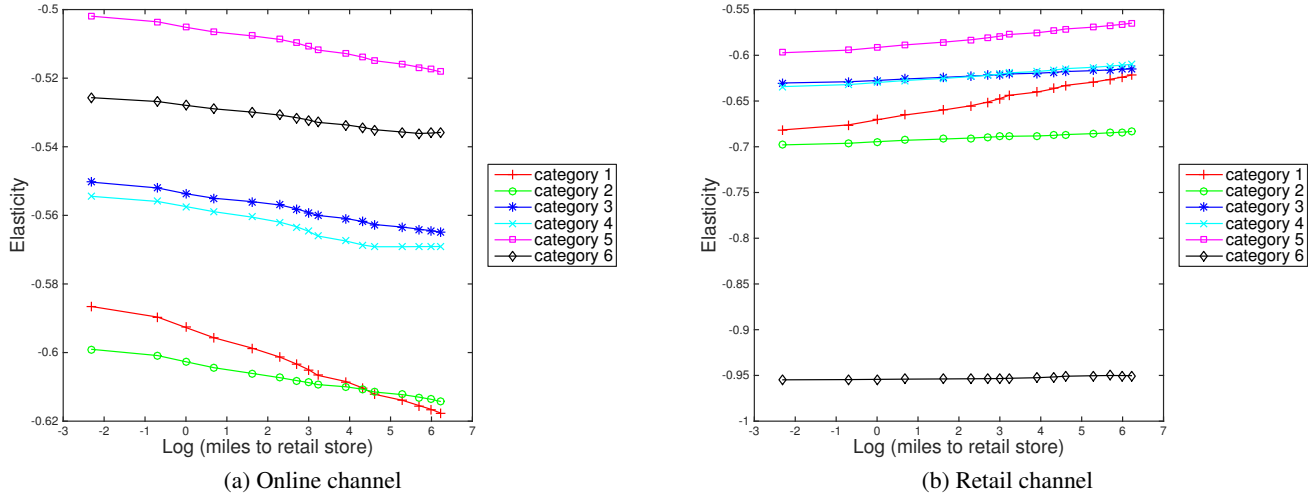


Figure 12: Unconditional quantity index price elasticity measures, homogenous model

Our final set of elasticity measures considers the impact of price changes on total expected annual expenditures (revenues). Again for brevity we use the homogeneous model estimates, and analogous to our definition in Section 5.1, we define total expected expenditures as $E[\mathbb{R}(d)] = E\left[\sum_{l=1}^4 L_l(d) \sum_{i=1}^4 b_{il}^*(d)\right]$. Our measure is thus computed as $\varepsilon_{ck} = \frac{\partial E[\mathbb{R}]}{\partial p_{ck}} \frac{p_{ck}}{E[\mathbb{R}]}$, where the expectation is taken with respect to the same information set as in the previous measure. These elasticities are summarized graphically in Figure 13. An important difference between Figures 12 and 13 is that the latter considers the impact on total basket size rather than the quantity of a single category. Accordingly, the magnitude of these effects is smaller than those in Figure 12, and the magnitude variation across categories reflects the relative preference ordering of the categories.

