Pennies for Your Thoughts: Costly Product Consideration and Purchase Quantity Thresholds*

Yufeng Huang† and Bart J. Bronnenberg‡

March 15, 2018

Abstract

Individual demand for consumer packaged goods shows discrete jumps between zero and large quantities, under a marginal change in price. Ruling out multiple alternative explanations, this paper provides evidence from micro-data in the yogurt category that these jumps are caused by consumer fixed purchasing costs per product. We formulate and estimate a model in which (1) such fixed costs limit the number of different products considered, and (2) consumers use prices to screen a product in and out of their consideration set. Our structural estimation finds that the consumer incurs fixed costs of \$0.81 to consider a product. These costs are increased by 280% if she has not purchased the product for a year, and are decreased by 59% when the product is featured in the store; the dependence of fixed costs on information shifters suggests that these costs are incurred because of consideration. Consideration being scarce at the shelf, firms compete fiercely for customers: we simulate counterfactual markups in a world full of feature advertising and find that firms enjoy higher equilibrium markups because the provision of information softens competition for consideration.

Keywords: Love for variety, promotion threshold, consumer fixed costs, limited consideration

*We thank Jaap Abbring, Marnik Dekimpe, Jean-Pierre Dubé, Els Gijsbrechts, Alessandro Iaria, Jun Kim, Keyvan Dehmamy, Tobias Klein, Mitch Lovett, Stephan Seiler, Thomas Otter and Kosuke Uetake for comments and suggestions. We are also grateful to seminar/conference participants at the 2015 CEPR/IO Summer School, 2016 Marketing Science Conference, 2016 Stanford Digital Marketing Conference, Jinan University, Sun Yat-sen University, and the University of Rochester. The authors thank the Netherlands Organization for Scientific Research (NWO Vici Grant 453-09-004) for financial support.

†University of Rochester Simon Business School, Rochester, NY 14627, USA. E-mail: yufeng.huang@simon.rochester.edu.

‡Graduate School of Business, Stanford University, Stanford, California 94305, and CEPR, London EC1V 0DX. E-mail: bart.bronnenberg@stanford.edu.

1 Introduction

A widely-held belief among practitioners is that small price discounts do not persuade consumers to buy, i.e., that consumers have a "promotion threshold". For example, Della Bitta and Monroe (1981) document that many retailers believe that there is a minimum discount threshold of 15% before one can attract consumers to a sale. In the left-hand-side panel of Figure 1, we plot the demand for a yogurt product using consumer-level purchase data. From this graph, consumers show little to no response to price changes when prices are high, but are very elastic to price changes when prices are low.[1] Further, the possibility that elasticity can be increasing in discount depth cannot be rationalized by commonly used demand functions such as the log-log model or the multinomial logit model. This paper (1) documents that individuals are more elastic at intermediate prices than at high prices, (2) proposes an explanation for such a pattern consistent with the theory of consideration sets, (3) empirically tests the explanation against alternatives, and (4) structurally quantifies the impact of costly consideration sets on consumer behavior and market prices.

Our explanation for this convex aggregate demand curve becomes clear once we decompose it into individual demand curves. To exemplify, in the right-hand-side panel of Figure 1, we plot demand curves for the same product separately for two groups of consumers having different "price thresholds" – in this case, observed maximum prices for buying at the 25th and 75th percentiles of the overall price distribution. We find (1) that consumer demand features discontinuities at the individual price threshold, (2) that around this threshold the consumer either does not buy at all or buys in large quantities, and (3) that the transition between these two decisions is driven by marginal shocks to price.[2] The convex demand function is thus a linear combination of discontinuous individual demand functions, smoothed out from integration over heterogeneous thresholds.

After ruling out alternative explanations, e.g., consumer responses to volume deals, stockpiling, or across-consumer heterogeneity, these jumps in individual demand curves at price thresholds suggest that there exists consumer fixed purchasing costs per product, which are incurred after prices are known. We interpret these fixed costs as the mental costs to collect and process the information needed to make a purchase decision, i.e. costs of consideration.[3] Take yogurt as an

[1] This pattern can also explain why price discounts are infrequent but large (Blattberg et al., 1995). An analogy on advertising response threshold is discussed in detail by Dubé et al. (2005).

[2] We find similar patterns across the top 4 products as shown in Appendix Figure 11.

[3] In this paper, we view this terminology as interchangeable with costly evaluation or costly thinking.

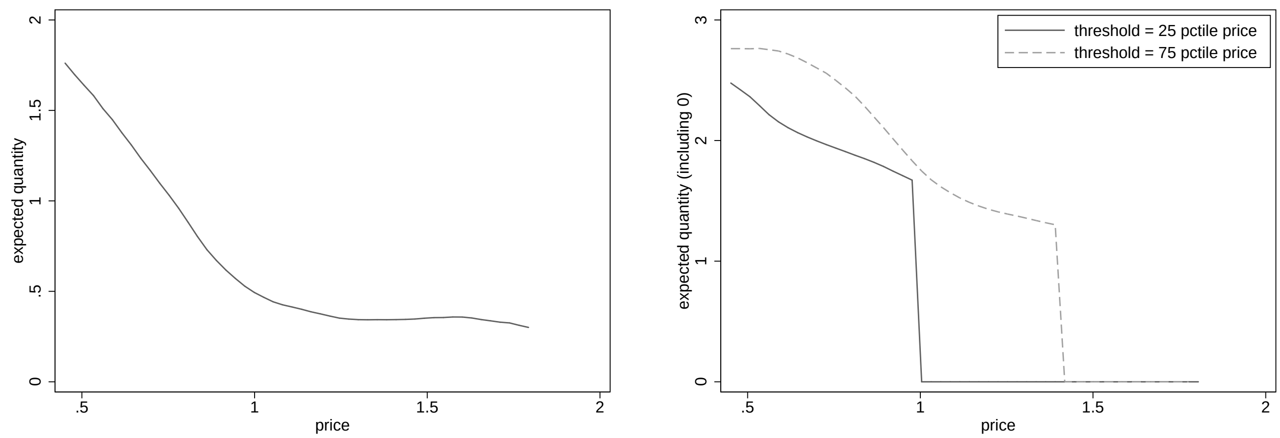

Figure 1: Overall- and by-cohort- demand for Dannon Light

Notes: The left panel pictures quantity demand for Dannon Light (including zeros) as a function of price, pooled over all individuals. The right panel portrays demand by "cohort" defined as consumers whose observed price thresholds (max accepted price) are similar. In particular, the right panel plots demand for two cohorts whose price thresholds are within $0.1 from the 25th and 75th percentile of price distribution.

example. A consumer with some interest to shop for yogurt not only needs to know the price, but also needs to know, understand, and weigh information about flavor, brand, nutritional content, available fridge space at home, and other information, before she is able to determine whether and which product to purchase and which quantity of it. If significant effort is required to collect and evaluate these non-price attributes of each product, the consumer will likely restrict attention to a smaller consideration set, and – more importantly – use prices as cues to screen products in and out of the set. Therefore, marginal price changes are able to push a consumer in choosing between zero and large quantities, generating individual demand curves with jumps at the price threshold.

We start the paper by empirically describing individual consumer demand functions, focusing on the yogurt category of the IRI Academic Dataset. We show that individual consumers purchase large quantities concentrated in small subsets of available products, but their purchase sets change trip-by-trip depending on how products are priced. These patterns suggest that, if a marginal price change is able to induce the consumer to start buying a product (referred to as a "price threshold"), she buys the product in large quantities, resulting in a discontinuity in her demand function at this price. We formally document that these quantity discontinuities are large and the magnitude cannot be explained by quantity discounts or product indivisibility. Thus, assuming decreasing marginal utility of consumption, these quantity jumps suggest the existence of a consumer fixed cost to acquire each product. We also test against alternative explanations, in particular (1) consumer and product heterogeneity and (2) forward buying and storing goods.[4]

Next, through the lens of a structural model, we conjecture that an important part of this fixed cost as associated with consumer's consideration of each product. In our model, consumers have full information on price, but decide to spend effort to discover match values to products before making the decision of whether and how much to buy. We first construct an illustrative model and show numerically that it implies demand curves that are consistent with our data. Then, we discuss the extent to which this consideration set model can be identified by purchase quantity data. It is difficult to distinguish between lack of consideration and low preferences. Our identification strategy builds on the idea that the mental effort of consideration (in the sense of incurring fixed costs per product) creates scale economies and encourages consumers to either buy large quantities or not buy at all. With variation on price, this mechanism creates jumps in quantity that are not rationalized by a standard model. We provide evidence from Monte Carlo experiments showing that consideration costs in our setup are identified from the observed discontinuities in quantity purchases, without having to resort to exclusion restrictions in promotion and advertising, and even when the functional form of utility is not parametrically known.

We then estimate the model to measure the magnitude of consumer fixed costs which we interpret as the cost of consideration. Our estimates suggest that, for consumers who have just purchased the product the week before, fixed costs are \$0.81 on average. We further infer from our model that these fixed costs increase with inter-purchase time and decrease with feature advertising: they are 280% higher for consumers who have not bought the product for a year, and 59% lower for products on feature. The empirical dependence of the estimated fixed costs on information shifters supports our conjecture that a large fraction of these costs can be attributed to costly consideration, i.e. consumers' effort on information collection or processing. However, the usual caveat applies that the attribution of fixed purchasing costs to consideration is to some extent an interpretation, made through the lens of a structural model.

Finally, we evaluate the role of price discounts in our model. Because consideration is costly, discount strategies help overcome a consumer's consideration barrier. We thus use our model to decompose price elasticities into an effect on inclusion of the consideration set, and another effect

[4]We condition on category purchase and this likely rules out price search as well as travel costs within the store.

on quantity choice given consideration-set membership. We find that roughly a third of the overall price elasticities stem from additional consideration rather than increasing purchase quantities. This decomposition implies that a price discount is less effective on quantity choice once a consumer starts to think about the product. Therefore, putting a product on feature decreases the price elasticity in magnitude and alleviates the firm's burden to motivate consumers to consider and evaluate its product by using prices. We use a static supply side model to find that the consideration cost decrease from feature advertising implies a 5-9% increase in equilibrium markup.

Our paper contributes to both the descriptive literature on promotion threshold effects and to the structural literature on limited consideration. First, on the descriptive side, the previous literature has long noted that consumers show little response to small price discounts (Gupta and Cooper, 1992; Blattberg et al., 1995; Van Heerde et al., 2001; Briesch et al., 2002), but does not provide a deep explanation that relates to rational consumer decisions.[5] Our model of costly consideration contributes to the understanding of this empirical regularity, by providing a rational explanation to it.