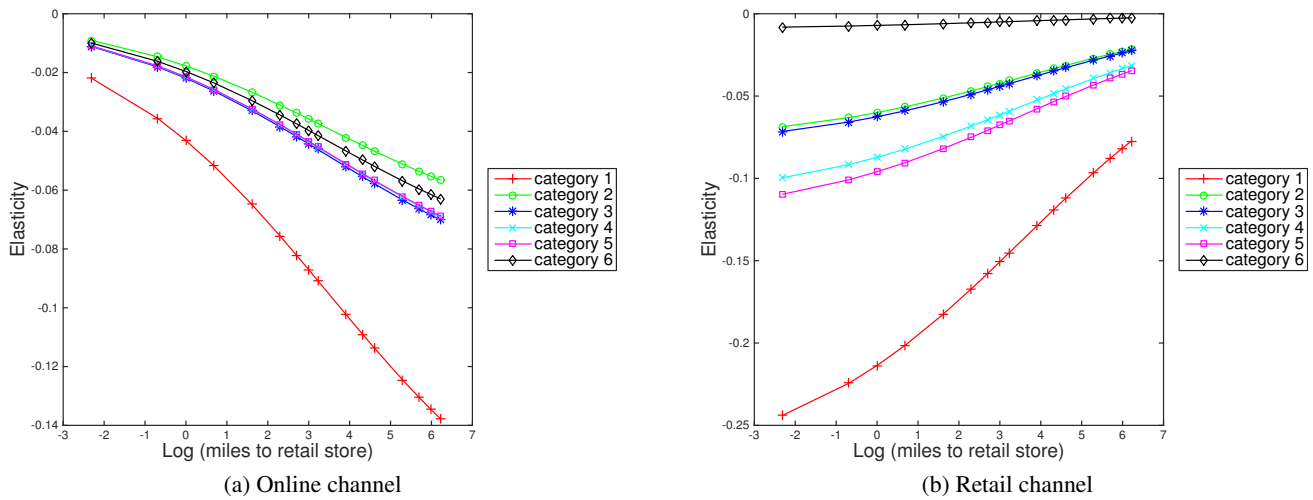


Figure 13: Unconditional total revenue price elasticity measures, homogenous model

5.3.2 Price optimization example

Figures 12 and 13 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 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.²² 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 (that we denote by λ), holding online prices fixed at their values in the final period of observation.

While computing optimal budgets was sufficient for our previous analyses of revenues, determining profits requires that we work directly with purchase quantities because category margins will differ across the channels. As shown in the previous section, optimal quantities will depend upon prices and retail store distance. We therefore write the firm’s objective function (expected profits) as:

$$\pi = E \left[\sum_{t=1}^4 \sum_{l=1}^{L_{it}(d)} \sum_{k=1}^K q_{itlk}^*(d_i, \vec{p}_1, \lambda) ((p_{1k} - \omega_k)I(c_{itl}^*(d_i) = 1) + (\lambda p_{1k} - \omega_k)I(c_{itl}^*(d_i) = 2)) \right] \quad (27)$$

where the firm’s problem is: $\max_{\lambda} \pi$, *subject to* : $\lambda > 0$. In (27), ω_k represents category costs computed as described above and d_i denotes the customer’s distance to the retail store (in log scale). Our approach to handling the distance dependence in (27) is to compute a weighted average of profits, where the weights are generated from the empirical distribution of customer retail distances (in the final period of observation). In other words, the expectation in (27) incorporates an integration over the probability that a customer lives at distance d from the nearest retail store, $Pr(d_i = d)$. As in previous calculations, the expectation is also taken over all econometric shocks ($\varepsilon, \nu, \eta, \varepsilon_0$), as well as the probability of belonging to a particular segment (we use our two segment results to evaluate this counterfactual). 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 14. As may be seen, a lognormal distribution provides a reasonable fit to the empirical distribution of customer

²²An 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.

retail distances.

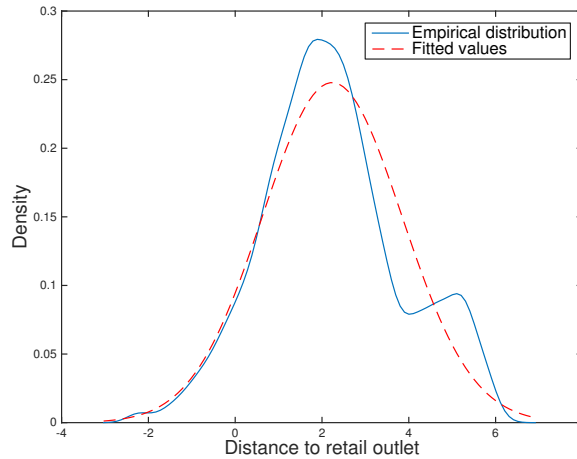


Figure 14: Distribution of customer retail proximity in the final period

We summarize the results of our experiment graphically in Figure 15 below. In addition to searching for the optimal retail/online price ratio for the in-sample distribution of customer retail distances, we explore the sensitivity of the results to alternative assumptions about the retail distance distribution. Specifically, we scale the mean of the distance distribution by factors of 1/2 and 2, and recompute expected profits – the three level curves in Figure 15 correspond to these different scenarios. For each scenario, the optimal price ratio is indicated by a red 'x' on the corresponding curve in the right hand panel of Figure 15. Optimal retail prices range between 67% and 70% of online prices, with higher prices (lower discounts) corresponding to closer average distances. As may be seen, while total expected profits are higher when customers are located closer to retail store, the firm's ability to price discriminate across channels goes down, since retail discounting under optimal pricing is lower at smaller average retail distances. These price ratios are somewhat smaller than the average retail/online price ratio observed in the data (approximately 83%), which, if our cost assumptions are correct, would reflect sub-optimal pricing by the firm.

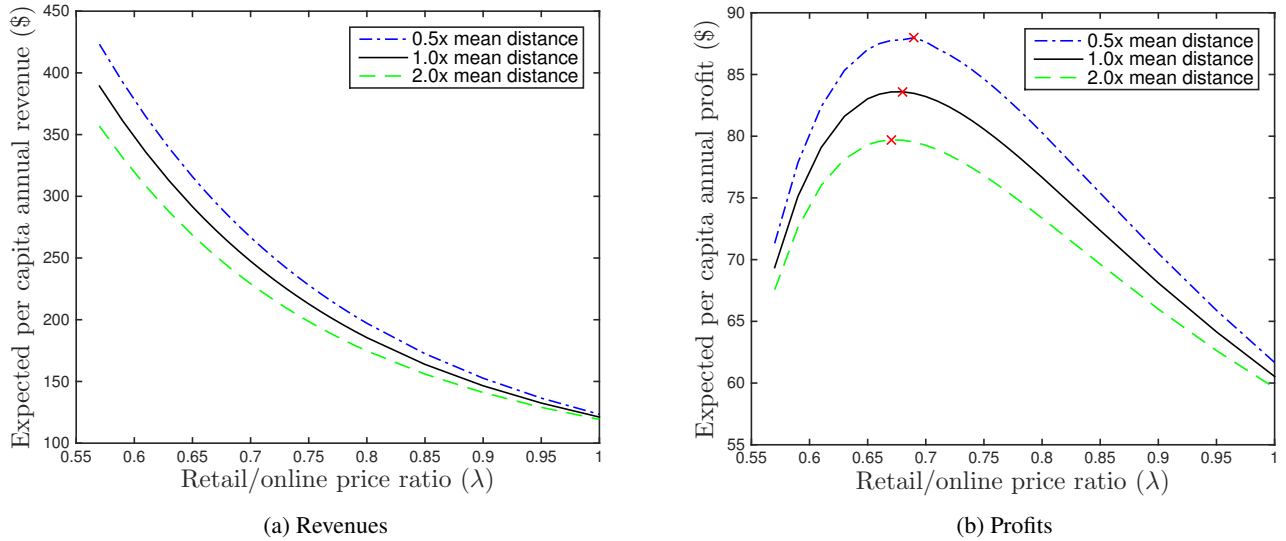


Figure 15: Pricing counterfactual

6 Conclusion

Our intended contributions from this research are threefold. Substantively, the paper adds to the 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 a 10% reduction in retail store distance increases total per-capita revenues 0.53%, by increasing retail revenues 1.96% and decreasing online revenues by 1.43%. 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, this formulation allows us to draw inference on the mechanisms by which channels 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. While our application is to apparel categories, our model and estimation methodology are quite general – they can be employed in any setting where a firm operates both online and retail channels and has access to historical customer transaction data.

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.

There are several potential avenues to extend the current work. One approach might compare our model performance to those that impose alternative assumptions about the consumer's shopping decision sequence. Another possibility for researchers with access to firm inventory data could be to disentangle the dependencies of channel choice on perceived stockout risk and the provision of product information. Those with panel data containing high purchase frequencies or long time dimensions could tackle dynamic considerations such as state dependence in channel choices or learning about product categories over time. Relatedly, the role of channel interactions in the consumer's search process could be explored. While we believe that our focus on apparel categories to some extent limits the scope for dynamics in our analysis (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 direction for future work.