Second, to the demand estimation literature with limited consideration (Goeree, 2008; Van Nierop et al., 2010; Dehmamy and Otter, 2014; Kawaguchi et al., 2016), a key question is whether one can separately identify lack of consideration from lack of preference, when the data only shows no purchase. The question is managerially important, because in the former case a firm can provide information to foster consideration and purchase, whereas the latter case is associated with product design or brand value. Yet identification between these two mechanisms is difficult in general without excluded variables that vary consideration costs independently from preference. Relative to the literature, we provide an alternative identification strategy that relies on quantity jumps from price variations. In particular, our identification strategy relies on observing "price thresholds," i.e. the highest price at which the consumer would purchase positive quantities, and the extent to which purchase quantity show discontinuities at these thresholds. Our usage of consumer level scanner data allows us to obtain a dense empirical support of price, which enables measurement of these price thresholds. We provide strong evidence of discontinuities in purchase quantity around these thresholds. With this identification strategy, we also propose direct tests for costly consid-

[5] Della Bitta and Monroe (1981) and Gupta and Cooper (1992) relate this phenomenon to reference point theories in psychology.

eration using standard marketing data, as well as a structural model that can be used to quantify these costs. In addition, substantively, our empirical results give insights into the effects of price discounts, and marketing strategies that aim at overcoming consumers' consideration barriers.

The rest of this paper is organized as follows. Section 2 briefly surveys the related literature. Section 3 presents the data. Section 4 then discusses the model. First, we outline an illustrative model and use it to discuss both identification of, and testable implications for costly consideration. Then, section 5 parametrizes the model and discusses estimation details. Sections 6 and 7 discuss parameter estimates and counterfactual implications. Finally, Section 8 concludes.

2 Related literature

This paper draws primarily from three strands of literature. First, it relates to the literature on promotion effects. Van Heerde et al. (2001) estimate a semi-parametric model using sales and price data, conditional on feature and display, and find that sales is unresponsive to small price changes, and is most responsive to moderately large discounts. Briesch et al. (2002) estimate a discrete choice model where the effect of deals is non-parametric, and find that utility on deals can be convex for yogurt. We complement their work by providing individual level evidence and a structural model that explains these findings. To this end, we adjust for consumer heterogeneity, product heterogeneity and store heterogeneity, and the interaction across these. Using detailed individual-level scanner data, we find clear evidence that an individual is less responsive to price discounts less than 30% off the regular price. We contribute to this literature also in the sense that we provide an explanation that rationalizes this behavior. This explanation further complements the psychology literature that attributes the unresponsiveness to a consumer's innate threshold (Gupta and Cooper, 1992).

Second, our work is related to the literature on limited consideration. Among earlier works, Shugan (1980) provides psychological justifications for the existence of a consumer thinking cost and an analytical framework of its implications. The empirical literature teases apart lack of attention (from e.g., large thinking costs) from lack of preference, in several ways. One way is to obtain a direct measure of attention: for example, Roberts and Lattin (1991) and Draganska and Klapper (2011) utilize survey data and directly elicit consideration decisions. Another way is to provide exclusion restrictions that only enter consideration but not purchase: Goeree (2008) assumes that advertising is informative, and advertising expenditure acts as an exclusion restriction in the utility function (given consideration); Ranjan et al. (2016) leverage natural-experimental variations in shelf-space allocation, observed from a retailer's planogram data, to identify the impact of consideration set on consumer choice and store revenue. A third way to test for the lack of attention is to examine consumer behavior that is inconsistent with any full information model: Clerides and Courty (2015) find that consumers continue to purchase large packages in the event that per-unit price of large packages exceed that of smaller packages; Kawaguchi et al. (2016) focus on variations in product availability and show that responses in consumer choices are inconsistent with a full information model.[6]

The most closely related work to us is Dehmamy and Otter (2014). This paper utilizes the sunk cost property of consideration in a consumer's decision of purchase quantity. Fixed costs do not enter quantity choice because they are sunk. Therefore, they propose that one can test for whether variables (such as the number of facings) affect consumption utility directly, by testing whether they affect quantity choice conditional on purchase. In their experimental application, they provide evidence that the number of shelf facings and the location on the shelf only affect consideration, and therefore are good excluded variables from utility. However, their methodology requires data on shelf facings and planograms, which are not always available.[7] In our framework, we highlight the discontinuity in quantity choices due to fixed consideration cost, and endogenous consideration decisions. Our model implies that prices (and potentially other product characteristics) affect these decisions, and generates a quantity jump at the price acceptance threshold. We provide reduced-form tests that uses standard data-sets, and propose an alternative structural model that takes consideration as a first-stage decision.

Further, our work is related to the literature on variety seeking (Kahn et al., 1986; McAlister, 1982) and multiple discrete choice. Hendel (1999), Kim et al. (2002), and Dubé (2004) model a consumer's product and quantity choice, and make simplifying assumptions to isolate the choice from quantity decisions. While this approach eases computation burden, it abstracts from competition for consideration set membership. In our paper, we assume that the consumer takes into

[6]Relatedly, Caplin and Dean (2015) propose a theory where agents rationally trade-off information acquisition costs with probabilities of ex-post mistakes.

[7]Ranjan et al. (2016) use planogram data to identify consideration costs, but do not utilize variations in quantity.

account the expected gain from purchase when she decides which product to include in her consideration set. Our model is in line with our proposed tests for limited consideration, but can also serve as an alternative model to characterize multiple discrete choice, when quantity decisions are isolated from choice. In addition, the model allows for the presence of nonlinear prices (quantity discounts) and discrete quantities (cf. Allenby et al. 2004).

Finally, Gilbride and Allenby (2004) discuss estimation issues related to two-stage decision models with decision heuristics. Our model falls into their characterization of a compensatory screening rule, where the utility of an alternative must exceed a given threshold. Their paper proposes a model and related estimation strategies, while our paper focuses on rationalizing the existence of such thresholds. Dzyabura and Hauser (2011) study consideration heuristics using machine learning methods.

3 Data and descriptive statistics

3.1 Construction

We use the 2001-2003 Behavioral Scan panel data from Information Resource Inc.'s (IRI) Academic Data Set (Bronnenberg et al., 2008).[8] We focus on the yogurt and yogurt drink categories. A "store visit" is recorded when a household purchases yogurt or yogurt drink in a trip. The data records, at the SKU level, the number of units the individual purchased in a given store-week, the total amount paid for the purchase, store level weekly data on the total units sold and revenue received on the given SKU, as well as product characteristics such as package size.

At the SKU level, prices are defined as store level revenue divided by units sold. Regular prices are defined as the 95th percentile of prices for an item in a given store-year. Discounts are defined as either absolute or percentage changes of prices from the regular price.[9] The data-set also records whether the product is on feature advertising, on in-store display, or on both.

Next, we aggregate the SKU level data to the level of a "product," defined as all SKUs with a

[8]In earlier versions, we checked to make sure that our reduced form and structural results are robust when adding data from 2004 to 2007.

[9]In an earlier version, we also defined regular prices as max price in the past 4 weeks. We also distinguished between "temporary" price discounts, where prices drop in one week but revert back to their previous regular prices in the following weeks, and "permanent" ones which are associated with a long-lasting price change. We do not find results to be different.

specific name recorded by the data regardless of the flavor or package size. For example, “General Mills Colombo Light” is considered as a “product,” whereas “General Mills Colombo Light, berry flavor, 8 oz” is a distinct SKU.[10] We consider the same product with different package sizes as different quantity options of a homogeneous product. To this end, we find the minimum *available* package size of a product in a given trip, and define “equivalent units” as total purchased volume divided by the minimum package size. For example, for a product with the minimum package size of 8 oz, an individual who purchased 1 unit of 8 oz, and 3 units of 16 oz ($8+3\times16=56$ oz), is considered to have purchased *7 equivalent units* ($7\times8=56$ oz). Since few consumers bought non-integer equivalent units,[11] we characterize quantity choice as discrete and round the non-integer choices to the nearest integer.

We average feature advertising and in-store display incidence to the product level, as sales-weighted averages in a given store and week from the SKU-level data. We use the same approach to aggregate discounts to the product level, which we use in our descriptive analysis only. We characterize price for a given quantity as follows. Starting with prices for each available SKU of a product, we find for each possible purchase quantity, the least expensive combination of different package size that achieves it. For example, price of two equivalent units is the less expensive of buying two units of the smallest package, or one unit of the twice-as-large package, of the same product. This captures discounts due to bulk purchase, frequently observed in this category.

As stated in the introduction, we subset the sample and only focus on trips with yogurt purchases. This condition assures that the consumer has been at the yogurt section and likely rules out costly price search.[12] When we later discuss alternative explanations to consumer fixed costs, this condition also rules out travel costs to the store and across different sections in the store (Pinna and Seiler, 2014).

Finally, to manage the computational burden of the structural analysis, we restrict our attention to products with a minimum package size smaller than 1 pint (16 oz). This precludes 24 out of 84 product but only removes 1/6 of the category sales volume. We also focus on the top 10 product (among the remaining 60), ranked in terms of sales volume, in order to maintain consistency

[10]In earlier versions of the paper, we defined a “product” as a combination of item name and flavorand reported similar descriptive results.

[11]A total of 89% of non-zero quantity choices involve integer equivalent units.

[12]We use the argument in Seiler (2013) that such consumers can find price information easily, compared to the consumers who did not buy products in the yogurt section.

Table 1: Set of products in the analysis

	vol share	share sold on disc.	reg price	price on disc.
Dannon Light N Fit	0.13	0.09	1.53	1.04
Yoplait Original	0.13	0.06	0.89	0.72
Colombo Classic	0.07	0.05	1.09	0.83
Colombo Light	0.07	0.02	0.82	0.55
Yoplait Light	0.07	0.03	0.83	0.68
Dannon Regular	0.06	0.03	0.81	0.53
Yoplait Thick	0.05	0.02	0.72	0.48
Danon Stonyfield Farm	0.04	0.02	2.48	1.77
Wells Blue Bunny	0.03	0.01	0.65	0.56
Coolbrands International Breyers	0.03	0.02	0.76	0.54

Note: This table presents the set of products, their in-sample sales volume share and share of volume sold on discount (defined as occasions when prices are 5% or more below the regular price). It also reports regular price and price given discount. These products are used both in descriptive and structural estimates.

between the reduced form and structural analysis. Table 1 summarizes the set of products in our analysis.

3.2 Summary statistics

3.2.1 Demographics

There are 8,397 households in the 2001-2003 sampling period. Taking a cross-section for the year 2003 (which consists of 6,558 unique households), we find that these households have an average size of 2.5 members, an average age of the household head of 53.7 years, and an annual income of $44,114. Household characteristics in other years are very similar.