References

- Ansari, A., C. Mela, and S. Neslin (2008). Customer channel migration. *Journal of Marketing Research* 45(1), 60–76.
- Avery, J., T. J. Steenburgh, J. Deighton, and M. Caravella (2012). Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time. *Journal of Marketing* 76(3), 96–111.
- Bell, D. R., S. Gallino, and A. Moreno (2013). Inventory showrooms and customer migration in omni-channel retail: The effect of product information. Technical report, University of Pennsylvania.
- Bell, D. R., T.-H. Ho, and C. S. Tang (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research* 35(3), 352–369.
- Bell, D. R. and J. M. Lattin (1998). Shopping behavior and consumer preference for store price format: Why large basket shoppers prefer EDLP. *Marketing Science* 17(1), 66–88.
- Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological* 42(3), 274–303.
- Biyalogorsky, E. and P. Naik (2003). Clicks and mortar: The effect of on-line activities on off-line sales. *Marketing Letters* 14(1), 21–32.
- Chevalier, J. A., A. K. Kashyap, and P. E. Rossi (2003). Why don't prices rise during periods of peak demand? Evidence from scanner data. *American Economic Review* 93(1), 15–37.
- Deleersnyder, B., I. Geyskens, K. Gielens, and M. Dekimpe (2002). How cannibalistic is the internet channel? A study of the newspaper industry in the United Kingdom and the Netherlands. *International Journal of Research in Marketing* 19(4), 337–348.

- Dzyabura, D., S. Jagabathula, and E. Muller (2015). Using online preference measurement to infer offline purchase behavior. Technical report, New York University.
- Ehrenberg, A. S. (1959). The pattern of consumer purchases. *Applied Statistics* 8(1), 26–41.
- Forman, C., A. Ghose, and A. Goldfarb (2009). Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Science* 55(1), 47–57.
- Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review* 97(3), 713–744.
- Geyskens, I., K. Gielens, and M. Dekimpe (2002). The market valuation of internet channel additions. *Journal of Marketing* 66(2), 102–119.
- Gordon, B. R., A. Goldfarb, and Y. Li (2013). Does price elasticity vary with economic growth? A cross-category analysis. *Journal of Marketing Research* 50(1), 4–23.
- Hanemann, W. M. (1984). Discrete/continuous models of consumer demand. *Econometrica* 54(3), 541–561.
- Kim, J., G. M. Allenby, and P. E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science* 21(3), 229–250.
- Morrison, D. G. and D. C. Schmittlein (1988). Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *Journal of Business & Economic Statistics* 6(2), 145–159.
- Neslin, S. A., D. Grewal, R. Leghorn, V. Shankar, M. L. Teerling, J. S. Thomas, and P. C. Verhoef (2006). Challenges and opportunities in multichannel customer management. *Journal of Service Research* 9(2), 95–112.
- Pauwels, K. and S. Neslin (2011). Building with bricks and mortar: The revenue impact of opening physical stores in a multichannel environment. Technical report, Marketing Science Institute.
- Pinjari, A. R. and C. Bhat (2010). An efficient forecasting procedure for Kuhn-Tucker consumer demand model systems. Technical report, University of Texas.
- Soysal, G. and A. Zentner (2014). Measuring e-commerce concentration effects in the apparel industry. Technical report, University of Texas at Dallas.
- Swait, J. and J. Louviere (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research* 30(3), 305–314.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Van Nierop, J., P. Leeflang, M. Teerling, and K. Huizingh (2011). The impact of the introduction and use of an informational website on offline customer buying behavior. *International Journal of Research in Marketing* 28(2), 155–165.
- Wales, T. J. and A. D. Woodland (1983). Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics* 21(3), 263–285.
- Wang, K. and A. Goldfarb (2014). Can offline stores drive online sales? Technical report, University of Toronto.
- Yoganarasimhan, H. (2013). The value of reputation in an online freelance marketplace. *Marketing Science* 32(6), 860–891.
- Yoganarasimhan, H. (2015). Estimation of beauty contest auctions. *Marketing Science, Forthcoming*.

Appendices

A 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 (ε, ν) , without loss of generality assume the first M categories are chosen. For chosen categories, the Kuhn-Tucker condition

$\frac{\partial \mathcal{L}}{\partial e_k} = 0$ implies $\frac{\Psi_{ck}}{p_{ck} \left(\frac{e_k^*}{p_{ck} \gamma_k} + 1 \right)} = \lambda$, which may be solved for the optimal expenditure: $e_k^* = \gamma_k p_{ck} \left(\frac{\Psi_{ck}}{\lambda p_{ck}} - 1 \right)$. Next,

we use the budget constraint equation to eliminate the Lagrange multiplier λ : $b = \sum_{k=1}^M e_k^* = \sum_{k=1}^M \gamma_k p_{ck} \left(\frac{\Psi_{ck}}{\lambda p_{ck}} - 1 \right)$.

Solving for λ gives $\lambda = \frac{\sum_{k=1}^M \gamma_k \Psi_{ck}}{b + \sum_{k=1}^M \gamma_k p_{ck}}$. Then, substituting this expression for λ back into the optimal expenditure

equation above gives: $e_k^* = \gamma_k p_{ck} \left(\frac{\Psi_{ck} \left(b + \sum_{k=1}^M \gamma_k p_{ck} \right)}{p_{ck} \left(\sum_{k=1}^M \gamma_k \Psi_{ck} \right)} - 1 \right)$. Equivalently, the optimal quantities are:

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

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 $\left(\frac{\Psi_{ck}}{p_{ck}} \right)$ 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 the product shocks (ε, ν)
2. Compute the price normalized baseline utilities for all K categories: $\Upsilon_k = \frac{\Psi_{ck}}{p_{ck}}$
3. Sort the categories from highest to lowest Υ ; denote quantities sorted in this order with a tilde, e.g. $\tilde{\Upsilon}$
4. Iteratively compute the Lagrange multiplier, assuming the first m categories are chosen: $\lambda^m = \frac{\sum_{j=1}^m \tilde{\gamma}_j \tilde{\Psi}_{cj}}{b + \sum_{j=1}^m \tilde{\gamma}_j \tilde{p}_{cj}}$
5. Determine the number of chosen categories by the relation: $M = \sum_{m=1}^K I(\tilde{\Upsilon}_m \geq \lambda^m)$
6. Compute the optimal quantities for the chosen categories as $\tilde{q}_m^* = \tilde{\gamma}_m \left(\frac{\tilde{\Psi}_{cm} \left(b + \sum_{k=1}^M \tilde{\gamma}_m \tilde{p}_{cm} \right)}{\tilde{p}_{cm} \left(\sum_{m=1}^M \tilde{\gamma}_m \tilde{\Psi}_{cm} \right)} - 1 \right)$ 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 (6a).

Online Appendices - Not Intended for Publication

B Details of the product model

B.1 Kuhn-Tucker conditions

As stated in the text, the Kuhn-Tucker conditions are: a) $\frac{\partial \mathcal{L}}{\partial e_k} \leq 0$ (stationarity), b) $e_k \geq 0$ (primal feasibility), and c) $\frac{\partial \mathcal{L}}{\partial e_k} e_k = 0$ (complementary slackness). Formulating the Lagrangian in terms of expenditures yields: $\mathcal{L} = \sum_{k=1}^K \Psi_{ck} \gamma_k \log\left(\frac{e_k}{p_{ck}\gamma_k} + 1\right) - \lambda \left(\sum_{k=1}^K e_k - b\right)$. 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 $e_1^* > 0$. By condition (c) $\frac{\partial \mathcal{L}}{\partial e_1} = 0$, which implies the Lagrange multiplier is given by $\lambda = \frac{\Psi_{c1}}{p_{c1}\left(\frac{e_1^*}{p_{c1}\gamma_1} + 1\right)}$. For other chosen categories (where $e_k^* > 0$), we must have $\frac{\partial \mathcal{L}}{\partial e_k} = \frac{\Psi_{ck}}{p_{ck}\left(\frac{e_k^*}{p_{ck}\gamma_k} + 1\right)} - \lambda = \frac{\Psi_{ck}}{p_{ck}\left(\frac{e_k^*}{p_{ck}\gamma_k} + 1\right)} - \frac{\Psi_{c1}}{p_{c1}\left(\frac{e_1^*}{p_{c1}\gamma_1} + 1\right)} = 0$. Rearranging, taking logarithms, and substituting in from equation (6b) yields:

$$\begin{aligned} \log(\Psi_{ck}) - \log\left(\frac{e_k^*}{p_{ck}\gamma_k} + 1\right) - \log(p_{ck}) &= \log(\Psi_{c1}) - \log\left(\frac{e_1^*}{p_{c1}\gamma_1} + 1\right) - \log(p_{c1}) \\ \beta_{sk} + v_{ck} - \log\left(\frac{e_k^*}{p_{ck}\gamma_k} + 1\right) - \log(p_{ck}) + \varepsilon_k &= \beta_{s1} + v_{c1} - \log\left(\frac{e_1^*}{p_{c1}\gamma_1} + 1\right) - \log(p_{c1}) + \varepsilon_1 \\ V_{ck} + \varepsilon_k &= V_{c1} + \varepsilon_1 \end{aligned}$$

Where $V_{ck} \equiv \beta_{sk} + v_{ck} - \log\left(\frac{e_k^*}{p_{ck}\gamma_k} + 1\right) - \log(p_{ck})$. For non-chosen categories, a similar derivation yields $V_{ck} + \varepsilon_k < V_{c1} + \varepsilon_1$. These conditions correspond to those stated in equation (9) in the text.