3.2.2 Trips

On average, a household is observed for 56.5 weeks between the first and the last recorded trip containing yogurt purchases at a given retailer. In this period, there are 9.5 weeks on average with yogurt purchases. Within those 9.5 weeks, 5.3 weeks are associated with purchase of the top 10 products, and 3.0 weeks are with the top 4 products.

3.2.3 Products, prices and concentration

We compute in-sample market share, based on the shares of total consumer expenditure in yogurt, and find that concentration is moderate: overall, the top 10 products represent 61% of category sales. Among the top 10 products, average per-volume price is 0.14 dollar/oz, with a standard deviation of 0.07. The most popular sizes are 6 oz (51% of purchase occasions) and 8 oz (41%). The remaining 8% purchase occasions involve sizes of 1 or 2 pints.[13]

3.2.4 Discounts, feature and display

The use of feature and display is infrequent in the yogurt category. 8.4% of all product-store-week combinations have 50% or more of the SKUs on feature.[14] At 1.8% of all observations, displays are less common.

Figure 2 shows the distribution of price discounts separately for products on and not on feature or display. Discounts are frequently aligned with feature or display. That is, 80% of the products on feature or display are on a discount with no less than 5% price drop. On the contrary, conditional on being on feature or display, the discount distribution is bimodal: over 20% products are priced at regular prices whereas the products on discount have a mode discount of around 30%. Still, discounts are widely spread and a large fraction of discounts are between 0% and 30%.

We also report, in Appendix Table 6, that there are differences in the frequency and depth of discounts offered to different package sizes. Smaller pack-sizes (12 and 16 oz) are offered more frequent and deeper discounts than larger ones.

[13]Mostly from Dannon Light and Stonyfield. We adjust for availability of smaller sizes both in reduced form and structural analysis.

[14]The percentage of feature is calculated as a sales volume weighted average.

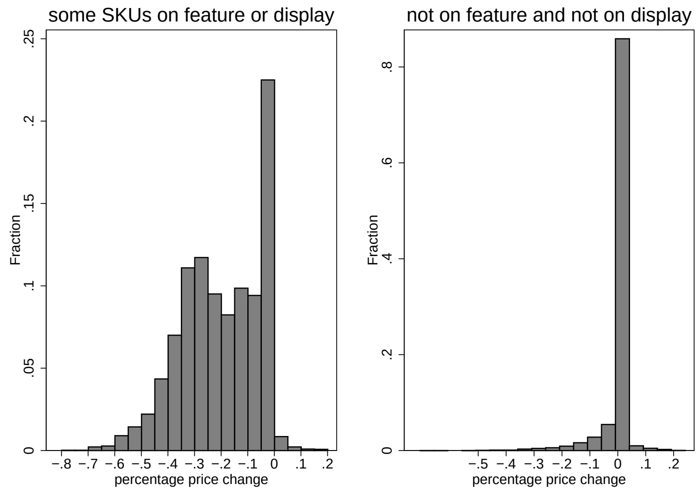

Figure 2: Distribution of price changes

Notes: This figure plots the distribution of percentage price changes. Regular price is defined as the 95% percentile of prices within product-retailer-quarter. Feature and display are defined as at least half the SKUs for a given product are on feature or display, respectively.

3.2.5 Variety and quantity per trip

Within-trip expenditure is heavily concentrated. In 76% of all shopping trips, consumers buy more than one (equivalent) unit of purchase, and 23% more than 5 equivalent units. Note that equivalent units are defined such that unit 1 is always available. However, consumers are not willing to spread the expenditure across different products. Across all trips with purchase, a consumer buys *one or two* products in 97% of the time. This pattern agrees with the literature on soft drinks (Dubé, 2004; Chan, 2006).

In sharp contrast to the concentration of products within a trip, we find that the total number of products purchased in the entire sample duration is much higher. Focusing on households who we observe between 20 to 40 trips that involve yogurt purchases,[15] we find that on average, a household purchases 1.2 distinct products per trip, while purchasing no less than 7.9 distinct products overall.[16] The large difference between these numbers suggests that a household does not focus on a narrow set of products in any given trip because of time-invariant preferences, which is in

[15]This selects a sub-sample of between median and 75th percentile in the trip distribution.
[16]For a full distribution of within and across trip variety, see Figure 10 in the appendix.

line with our theory where per-trip fixed costs will limit the amount of varieties in the short run. However, it could also be explained by variations of product characteristics such as price, feature or availability.

3.3 Aggregate price response curves

In this section, we formalize that demand is more responsive to large discounts. With individual level data, we first calculate the average purchase quantity, $\bar{q}_{ij}$, of an individual for a given product j (in a given store), when prices are within 15% of the regular price. We then compute the percentage change of demand in each given trip, $\frac{q_{ijt}-\bar{q}_{ij}}{\bar{q}_{ij}}$, and likewise, percentage change in price from the regular price. We then adjust for year and month fixed effects for the relative quantity measure we constructed and for price, and plot the two variables against each other.[17] Figure 3 presents local polynomial fit of such relations, pooled over all products and all occasions where the focal product is not on feature or display.

Because the axes represent percentage changes in quantity and price, the slope of the polynomial fit can be viewed as the point elasticity for a given price drop from the regular price, and the average slope of the figure can be viewed as average elasticity. We find that the unweighted average price elasticities is around -3.4, measured by the slope of a line cutting through (0.02, 0) and (0.6, 2). When we examine price elasticities conditional on discount depth, however, we find that the average price elasticity is -0.85 below 20% discount, -1.60 between 20% and 40% discount, and -8.02 between 40% and 60% discount. This comparison of elasticities shows that demand is "log-convex" in the sense that price elasticity can be increasing in the depth of discounts.

Figure 2 shows that the discounts offered by retailers are often "shallow." In particular, given that the product is not on feature or display at all (which is the same condition generating Figure 3), we find that 84% of the discounts offered are below 20%,[18] i.e., in the inelastic region of the individual price response curve. This finding suggests that the retailers in our sample might have considered the empirical consumer discount response when setting discounts.

It is also useful to highlight that such increasing elasticity patterns are not shared with typical

[17] We regress each of the percentage changes of quantity and price against year and month dummies, for the part of the sample where the percentage decrease in price is between [0.02, 0.6]. We then take the residuals, and adjust each by a constant so that Figure 3 goes through the origin.

[18] The standard definition of discount in the IRI data is that the discount is at or above 5%.

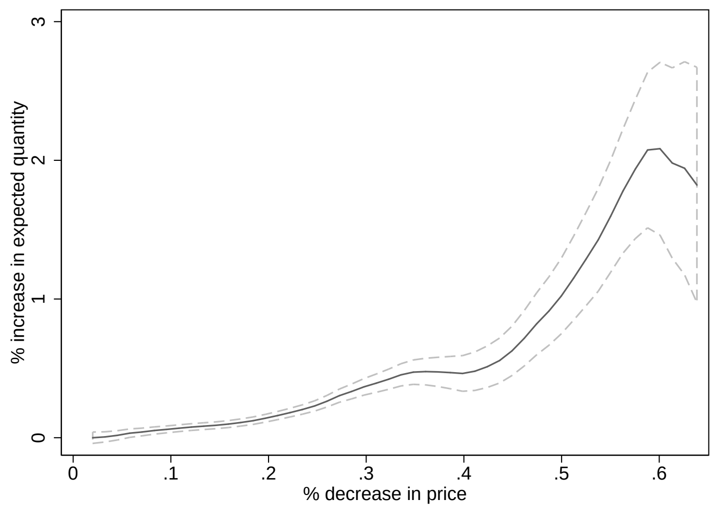

Figure 3: Convex response to discount

Notes: This figure shows the observed relationship between price and quantity in the yogurt category. The horizontal axis shows discounts from the regular price (in US dollars). The vertical axis shows the corresponding average increase in purchase quantity compared to the average purchase quantity under regular prices. We construct the figure conditioning on the (lack of) provision of additional price information (i.e. no feature or display) and normalizing the changes in quantity relative to the average quantity within consumer-store-product. We further control for year and month fixed effects.

demand system such as log-log or multinomial logit. A log-log regression implies constant elasticity, or a linear relationship in Figure 3. A multinomial logit model on product (but not quantity) choice implies that elasticity is $-\alpha \cdot price \cdot (1 - share)$ where $-\alpha$ is the price coefficient.[19] This means that logit elasticity is decreasing in market share, and the relationship in Figure 3 should be concave.[20]

4 Illustrative model and identification

4.1 Overview

Figure 3 shows that consumer demand is increasingly responsive to price discounts. In this section, we first construct an empirical model of endogenous consideration set formation that rationalizes

[19] Under the condition that utility is linear in price.

[20] An important point of difference with our model is that the logit demand system captures choice but not quantity, i.e., the comparison is not completely informative if quantity demand to rises disproportionally in price discounts.

this pattern. In our model, the consumer costlessly observes prices of a product, but needs to incur costs to "consider the purchase" – i.e., to acquire other information before the purchase decision. Because the consumer decides on quantity after consideration, such consideration costs are fixed to quantity. This makes consideration more likely at high discounts, because of the potential to purchase multiple units in such occasions. We numerically illustrate that such a model with endogenously-determined consideration sets after observing prices can accommodate the convex demand function, similar to Figure 3.

Next, we provide a testable implication from the model: a consumer would rather not consider at all than considering the product and buying a very small quantity. This is because a small purchase quantity does not generate sufficient utility to justify the consideration cost. We test this hypothesis by reporting the purchase quantity distribution at each consumers' maximum accepted price, and find that the purchase quantity has a median of 4 units when the choice of 1 unit is feasible. In addition, in Appendix C, we numerically illustrate that with large fixed costs, the purchase probability (incidence) will drop to zero at high prices but the average quantity given purchase remains high. We also empirically find that, controlling for full consumer heterogeneity, individual consumer demand shows the same discrepancy between incidence and quantity.

Then, we show that the model can be identified when the only source of variation is price. The intuition is that incidence and quantity responds differently to price changes and the response pattern resembles a model with fixed costs. We run Monte Carlo simulations to show that parameters are very precisely estimated when we use the correct parametric model. When we do not know the correct consumption utility functional form (but instead approximate it by finite order polynomial), we show that key parameters including fixed costs can be estimated precisely. This suggests that consideration costs can be identified from price variation alone.

4.2 An illustrative model: setup

In this section, we present a simplified model that characterizes the purchase decisions of a given product for a representative consumer. We simplify the model in order to illustrate key testable implications of it, as well as to show how parameters of the model are identified. We extend the model to incorporate multiple varieties in Section 5.