B.2 Likelihood function

This derivation follows Bhat (2008). We begin by evaluating the integrals with respect to the ε_k terms in equation (10) :

$$\begin{aligned} Pr(\vec{e}^* | c, b) &= |J_{\vec{\varepsilon}_M \rightarrow \vec{e}_M}| \int_{\mathbf{v}} \int_{\varepsilon_1 = -\infty}^{+\infty} \int_{\varepsilon_{M+1} = -\infty}^{V_{c1} - V_{c,M+1} + \varepsilon_1} \cdots \int_{\varepsilon_K = -\infty}^{V_{c1} - V_{cK} + \varepsilon_1} f(\vec{\varepsilon}) f(\vec{\mathbf{v}}) d\vec{\varepsilon}_M d\vec{\mathbf{v}} \\ &= |J_{\vec{\varepsilon}_M \rightarrow \vec{e}_M}| \int_{\varepsilon_1 = -\infty}^{+\infty} \int_{\mathbf{v}} \left\{ \prod_{j=2}^M \frac{1}{\sigma_\varepsilon} \lambda \left[\frac{V_{c1} - V_{cj} + \varepsilon_1}{\sigma_\varepsilon} \right] \right\} \\ &\quad \times \left\{ \prod_{j=M+1}^K \Lambda \left[\frac{V_{c1} - V_{cj} + \varepsilon_1}{\sigma_\varepsilon} \right] \right\} \frac{1}{\sigma_\varepsilon} \lambda \left[\frac{\varepsilon_1}{\sigma_\varepsilon} \right] d\varepsilon_1 f(\vec{\mathbf{v}}) d\vec{\mathbf{v}} \end{aligned}$$

where $\lambda(x) = e^{-(x+e^{-x})}$ and $\Lambda(x) = e^{-e^{-x}}$ are the standard extreme value pdf and cdf, respectively. Substi-

tuting and rearranging yields:

$$\begin{aligned} Pr(\vec{e}^*|c, b) &= |J_{\vec{e}_M \rightarrow \vec{e}_M}| \frac{1}{\sigma_\varepsilon^{M-1}} \int_{\mathbf{v}} \prod_{j=2}^M \exp\left(\frac{V_{c1} - V_{cj}}{\sigma_\varepsilon}\right) \\ &\quad \times \int_{\varepsilon_1 = -\infty}^{+\infty} e^{-\sum_{j=1}^K \exp\left(-\left(\frac{V_{c1} - V_{cj} + \varepsilon_1}{\sigma_\varepsilon}\right)\right)} \left(\frac{1}{\sigma_\varepsilon} e^{-\frac{\varepsilon_1}{\sigma_\varepsilon}}\right)^{M-1} \frac{1}{\sigma_\varepsilon} e^{-\frac{\varepsilon_1}{\sigma_\varepsilon}} d\varepsilon_1 f(\vec{v}) d\vec{v} \end{aligned}$$

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

$$\begin{aligned} \int_{\varepsilon_1 = -\infty}^{+\infty} e^{-\sum_{j=1}^K \exp\left(-\left(\frac{V_{c1} - V_{cj} + \varepsilon_1}{\sigma_\varepsilon}\right)\right)} \left(\frac{1}{\sigma_\varepsilon} e^{-\frac{\varepsilon_1}{\sigma_\varepsilon}}\right)^{M-1} \frac{1}{\sigma_\varepsilon} e^{-\frac{\varepsilon_1}{\sigma_\varepsilon}} d\varepsilon_1 &= \int_{u=0}^{+\infty} e^{-u \sum_{j=1}^K \exp\left(-\left(\frac{V_{c1} - V_{cj}}{\sigma_\varepsilon}\right)\right)} u^{M-1} du \\ &= (M-1)! \left(\sum_{j=1}^K \exp\left(-\left(\frac{V_{c1} - V_{cj}}{\sigma_\varepsilon}\right)\right) \right)^{-M} \end{aligned}$$