An individual i wishes to purchase at least one unit of yogurt and travels to the refrigerated product shelf in trip t. To convey the essence of our argument, it suffices to focus on the purchase decisions of a single product. The consumer first observes the unit price of the product and decides whether it is optimal for her to consider the product for purchase. From this she discovers her match value to the product but incurs effort costs. Upon consideration, she can then choose what quantity to purchase, q_{it}, with the possibility of $q_{it} = 0$ symbolizing the choice of "other varieties" (the outside option).

The consumer solves the problem backwards. Given consideration, we specify her indirect consumption utility as

$$c_{it}\left(p_t, q_{it}\right) = \beta \log\left(q_{it}+1\right) - \alpha p_t q_{it} + \mu_{it}\left(q_{it}\right) \tag{1}$$

where $\beta \log\left(q_{it}+1\right)$ captures that consumption utility is increasing in quantity and that marginal utility is decreasing, $-\alpha p_t q_{it}$ captures the dis-utility on expenditure, and $\mu_{it}\left(q_{it}\right)$ captures other product characteristics or match value that are unobserved to the researcher, and unobserved to the consumer prior to consideration. For analytical simplicity, we model these match values as quantity-specific utility shocks that are Type I Extreme Value with scale parameter σ_μ.

Match values $\mu_{it}\left(\cdot\right)$ are revealed through effort or costly consideration. For example, consideration might involve picking up a product and evaluate its caloric or dietary content. Alternatively, consideration might involve planning ahead on which days in the coming week to consume yogurt. Yet different, it may involve comparison to substitute products or categories. The effort involved in consideration is a fixed cost $f + \Delta\varepsilon$, and the decision is rational in the sense that the consumer weighs the expected net consumption utility against this fixed cost. Specifically, the consumer observes prices, and expects to receive total utility

$$u_{it} = \underbrace{\mathbb{E}\left[\max_{q\in Q}\left(c_{it}\left(p_t, q\right)\right)|p_t\right]}_{\text{option value from consideration}} - \underbrace{\left(f + \Delta\varepsilon_{it}\right)}_{\text{fixed costs}} \tag{2}$$

if she considers the product, or zero if she does not consider. Note that the expectation term integrates over potential information on μ –unobserved prior to consideration– therefore also the optimal consumption quantity.

The consumer chooses to consider the product if she gets positive expected total utility, i.e. $u_{it} > 0$. Given consideration, she chooses quantity that maximizes (1).

4.3 Testable implication: Quantity jump at the threshold price

We present and test the key implication for our model. For a consumer with a given draw of fixed cost shocks, $\Delta\varepsilon_{it}$, consideration reduces to a threshold price rule in the sense that the consumer will consider with certainty when prices are below a threshold $\bar{p}$. This threshold price property eliminates small quantity demand, since the expected utility from consumption at high prices does not justify spending the fixed cost. Our model implies that her purchase quantity distribution at the maximum "accepted" price, as a proxy for the threshold price, will be bounded away from low values.

To numerically show this, we simulate optimal quantity choice from the model defined in (1) and (2) at parameters $\beta = 3$, $\alpha = 1$, $f + \Delta\varepsilon_{it} = 4$ and $\sigma_\mu = 2$, and allow quantity choices to take values in $Q = \{0, 1, ..., 12\}$. We take 100,000 draws of μ_{it} and prices, and plot the average purchase quantity conditional on price in Figure (4). The figure shows that there is a discrete jump in quantity at the threshold price because quantity demand at higher prices are "suppressed" – as they do not generate high enough expected consumption utility to justify the consideration fixed costs. Therefore, if the researcher knows the threshold price $\bar{p}_t$ for each consumer-trip, she can then test for the presence of a consumer fixed cost f, by testing whether the quantity at a price *slightly* below $\bar{p}_t$ – i.e. the threshold quantity $\bar{q}_{it}$ – is significantly larger than zero.

Taking this intuition to our data, we construct the threshold price $\bar{p}_{ij\tau}$ as the maximum accepted price, i.e. the highest price at which we observe individual i making a purchase of product j within the given year τ. We do this by conditioning on observing at least 5 purchase occasions for a given product within the year, and we take the highest purchase price among the 5 or more occasions. We compute the threshold discount level by taking regular price minus the threshold price, in order to be able to compare across products having very different price levels. Thus, the threshold discount is the price discount that just makes a consumer buy. Next, we characterize the purchase quantity distributions for each individual and product, at different threshold values, given that the discount level is within 2.5 cents difference to her discount threshold, and that the individual purchases at

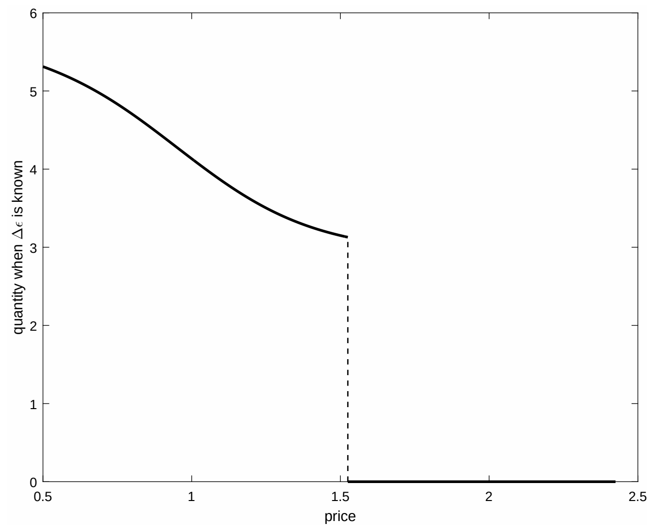

Figure 4: Quantity jump at the threshold price

Notes: The solid curve illustrates the average purchase quantity at different prices, from simulation exercises where demand is implied by (1) and (2). We take $\beta = 3$, $\alpha = 1$, $f + \Delta\varepsilon = 4$ and $\sigma_\mu = 2$. Choice set is discrete with $Q = \{0, 1, ..., 12\}$.

the current discount level. Recall that quantity is measured in multiples of the minimum available package size. Therefore, a quantity of 1 is available by definition for each product in each trip.

We plot the quantity distribution at the threshold discount across combinations of individuals and products with different discount thresholds, resulting in Figure 5. This figure shows that median purchase quantity at the discount threshold is 3-5 times the minimal available package size. The 25th percentile is 2 or 3, and 75th percentile is mostly 5, 6, or 7. This figure clearly documents the prevalence of quantity jumps at the discount thresholds. These jumps are consistent with a theory of (endogenous) costly consideration.

4.4 Identification

4.4.1 Intuition

Figure 4 shows that, when the fixed cost is deterministic, the researcher observes under which threshold price the consumer considers the product with probability one. Therefore, below the threshold price, consumption utility is identified by the distribution of purchase quantities as a function of prices. Furthermore, at the threshold price, consideration cost equates the total ex-

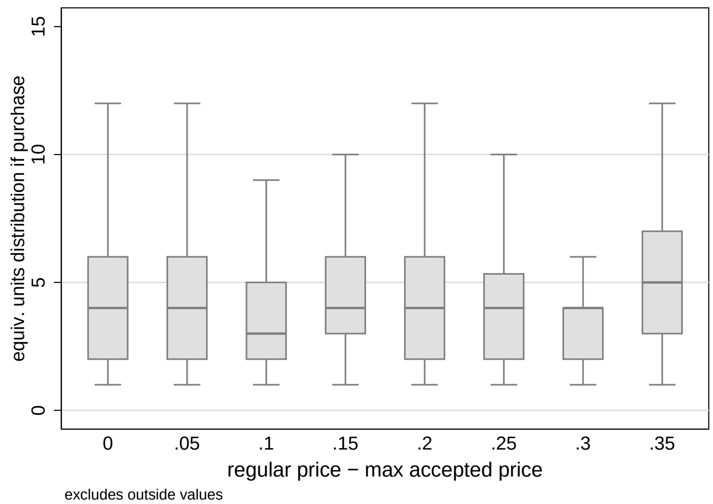

Figure 5: Quantity response to price for the marginal consumer

Notes: On the horizontal axis, we plot the difference between regular price and the maximum accepted price of a consumer, for a given product-retailer-year. On the vertical axis, we plot quantiles in the purchase quantity distribution, given that the consumer purchases at prices that are within $0.025 of the max accepted price. The grey box represents 25% and 75% percentile, the center bar represents the median, and the outer range represent 5% and 95% percentiles.

pected value from consumption. Therefore, the location of threshold price identifies the size of consideration cost.

When there are random components in the consideration costs ($\Delta\varepsilon$), these costs generate a gap between quantity given purchase and purchase probability. Specifically, when prices are high, purchase incidence is low despite that quantity given purchase is high. In our simple model, the only element that explains this gap is a positive consideration cost f. A high f results in most consumers not thinking about the product when prices are high, therefore lowering the purchase probability considerably while not changing quantity at purchase.

4.4.2 Monte Carlo simulations

We run Monte Carlo simulations to confirm that we can uniquely pin down consumption utility parameters and fixed cost parameters separately (see Appendices C and D). We first simulate choices of quantity under different prices by a homogeneous set of individuals, under the same model and parameters that generate Figure 4. We find that all parameters can be estimated precisely and without bias, implying that consideration cost is identified by price variations being the only excluded variable from the fixed costs. That is, one does not have to resort to additional exclusion restrictions such as advertising. Appendix Table 3 presents the results.

We further confirm that our results are not driven by specific functional form assumptions, by estimating models with a more flexible utility function than our data-generating process. Our results confirm that one can reliably estimate model parameters when consumption utility functional form is unknown but approximated by a polynomial. These results are presented in Appendix D.

4.5 Alternative explanations

4.5.1 Stockpiling

If the underlying product is storable, and if consumers are dynamic optimizers with correct beliefs about the arrival of discounts, their demand might follow a threshold rule such as the one in Figure 4.

We therefore test for stockpiling in our yogurt data. Specifically, we regress purchase incidence and quantity given purchase, either on time since last purchase or on the past 2 weeks of prices, controlling for current prices, individual fixed effects, and flexible specifications of time dummies. Consumers who take advantage of a price discount to stock up will purchases less in the near future; therefore, past prices should have positive effects on current sales if stockpiling is a concern. Similarly, consumers who stock up should purchase more in time elapsed since last purchase. We estimate linear specifications both at the category- and the product level, shown in Table 2 and Appendix Table 7 respectively. Our results do not support stockpiling behavior.[21]

4.5.2 Unobserved individual and product characteristics

In Figure 3, we measure the consumer response to price discounts by purchase quantity deviations from their means within consumer-store-product. This normalization absorbs, in a multiplicative way, time invariant heterogeneity at the consumer-store-product level. One might in addition worry that the figure pools over consumers with different price sensitivities or products (or time periods)

[21]Columns 1 and 3 are similar to tests in Hendel and Nevo (2006), who find weak evidence for stockpiling in the yogurt category compared to laundry detergent and soft drinks. Our evidence is opposite to what a stockpiling model would predict, but suggests that consumers with stronger category preference buy more frequently. In earlier versions, we instrument inter-purchase time by past preferences and do not find any effect on current purchase decisions, either. Product-level evidence tells the same story.

Table 2: Test for consumer stockpiling: category level

	incidence	incidence	logvolume	logvolume
months since last purchase	-0.03***		-0.03***	
	(0.00)		(0.00)	
– squared	0.00***		0.00***	
	(0.00)		(0.00)	
log price	-0.21***	-0.17***	-0.41***	-0.34***
	(0.01)	(0.02)	(0.02)	(0.04)
– past week		-0.02		0.03
		(0.02)		(0.04)
– past 2 weeks		0.01		0.01
		(0.02)		(0.04)
years since 200101	0.05***	0.05***	0.02***	0.01
	(0.00)	(0.00)	(0.00)	(0.01)
Apr-Jun	-0.02***	-0.01*	-0.02***	0.02
	(0.00)	(0.01)	(0.00)	(0.01)
Jul-Sep	-0.03***	-0.02***	-0.01***	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)
Oct-Dec	-0.03***	-0.02**	-0.04***	-0.04***
	(0.00)	(0.01)	(0.00)	(0.01)
individual FE	Yes	Yes	Yes	Yes
obs.	152631	26361	126312	22087

Notes: Category level evidence for stockpiling both in the extensive and intensive margin. Standard errors are reported beneath the parameter estimates. 1, 2 and 3 stars indicate significance levels at 10%, 5% and 1%, respectively.

with different price elasticities. In Appendix B, we flexibly test the average price elasticities within the same retailer-product and within a set of similar consumers and find robust evidence for log-convex demand.

4.5.3 Seasonality

Another alternative explanation to Figure 3 is that demand and prices are both seasonal. If low prices are set in seasons with high price elasticity (e.g. if pricing is counter-cyclical, as in Nevo and Hatzitaskos, 2006 and Haviv, 2015), we will see a log-convex relationship between quantity and price. While we do control for seasonality in Figure 3, in Appendix A (Table 1), we further examine the presence of seasonality in demand level, price elasticity and prices. We find that while there are some seasonality in the demand level, price elasticity do not display seasonal changes. At the same time, there is no economically significant seasonal variations in prices. Thus, we conclude that Figure 3 can not be generated from seasonality in our data.

4.5.4 Quantity discounts

The final alternative explanation is associated with different promotion strategies across package sizes. If large packs are discounted heavily such that consumers mainly buy large packs and do so when they are on sale, one might find that some consumers buy much higher volume at slightly reduced average (across package size) prices. Whereas our structural model fully controls for this explanation through flexible functions of price on quantity, we further examine the discount depth and frequency between small and large pack sizes. We find that smaller sizes are on discount more frequently and are discounted more heavily when so. They are also put on feature and display more frequently.[22] The differences between promotions across package sizes do not support the aforementioned alternative explanation to our descriptive evidence.

[22] See Appendix Table 6 for detailed results.

5 Structural model and implementation

5.1 Overview

We generalize the model in Section 4.2 to be able to (1) accommodate more products in the choice set, (2) allow for choice of multiple products at the same time, and (3) account for observed characteristics and demographics as covariates. This section provides details on parametrization of each part of the model, the solution to the consumer problem, and implementation in estimation.

Consumer i maximizes utility by choosing quantities, denoted $\mathbf{q}_{it} = (q_{i1t}, q_{i2t}, ..., q_{iJt})'$. To choose any product the consumer has to first consider it, and we denote the (endogenous) consideration set as $K_{it} \subset J$. Not choosing any of the J products means buying from the outside option, which contains all other products in the yogurt category. Recall that because of the way we construct the sample, all individuals purchase some yogurt products in the trip.

In our empirical implementation, we limit dimensionality by restricting the total number of products in the consideration set to be at most 2, i.e. $||K|| \leq 2$. Recall from our data description that this still captures 97% of all trips with yogurt purchases. Limiting the consideration set size to 2 is therefore not far from reality in our context and greatly reduces computation burden.

5.2 Parametrization

5.2.1 Consumption utility

We specify the (indirect) consumption utility as

$$c_{it}(\mathbf{q}_{it}, \mathbf{p}_t) = \sum_{j \in K_{it}} \mathbf{x}'_{ijt} \beta_i \log(q_{ijt} + 1) + \gamma \prod_{j \in K_{it}} \log(q_{ijt} + 1) + \alpha \sum_{j \in K_{it}} p_{jt}(q_{ijt}) \cdot q_{ijt} + \mu_{it}(q_{ikt}, q_{ilt}) \tag{3}$$

where the consumption sub-utility is specified in log, and defined on discrete quantities $q_{ijt} \in \{0, 1, ..., \bar{q}\}$. The functional form of consumption utility is similar to Kim et al. (2002) and Dehmamy and Otter (2014). The log specification in quantity implies decreasing marginal utility, and is consistent with love-for-variety preference. To make the model less restrictive, we allow for additional love for variety captured by γ multiplying interactions between (log) utility of different products. A positive γ means that love for variety is stronger than implied by the "sum of log quantity" specification, while a negative γ implies weaker love for variety preference, i.e., a negative γ captures the degree of substitutability between products. Finally, as a matter of definition, when the choice set is singleton, one of the log-quantity terms becomes zero, effectively setting the interaction effect to zero.

$\mathbf{x}_{ijt}$ is a *vector* of indicators of product characteristics, individual characteristics and time. Specifically, it contains a vector of 4 brand dummies, an indicator for the characteristic "light", household size, and a linear time trend. The product $\mathbf{x}'_{ijt}\beta_i$ captures the marginal utility for percentage increases in quantity. For example, Dannon Light has characteristics "Dannon" ($x_{j1} = 1$) and "light" ($x_{j5} = 1$), and therefore the marginal utility for log quantity is $(1,0,0,0,1,hhsize_i,t) \cdot (\beta_{i1},\beta_{i2},\beta_{i3},\beta_{i4},\beta_{i5},\beta_6,\beta_7)' = \beta_{i1} + \beta_{i5} + \beta_6 \cdot hhsize_i + \beta_7 \cdot t$. Note that demographic and time coefficients are constant across households.

Prices affect decisions via the indirect utility function. We allow the per-product expenditure $p_{jt}\left(q_{ijt}\right) \cdot q_{ijt}$ to be non-linear in quantity, to capture the potential quantity discounts that consumers could benefit from, by buying in large quantities. Recall, $p_{jt}\left(q_{ijt}\right)$ is the *lowest* (across different quantity combinations) per-unit price one could get when choosing total quantity q_{ijt}. α is the price coefficient, held fixed across households.

The consumer also receives random consumption utility shocks, $\mu_{it}\left(q_{ikt}, q_{ilt}\right)$, unobserved by the researcher, which capture taste shocks that favor, or oppose, purchasing a specific quantity combination of products k and l. We assume, for a particular combination of quantity (q_1, q_2), that $\mu_{it}\left(q_1, q_2\right)/\kappa$ is i.i.d. type-1 extreme value distributed, where κ is a scale coefficient. We make the i.i.d. assumption for model tractability: in presence of the computation burden in estimation, this simplifying assumption is crucial as in this way we can explicitly solve for the inclusive value from quantity choice rather than resorting to simulation. Gentzkow (2007) uses a similar specification of utility shocks over combinations of newspapers.[23] With this assumption, one should interpret the μ's as shocks to match value, rather than unobserved product characteristics.[24] We also assume

[23]As in Gentzkow (2007), we model utility shocks as bundle specific to allow for a flexible specification of substitution or complementarity in consumption. Our approach allows one to characterize choice probabilities directly instead of first order conditions and therefore can be used to estimate fixed costs. An alternative approach to modeling quantity choices is Kim et al. (2002), who model marginal utility shocks as error terms to quantity choice. Although easier to interpret, estimation of their model relies on first order conditions on the optimal quantity, which cannot be used to estimate fixed costs. In addition, their model is less flexible in capturing non-smoothness in the quantity distributions in the data.

[24]In an earlier version, we controlled for product fixed effects both in $c_{it}(\cdot)$ and consumer fixed costs and obtained

that μ_{it} is realized *after* the consumer forms consideration set K_{it}. One can interpret the fixed costs F_i, specified next, as the costs of drawing utility shocks.

5.2.2 Consideration costs

The consumer also incurs a consideration cost, $F_{it}(K_{it})$, as a function of her consideration set. Denote M_{ijt} as the number of months since the consumer purchased j last time, and A_{ijt} whether a product is on feature advertising. In addition, we denote $\mathbf{M}_{it}$ and $\mathbf{A}_{it}$ as the vector of M_{ijt} and A_{ijt} over all products, and we adopt the convention that $M_{ijt} = \infty$ if the consumer has not purchased j prior to t. Now, we parametrize the fixed cost as

$$\begin{aligned} F_i(K_{it}, \mathbf{A}_{it}, \mathbf{M}_{it}) \quad = \quad & \sum_{j \in K_{it}} \left[f_{i0} + f_M \cdot \log\left(M_{ijt}+1\right) \mathbf{1}\left(M_{ijt} \neq \infty\right) + f_N \cdot \mathbf{1}\left(M_{ijt} = \infty\right) + f_A \cdot \mathbf{1}\left(A_{ijt} = 1\right) \right] \\ & + f_2 \cdot \mathbf{1}\left(||K_{it}|| = 2\right) - \varepsilon_{iKt}, \end{aligned} \tag{4}$$

where f_{i0} denotes the baseline per-product consideration cost for individual i, which is common across products. f_M is the change in fixed cost in log months since purchase of the product, f_N is the additional fixed cost for consumers who never purchased the product before, f_A is the increase (negative means reduction) in fixed cost when the product is on feature, and f_2 is the additional total consideration cost when considering two products.

The consumer also incurs an unobserved (by the researcher) cost shock ε_{iKt}, specific to set $K = K_{it}$, which are independent type-1 extreme value random variables, across individual, trip and all potential sets $K \subset J$. In addition, for tractability, ε_{iKt}'s are independent of $\mu_{it}(q_{ikt}, q_{ilt})$.

5.3 Solution of optimal choice rules

5.3.1 Quantity given consideration set

The consumer maximizes expected utility and solves the consumer problem backward. In the second stage, given that the consumer decides to consider products in set K_{it}, she reveals consumption utility shock $\mu_{it}(.)$ and chooses the quantity combination within K_{it} that maximizes utility in equation (3). In other words, she chooses (q_{ikt}, q_{ilt}) given $K_{it} = \{k, l\}$ ($l = 0$ in case of a single-product

similar results to the ones reported below.

consideration set). Note that fixed costs and the cost shocks are sunk and are irrelevant to the purchase decision once the consideration set K_{it} is fixed.