Substituting back in and simplifying yields:

$$\begin{aligned} Pr(\vec{e}^*|c, b) &= |J_{\vec{e}_M \rightarrow \vec{e}_M}| \frac{(M-1)!}{\sigma_\varepsilon^{M-1}} \int_{\mathbf{v}} \prod_{j=2}^M \exp\left(\frac{V_{c1} - V_{cj}}{\sigma_\varepsilon}\right) \left(\sum_{j=1}^K \exp\left(-\left(\frac{V_{c1} - V_{cj}}{\sigma_\varepsilon}\right)\right) \right)^{-M} f(\vec{v}) d\vec{v} \quad (29) \\ &= |J_{\vec{e}_M \rightarrow \vec{e}_M}| \frac{(M-1)!}{\sigma_\varepsilon^{M-1}} \int_{\mathbf{v}} \left(\prod_{j=1}^M e^{V_{cj}/\sigma_\varepsilon} \right) \left(\sum_{j=1}^K e^{V_{cj}/\sigma_\varepsilon} \right)^{-M} f(\vec{v}) d\vec{v} \end{aligned}$$

Next, consider the elements of the Jacobian:

$$\begin{aligned} J_{ij} &= \frac{\partial \varepsilon_{i+1}}{\partial e_{j+1}^*} = \frac{\partial}{\partial e_{j+1}^*} (V_{c1} - V_{c,i+1} + \varepsilon_1) \quad \forall i, j = 1, \dots, M-1 \\ &= \frac{\partial}{\partial e_{j+1}^*} \left(\beta_{s1} + v_{c1} - \log\left(\frac{e_1^*}{p_1 \gamma_1} + 1\right) - \log(p_{c1}) - \left(\beta_{s,i+1} + v_{c,i+1} - \log\left(\frac{e_{i+1}^*}{p_{c,i+1} \gamma_{i+1}} + 1\right) - \log(p_{c,i+1}) \right) + \varepsilon_1 \right) \\ &= \frac{\partial}{\partial e_{j+1}^*} \left(-\log\left(\frac{b - \sum_{k=2}^K e_k^*}{p_{c1} \gamma_1} + 1\right) + \log\left(\frac{e_{i+1}^*}{p_{c,i+1} \gamma_{i+1}} + 1\right) \right) \\ &= \frac{1}{\frac{b - \sum_{k=2}^K e_k^*}{p_{c1} \gamma_1} + 1} \frac{1}{p_{c1} \gamma_1} + \frac{1}{\frac{e_{i+1}^*}{p_{c,i+1} \gamma_{i+1}} + 1} \frac{1}{p_{c,i+1} \gamma_{i+1}} I(i=j) \\ &= \frac{1}{e_1^* + p_{c1} \gamma_1} + \frac{1}{e_{i+1}^* + p_{c,i+1} \gamma_{i+1}} I(i=j) \end{aligned}$$

The Jacobian determinant is then:

$$|J_{\mathbf{e}_M \rightarrow \mathbf{e}_M}| = \begin{vmatrix} \frac{1}{e_1^* + p_{c1} \gamma_1} + \frac{1}{e_2^* + p_{c2} \gamma_2} & \frac{1}{e_1^* + p_{c1} \gamma_1} & \cdots & \frac{1}{e_1^* + p_{c1} \gamma_1} \\ \frac{1}{e_1^* + p_{c1} \gamma_1} & \frac{1}{e_1^* + p_{c1} \gamma_1} + \frac{1}{e_3^* + p_{c3} \gamma_3} & \cdots & \frac{1}{e_1^* + p_{c1} \gamma_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{e_1^* + p_{c1} \gamma_1} & \frac{1}{e_1^* + p_{c1} \gamma_1} & \cdots & \frac{1}{e_1^* + p_{c1} \gamma_1} + \frac{1}{e_M^* + p_{cM} \gamma_M} \end{vmatrix} = \left(\prod_{j=1}^M \frac{1}{e_j^* + p_{cj} \gamma_j} \right) \left(\sum_{j=1}^M e_j^* + p_{cj} \gamma_j \right)$$