Given the Type I Extreme Value assumption on μ, to the econometrician, quantity choice is multinomial logit on combinations of quantity. Denote $\bar{c}_{it}\left(q_{ikt}, q_{ilt}\right) = \sum_{j=k,l}\left(\mathbf{x}'_{ijt}\beta_i \cdot \log\left(q_{ijt}+1\right) - \alpha \cdot p_{jt}\left(q_{ijt}\right) \cdot\right.$ $\gamma \cdot \prod_{j=k,l}\log\left(q_{ijt}+1\right)$, choices over discrete quantity sets follow a multinomial logit probability

$$\Pr\left(q_{ijt}, q_{ikt} | K_{it}; \theta_i\right) = \frac{\exp\left(\bar{c}_{it}\left(q_{ikt}, q_{ilt}\right)/\kappa\right)}{\sum_{\left(q'_k, q'_l\right)\in Q^2}\exp\left(\bar{c}_{it}\left(q'_k, q'_l\right)/\kappa\right)}, \tag{5}$$

where θ_i denotes all relevant parameters. Because the quantity support Q includes zero, the set Q^2 of possible quantity combinations includes buying *nothing*, or buying from only one product.[25]

5.3.2 Choice of the consideration set

Consideration is costly and the set K_{it} is a choice. On the one hand, the consumer spends effort considering products in a given set K_{it}, incurring cost $F_i\left(K_{it}, \mathbf{A}_{it}, \mathbf{M}_{it}\right)$ as defined in (4). On the other hand, before considering the product and revealing μ_{it}, she does not perfectly predict her optimal choice, and evaluates the expected option value from a given consideration set in terms of the following "inclusive value" term, which is the expected maximum total utility from set K_{it}. Denote state $\mathbf{S}_{it} = \left(\mathbf{p}_t, \mathbf{A}_{it}, \mathbf{M}_{it}\right)$ for notational simplicity. Combining the expected gain from consideration and the expected fixed cost, the net expected utility from considering a set K_{it} is:

$$\bar{v}_i\left(K_{it}, \mathbf{S}_{it}\right) + \varepsilon_{iKt} = \mathbb{E}\left[\max_{\mathbf{q}\in Q^2}\left(c_{it}\left(\mathbf{p}_t, \mathbf{q}\right)\right) | K_{it}, \mathbf{p}_t\right] - F_i\left(K_{it}, \mathbf{A}_{it}, \mathbf{M}_{it}\right), \tag{6}$$

where $\mathbf{q}$ can have at most two positive elements because we restrict the size of consideration set K_{it}. Note that from the GEV assumption on μ, one can derive that

$$\bar{v}_i\left(K_{it}, \mathbf{S}_{it}\right) = 0.577 + \kappa \cdot \log\left(\sum_{\left(q'_k, q'_l\right)\in Q^2}\exp\left(\frac{c_{it}\left(q'_k, q'_l\right)}{\kappa}\right)\right) - F_i\left(K_{it}, \mathbf{A}_{it}, \mathbf{M}_{it}\right), \tag{7}$$

[25] For singleton consideration sets, the quantity support reduces to the one-dimension support Q.

where 0.577 is the Euler constant. Then, given the type I extreme value cost shocks ε_{iKt}, we can express the probability of choosing a consideration set K_{it} as:

$$\Pr\left(K_{it}|\mathbf{S}_{it};\theta_i\right)=\frac{\exp\left(\bar{v}_{it}\left(K_{it},\mathbf{S}_{it}\right)\right)}{\sum_{K'\in\mathscr{K}}\exp\left(\bar{v}_{it}\left(K',\mathbf{S}_{it}\right)\right)},\tag{8}$$

where $\mathscr{K}$ is the set of all possible consideration sets (up to the size limit of 2) – including $\emptyset$.

5.4 Construction of the likelihood function

5.4.1 Matching the observed choice probability

We have characterized the choice probability of consideration set K, and the probability distribution of purchase quantities given K. In our data, we observe the choice probabilities for specific quantity combinations. We treat these as the empirical realizations of the following probabilities from our model:

$$\Pr\left(q_{ikt},q_{ilt}|\mathbf{S}_{it};\theta_i\right)=\sum_{K'\supset\{k,l\}}\Pr\left(q_{ikt},q_{ilt}|K',\mathbf{S}_{it};\theta_i\right)\cdot\Pr\left(K'|\mathbf{S}_{it};\theta_i\right).\tag{9}$$

5.4.2 Likelihood with random coefficients

Given that each time series of choices by one individual is generated under each individual's independent realization of random coefficients, we can write the likelihood across the individual-trips, as

$$\mathscr{L}\left(\theta\right)=\prod_i\left(\int_{\theta_i}\left(\prod_t\Pr\left(q_{ikt},q_{ilt}|\mathbf{S}_{it};\theta_i\right)\right)dG\left(\theta_i;\theta\right)\right),\tag{10}$$

where $\left(q_{ikt},q_{ilt}\right)$ are observed quantities. The solver then minimizes $-\log\left(\mathscr{L}\left(\theta\right)\right)$ with respect to parameter θ.

5.4.3 Simulated maximum likelihood

Whereas the choice probabilities at the individual level exist in closed form and are exact, the integral over θ_i is not analytical and needs to be approximated by simulation. To implement the simulated maximum likelihood method, we first take D draws of random coefficients *shocks* on β_i and f_{i0}, denoted $\hat{\beta}_d$ and $\hat{f}_{d0}$ for draw d. Each dimension of the random coefficients is first independently drawn from standard normal distribution $\mathcal{N}(0,1)$, and then adjusted in scale and location by model parameters. For example, the d^{th} draw of fixed cost parameter is $f_{d0} = \bar{f}_0 + \sigma_f \cdot \hat{f}_{d0}$, where $\bar{f}_0$ and σ_f are mean and standard deviation of the coefficient. We restrict the interaction term coefficient γ and price coefficient λ to be homogeneous across individuals.

We then maximize the likelihood function with respect to the parameters, i.e., the mean and standard deviation of random coefficients, taking the draws as given. For a given parameter value, the empirical counterpart of the likelihood function is given by

$$\hat{\mathscr{L}}(\theta) = \prod_{i=1}^{N} \left(\frac{1}{D} \sum_{d=1}^{D} \left(\prod_{t} \Pr\left(q_{ikt}, q_{ilt} | \mathbf{S}_{it}; \theta_d\right) \right) \right). \tag{11}$$

5.5 Other details

5.5.1 Construction of the sub-sample

To restrict computation burden at a reasonable level, we implement the structural model on a random sub-sample of 5% of individuals in the data (422 households),[26] over all their in-sample trips. Because of dimensionality concerns, we focus on the 10 products that generate the highest overall sales (which consist of 56.2% of the total in-sample sales volume), and treat the rest as outside options. The set of products, with their respective share of sales volume, are listed in Table 1. Finally, as previously indicated, we only characterize consideration sets of sizes 0, 1 or 2.

5.5.2 Construction of the quantity set

We measure purchase quantity by volume in pints, so that consumption utility can be compared across products with different package sizes. Because the set of quantity available is not continuous, we define choice set to be multiples of the minimum package size, q_j^{min} – which is specific to the product but constant across trips. In the data, large quantity choices are scarce, but they should be allowed in the model. To balance computational burden and a realistic range of quantities, we discretize large quantity choices to coarse grids, so that the set of quantity one can choose from is $Q_j = q_j^{min} \cdot \{0,1,2,3,5,8,12\}$, where we bundle choices of 4-6 units into quantity 5, 7-9 into

[26] In earlier versions, we compared model estimates on this sample and on a sample with 10% households, and find that they are essentially the same.

quantity 8, and 9-20 units into quantity 12.[27]

5.5.3 Product, household and time characteristics

Given the choice of product set, we choose to focus on four characteristics – 3 brand indicators (Dannon, Yoplait and Colombo), an "other brand" indicator,[28] and the indicator for characteristics "light." We conjecture that demand for quantity depends on how many members consume yogurt within a household. Therefore, we include household size as a key predictor of the marginal consumption utility. We also include a (linear) time trend into the utility function. Given the mixed evidence for seasonality in demand, we do not add seasonality in the consumption utility.

5.5.4 Distribution of number of products and purchase quantity

We find that the distribution of the number of products in the sub-sample closely represents that in the full sample. In particular, as stated in Appendix Table 8, 0.95% of our sub-sample purchased more than 2 different products. Allowing for 3 products in the consideration set will increase the number of alternatives (consideration set - quantity combinations) from 1,680 (10 single-product cases and 45 duo-product cases, each product allowing for 6 quantity levels) to 27,600 (adding 120 triple-product sets, each with 6^3 quantity combinations), and the 0.95% observations cannot justify the additional computation burden. As a side note, zero products indicates purchase of another yogurt not in the set of interest – so we condition on category purchase.

5.5.5 Choice-based sampling

From Table 8, we find that there are many observations with no purchase, and few observations with 2-product purchases. However, the model structure demands rich information in quantity choice given purchase, in particular multiple-product choice. Relatedly, the "no purchase" observations are not very informative of the consumption utility functional form. In light of this, we under-sample these observations to gain computation speed, and correct for the stratified sampling approach in the likelihood function, in the way proposed by Manski and Lerman (1977).

[27]In the likelihood, we compensate for the width of quantity grids; for example, the observed probability of the (continuous) quantity falling into $[4,6)$ is 3 times the model probability of choosing (discrete) quantity 5. In other words, our model treats the world as if there are only 6 possible quantity levels (plus zero quantity).

[28]That is, those top-10 products that are not produced by the aforementioned 3 brands.

Table 3: Parameters estimates

	par. est.	std. err.
util. coef. for Dannon ($\bar{\beta}_1$)	4.61	0.38
util. coef. for Yoplait ($\bar{\beta}_2$)	4.85	0.39
util. coef. for Colombo ($\bar{\beta}_3$)	4.84	0.39
util. coef. for Other Manufacturers ($\bar{\beta}_4$)	3.59	0.38
util. coef. for Light ($\bar{\beta}_5$)	-0.09	0.13
util. coef. for household size ($\bar{\beta}_6$)	0.10	0.07
util. coef. for (100x)weeks ($\bar{\beta}_7$)	-0.01	0.00
interaction term in utility (γ)	-4.66	0.55
price coef. (α)	-1.76	0.12
baseline consideration cost ($\bar{f}_0$)	1.43	0.35
changes in consid. cost under feature (f_A)	-0.84	0.19
changes in consid. cost in log(months-since-purchased + 1) (f_M)	0.56	0.05
changes in consid. cost if never purchased (f_N)	7.65	0.24
changes in consid. cost for two products (f_2)	-1.22	0.46
scale of utility shock (κ)	2.01	0.16
std. dev. of brand coef. ($\sigma_{\beta,1}$)	1.02	0.08
std. dev. of characteristics coef. ($\sigma_{\beta,5}$)	0.12	0.26
std. dev. of mean consid. cost (σ_f)	1.40	0.11

Note: Estimates for the all parameters. Quantities are measured in pints. Standard errors are asymptotic (numerical).

Specifically, within the 5% sub-sample, we draw 30% random sample among observations with no purchase (of inside goods), and 80% random sample among single-product purchase occasions; at the same time, we keep all observations with two-product choice sets. This reduces the sample size used in estimation from 6,816, to 4,472, saving approximately 1/3 of the original computation time.

6 Estimation results

6.1 Consumption utility estimates

Table 3 reports all parameter estimates. The $\bar{\beta}'s$ capture the marginal consumption utility for log-transformed quantity, which depends on brand, light versus regular, household size, and a linear calendar time trend.

By defining random coefficients on product characteristics, we capture the within-consumer correlation in demand, so that, for example, consumers who like Yoplait products will have higher choice probabilities on both Yoplait Original and Yoplait Light. The mean brand coefficients suggest that, on average, consumers have a slightly higher marginal utility for Yoplait products or products from "other brands," than for Dannon and Colombo. The "light" coefficient is insignificantly different from zero: this means that low-calories or organic products are approximately equally favored. Expansion of household size results in higher consumption utility. The effect is not statistically significant, plausibly due to lack of variation of household size within a household. The time trend estimate suggests a small negative trend in quantity demand.

Using the log-transforms of quantity restricts the curvature of consumption utility for a single product, which implies love-for-variety. In addition, parameter γ captures how purchase quantity of multiple products substitute one another, conditional on consideration. The magnitude of γ suggests that products are close substitutes.

6.2 Fixed consideration or purchasing costs

$\bar{f}_0$ captures the baseline purchasing or consideration cost. Given the way we set up the cost structure in (4), the baseline represents fixed costs for consumers who just purchased the same product in the previous week, when the product is not under feature, and when the consumer only considers one product. Using the estimates for the baseline (money metric) fixed cost, $\bar{f}_0/\lambda$, we find that the average fixed costs are \$0.81 for consumers who just bought the product in the past week. This magnitude is close to a quarter of the average per-trip expenditure on a single yogurt product.[29]

However, for a consumer who is less familiar with the product, the fixed costs to include the product in her consideration set are estimated to be much higher. For new consumers who never purchase a given product, the fixed cost is $\left(\bar{f}_0 + f_N\right)/\lambda$, or \$5.16.[30] This is more than 6 times the consideration cost for a regular consumer. For those who purchased a while ago, their fixed cost is in between the two extremes: for example, the fixed cost for consumers who have not purchased the product in a year is \$3.08 on average,[31] which is 280% higher than when the consumer's

[29]The mean (and median) expenditure given purchase is \$3.26 (and \$2.56) per product.

[30]The estimates in f_N suggest that consumers who never purchased the product have much higher fixed cost. However, this pattern might be explained by unobserved heterogeneity in the product tastes beyond the individual random effects. We do not rely on this causal interpretation in our counterfactual exercise.

[31]We arrive at this number by $(1.43 + 0.56 * \log(12+1))/1.76$.

memory is fresh. Finally, putting the product on feature decreases consumer fixed costs by \$0.48, or 59% of the baseline fixed costs.[32] In addition, the second product to be considered incurs very small additional fixed cost, a form of scale economy when buying multiple products. Together, these findings suggest that variations in information – either from past experience or from retailer provision in the form of feature – reduce consumer fixed costs. In turn, these findings imply that a large part of the fixed costs estimated here are related to costly consideration or evaluation.

We also contrast the distribution of consideration and purchase set sizes. On average, 90% of consumers consider two products, 9% consider one, and almost no consumer does not consider any product. In contrast, 30% consumers purchase two products, 59% purchase one, and 11% does not make a purchase.

It should be emphasized that these results are produced under the restriction that consideration set sizes do not exceed two.[33] With this constraint, consideration of two products incurs additional opportunity costs because it precludes consideration of other products. Therefore, our estimates of consideration or purchasing fixed costs are conservative estimates.

6.3 Model fit

Figure 6 top panel presents the fit of our model with respect to the empirical distribution of quantity given purchase.[34] The model fits the empirical purchase probabilities to within 1% for most products. Given purchase, the model-predicted quantity distributions resemble similar shapes to that of the data. However, there are spikes in the data distribution of purchase quantities that the model does not rationalize. A potential explanation is that our model contains a continuous consumption utility function, and any particular spike in our data (that breaks smoothness) can only be rationalized by a particular low unit price at that quantity.

In the lower panel of the same figure, we predict price response separately for each product, by calculating the average purchase quantity given price. We find that the model predicts the convex

[32]Our model excludes feature in the consumption utility because there is no clear economic mechanisms for feature to drive the intensive margin given consideration. In an alternative version, we allow feature to enter consumption utility as well as fixed cost. We confirm that feature in fact does not drive the intensive margin in a statistically significant way and the rest of the parameter estimates are similar.

[33]Without such restriction, the computation burden makes the model intractable.

[34]Note that in the top panel of Figure 6, the choice probabilities for large quantities – say quantity 5 – are computed as the average of choosing one quantity in its interval (say 4, or 5, or 6). In this way, it matches with the model predicted choice probabilities according to the choice set $\{1, 2, 3, 5, 8, 12\}$ in the model.

price response reasonably well.

6.4 Price elasticities and decomposition into consideration and choice

We next present the price elasticities implied by the model computed as arc-elasticities. That is, we reduced, one at a time, prices for each of the 10 products by 15% and computed the changes in choice shares. To integrate out heterogeneity, we compute quantity choices based on each of 20 draws in the random coefficient, and then average them across draws.

We report the full set of price elasticities in Appendix Table 9 and summarize them here. Own and cross elasticities are conventional, and the magnitude intuitive. For example, the own price elasticities across all products are in the range $[-3.2, -2.2]$. Under the assumption that retailers act as monopolists over shoppers in the store, these elasticities imply a sizable markup, of roughly 30%-50% for most products. These are consistent with the literature on yogurt. For example, Villas-Boas (2007) finds markups to be around 40% in most settings.[35]

Cross elasticities show higher substitution for products that are closer in characteristics. For example, a change in price of Yoplait Original has (relatively) large impact on the market shares of Yoplait Thick and Yoplait Light, and vice versa. This is driven by the random coefficients interacted with product characteristics.

In Table 4, we decompose price elasticities into consideration incidence elasticity and quantity elasticity given consideration set. Specifically, for each product, we simulate the probability of which it falls into the consumer's consideration set, and measure the elasticity of such incidence with respect to a 15% change of the own price. We also simulate purchase quantity conditional on each product falling into potential 2-product consideration sets (weighted by the corresponding consideration probabilities), and use it to measure the elasticity of purchase quantity given consideration. We find that roughly a third of the total demand response to a price change can be attributed to consideration incidence, the rest to quantity choice given consideration.

This result offers a different view of incomplete information. While costly information of price (e.g. Diamond, 1971) will attenuate price response and soften competition, costly information of other product characteristics after the consumer knows price can intensify price response. In our

[35] Her own elasticity estimates are larger and cross-elasticities are smaller. The difference might come from that she models choices on the product-flavor level while we focus on the product level.

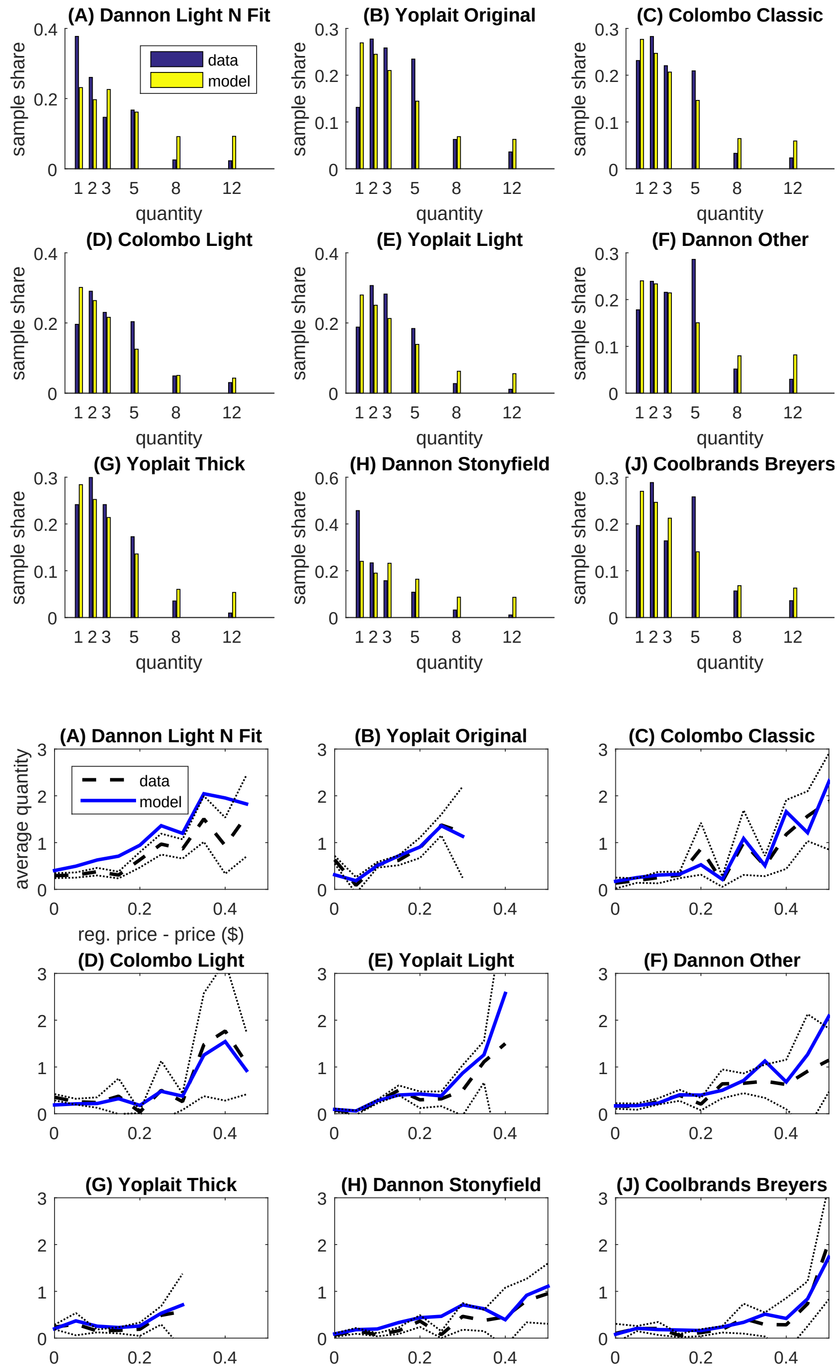

Figure 6: Model fit: purchase probability and quantity

Notes: The top graph compares the model-predicted quantities conditional on purchase with empirical distribution of purchase quantity, for 9 out of 10 products. Note that the quantity distribution is re-weighted by quantity grid width; for example, the sample frequency of choosing quantity 5 is 1/3 of the frequency of choosing 4, 5 or 6. The bottom graph plots model-predicted average purchase quantity against average quantity in the data, conditional on price.