Substituting this expression for the Jacobian into (29) above gives the fully simplified likelihood :

$$Pr(\vec{e}^*|c, b) = \frac{(M-1)!}{\sigma_\varepsilon^{M-1}} \left(\prod_{j=1}^M \frac{1}{e_j^* + p_{cj}\gamma_j} \right) \left(\sum_{j=1}^M e_j^* + p_{cj}\gamma_j \right) \int_{\mathbf{v}} \left(\prod_{j=1}^M e^{v_{cj}/\sigma_\varepsilon} \right) \left(\sum_{j=1}^K e^{v_{cj}/\sigma_\varepsilon} \right)^{-M} f(\vec{v}) d\vec{v}$$

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 $\Psi_0 = \exp(\beta_{0s} + \varepsilon_0)$ and $g(b) = \sigma_\eta \log \left(\exp \left(\frac{\bar{U}_1(b)}{\sigma_\eta} \right) + \exp \left(\frac{\bar{U}_2(b)}{\sigma_\eta} \right) \right)$. 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:

$$\frac{\partial g}{\partial b} - \Psi_0 \begin{cases} = 0 & \text{if } b^* > 0 \\ < 0 & \text{if } b^* = 0 \end{cases}$$

Rearranging, substituting for Ψ_0 and taking logs gives equation (16) in the text.

C.2 Likelihood function

From Section 3.3, the objective is to compute $Pr(b^* | b^* > 0) = \frac{Pr(b^* \cap b^* > 0)}{Pr(b^* > 0)} = \frac{Pr(b^*)}{Pr(b^* > 0)}$. Given the one to one mapping of the outside good shock (ε_0 , distributed $EV(0, \sigma_{\varepsilon_0})$) to the budget b^* , the numerator of the preceding equation is simply $|J_{\varepsilon_0 \rightarrow e}| \lambda \left(\frac{V_0(b^*)}{\sigma_{\varepsilon_0}} \right)$, where $\lambda(\cdot)$ is the extreme value pdf. Similarly, from the KT conditions (16), $Pr(b^* > 0) = 1 - Pr(b^* = 0) = 1 - Pr(\varepsilon_0 > V_0(0)) = Pr(\varepsilon_0 \leq V_0(0)) = \Lambda \left(\frac{V_0(0)}{\sigma_{\varepsilon_0}} \right)$, where $\Lambda(\cdot)$ is the extreme value cdf. Finally, for positive budgets, the Jacobian term is computed from (16) as $J = \frac{\partial \varepsilon_0}{\partial b} = \frac{\partial}{\partial b} \left(\log \left(\frac{\partial g}{\partial b} \right) - \beta_{s0} \right) = \left(\frac{\partial g}{\partial b} \right)^{-1} \left(\frac{\partial^2 g}{\partial b^2} \right)$, where the derivatives of g are evaluated at b^* .

Putting these sub-calculations together gives equation (17) 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 ε and \mathbf{v} , 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 (6a) to express the utility contribution in terms of the draw values (via the Ψ 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:

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

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

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

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:

$$Eu_c = E[u(\vec{q}^*|b, c)] = \frac{1}{D} \sum_{d=1}^D \sum_{k=1}^{M_d} u_{cdk}$$

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

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

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:

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

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

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

$$- \frac{2}{D^2} \left(\sum_{d=1}^D \sum_{k=1}^{M_d} u_{cdk} \right) \left(\sum_{d=1}^D \sum_{k=1}^{M_d} u''_{cdk} \right)$$

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:

$$\begin{aligned}\frac{\partial g}{\partial b} &= Pr(c = 1 | b) (Eu'_1 + \phi_3 Vu'_1) + Pr(c = 2 | b) (Eu'_2 + \phi_3 Vu'_2) \\ \frac{\partial^2 g}{\partial b^2} &= \frac{1}{\sigma_\eta} Pr(c = 1 | b) Pr(c = 2 | b) (Eu'_1 - Eu'_2 + \phi_3 Vu'_1 - \phi_3 Vu'_2)^2 \\ &\quad + Pr(c = 1 | b) (Eu''_1 + \phi_3 Vu''_1) + Pr(c = 2 | b) (Eu''_2 + \phi_3 Vu''_2)\end{aligned}$$