Table 4: Elasticities of overall quantity, consideration incidence and quantity given consideration

	overall	consideration	quantity if consider
(A) Dannon Light N Fit	-2.38	-0.95	-1.40
(B) Yoplait Original	-2.77	-0.96	-1.60
(C) Colombo Classic	-2.94	-0.71	-1.12
(D) Colombo Light	-3.15	-0.81	-1.11
(E) Yoplait Light	-3.07	-1.09	-1.67
(F) Dannon Other	-2.62	-0.59	-1.60
(G) Yoplait Thick	-3.17	-1.07	-1.68
(H) Dannon Stonyfield	-2.67	-0.48	-1.58
(I) Wells Blue Bunny	-2.24	-0.18	-1.33
(J) Coolbrands Breyers	-2.69	-0.47	-1.51

Note: The left column are own price elasticities of total demand. The middle column are consideration incidence elasticities measuring percentage changes in the probability that each product falls into the consideration set, to a change in price. The right column are elasticities measuring percentage quantity change conditional on consideration set membership, which is operationalized as quantity given a product is in each of the 10 possible two-product consideration sets, weighted by the probability that each of the 10 sets occur.

setting, prices are incentives for a consumer to incorporate a product into the consideration set, and we show that removing such incentive (through complete removal of the extensive margin) reduces price elasticities by roughly a third.[36]

6.5 Semi-elasticity to feature

We compute the differences in consumer purchase quantities while placing, in turn, each of the 10 products on and off feature, for all trips in the sample. The results show that having a product on feature increases sales by 25-38%. Because the effect of feature goes through consideration costs, these results imply large consideration cost sensitivity. Appendix Table 10 shows the full semi-elasticity table to feature.

[36]For the top 5 products, the elasticity in the extensive margin represents between 24% and 40% of the total elasticity.

7 The role of costly consideration on demand and prices

7.1 Counterfactual: all consideration costs changed by feature

From our estimates, we find that feature reduces consideration costs by 0.84 in utility terms, or \$0.48 in monetary-equivalence if weighted by the price coefficient. In this section, we switch feature for each product on and off simultaneously to study the impact of feature-driven consideration costs on consumer demand, price elasticities and implied markups.

7.2 Impact of consideration costs on demand

Comparing two worlds where all products are on and off feature at the same time, we separate the effect of feature on purchase incidence of a given product (extensive margin) and quantity choices conditional on purchase (intensive margin). We find that feature advertising increases the share of consumers buying two products and decreases the share of those buying zero or one, and as a result increases the total number of chosen products by 1.7%. This is a small effect compared to when there is only one product on feature (Section 6.5), showing that feature has strong business stealing effects but small category expansion effects on demand given category purchase.

7.3 Impact of consideration costs on market prices

In the classical search literature (e.g. Diamond, 1971), limited information on price attenuates price response and increases equilibrium price. Differently, in our model, consideration sets are smaller when fixed costs increase. Since consumers in our model use price information to make decisions about investing time considering the product for purchase, firms will intensify price competition to fight for costly (and thus scarce) consumer attention. Therefore, fixed consideration costs per variety or product lead to more intense competition for consideration set membership and puts downward pressure on equilibrium prices. Conversely, lowering such costs (for example, by feature) will increase market prices. In this section, we study the impact of setting all products to feature on static equilibrium prices. Although it is unlikely in reality that all products feature at the same time, our counterfactual experiment can be considered as simulating an information shock that lowers consumer fixed cost by \$0.48.

We simulate the equilibrium markup, following Berry et al. (1995), under the assumption that a retailer acts as a monopolist when setting prices for each of the products it carries. Specifically, we consider the static pricing decision of each retailer r trying to maximize flow profit Π_{rt} at week t, from all the products in its assortment J_r:

$$\Pi_{rt} = \sum_{j \in J_r} \left(p_{jt} - mc_{jt}\right) \cdot sales_{jt}\left(\mathbf{S}_t\right) \tag{12}$$

where state $\mathbf{S}_t$ includes all prices and other observables.[37] Take first order conditions on (12) and we have

$$\frac{\partial \Pi_{rt}}{\partial p_{jt}} = sales_{jt}\left(\mathbf{S}_t\right) + \sum_{k \in J_r} \left(p_{kt} - mc_{kt}\right) \cdot \frac{\partial sales_{kt}\left(\mathbf{S}_t\right)}{\partial p_{jt}} = 0. \tag{13}$$

The above can be re-written in vector form over all products in a time period:

$$\mathbf{p}_t = \mathbf{mc}_t + \Delta^{-1}\left(\mathbf{S}_t\right) \cdot \mathbf{sales}_t\left(\mathbf{S}_t\right) \tag{14}$$

where $\Delta\left(\mathbf{S}_t\right)$ is the ownership matrix defined as

$$\Delta_{jk} = \begin{cases} -\frac{\partial sales_{kt}\left(\mathbf{S}_t\right)}{\partial p_{jt}} & \text{if } j \text{ and } k \text{ are sold by the same retailer} \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

We first simulate the impact of the counterfactual change in fixed costs on own-price elasticities for all products. These are calculated in the same way as Table 9, and are presented in columns 1 and 3 in Table 5. We find that own-price elasticities are dampened in magnitude, by about 5%, when feature advertising lowers the fixed costs by \$0.48.

We also simulate equilibrium markup, across the two scenarios with different consideration costs. In line with our intuition and the elasticity calculations, it is not surprising to see that implied markup increases in a market with lower fixed costs – that is, when every product is on feature. Specifically, a \$0.48 decrease in fixed costs increases markup by 5-9% depending on the product.

[37] We compute sales quantity given observed states as the sum of expected choice probabilities, $sales_{jt}\left(\mathbf{S}_t\right) = \sum_i \Pr\left(q_{ikt}, q_{ilt} | \mathbf{S}_{it}; \theta_i\right)$.

Table 5: Own-price elasticities and implied equilibrium markup

	no feature: elasticity	markup	all feature: elasticity	markup	percent diff
(A) Dannon Light N Fit	-2.39	0.55	-2.30	0.59	0.08
(B) Yoplait Original	-2.80	0.51	-2.67	0.55	0.09
(C) Colombo Classic	-2.98	0.41	-2.89	0.43	0.06
(D) Colombo Light	-3.23	0.36	-3.13	0.38	0.06
(E) Yoplait Light	-3.10	0.34	-2.99	0.36	0.08
(F) Dannon Other	-2.66	0.39	-2.60	0.41	0.05
(G) Yoplait Thick	-3.25	0.34	-3.12	0.37	0.08
(H) Dannon Stonyfield	-2.69	0.38	-2.59	0.41	0.06
(I) Wells Blue Bunny	-2.24	0.27	-2.16	0.29	0.05
(J) Coolbrands Breyers	-2.79	0.32	-2.71	0.34	0.06

Note: The first two columns present own-price elasticity and implied markup $-\Delta^{-1}(\mathbf{S}_t) \cdot \mathbf{sales}_t(\mathbf{S}_t)-$ setting all products *off* feature. The third and fourth columns present elasticity and markup setting everything *on* feature. The last column calculates percentage differences in markup, setting the on-feature markup as baseline.

8 Concluding remarks

In this paper, we quantify the consumer's time or mental cost of considering and selecting a product. These costs generate scale economies in consumers' quantity choices and encourage purchase of larger quantities instead of many varieties. They also explain the existence of a quantity threshold and, relatedly, why consumers are unresponsive to small price discounts.

With many observations at the individual consumer level, we demonstrate how the existence of quantity thresholds – under price discounts that are just enough to convert them to purchase – can be used to empirically disentangle consideration from preference. That is, without costly consideration, consumers would gradually re-allocate their expenditure in response to gradual price changes, rather than switching in larger quantities from one product to another in a manner consistent with the presence of price thresholds. We demonstrate how one can test for the presence of these thresholds using standard marketing scanner data, without imposing a structural model.

We then estimate a two-stage structural demand model. In the first stage of this model, consumers select which products to consider for choice as the outcome of the trade-off between utility from variety and disutility from investments in decision effort. In the second stage, consumers make variety choices and subsequent quantity choices. Estimating this model using data from the yogurt category, our results indicate that consumers have large consideration costs associated with purchasing a product. These costs are even larger if the consumer has not purchased the product for a long time, or ever. Our estimates suggest that inertia in product choice is partly due to informational frictions. Price discounts act as incentives to invest time into overcoming these frictions. We also quantify the role of feature advertising and past-purchase experience in this framework.

Our model is manageable with a small number of varieties, but is still computationally intensive. Future work may focus on developing computational methods to estimate larger versions of our model. A substantive limitation of our paper is that it models the supply side in a simple way. A richer supply side model can formally address whether consumer fixed costs (and their discrete price response) can explain the way firms switch between regular price and deep discounts. We relegate this to future research.

References

Allenby, Greg M., Thomas S. Shively, Sha Yang and Mark J. Garratt (2004), 'A choice model for packaged goods: Dealing with discrete quantities and quantity discounts', *Marketing Science* **23**(1), 95–108.

Berry, Steven, James Levinsohn and Ariel Pakes (1995), 'Automobile prices in market equilibrium', *Econometrica: Journal of the Econometric Society* pp. 841–890.

Blattberg, Robert C., Richard Briesch and Edward J. Fox (1995), 'How promotions work', *Marketing Science* **14**(3), pp. G122–G132.

Briesch, Richard A, Pradeep K Chintagunta and Rosa L Matzkin (2002), 'Semiparametric estimation of brand choice behavior', *Journal of the American Statistical Association* **97**(460), 973–982.

Bronnenberg, Bart J., Michael W. Kruger and Carl F. Mela (2008), 'Database paper-the iri marketing data set', *Marketing Science* **27**(4), 745–748.

Caplin, Andrew and Mark Dean (2015), 'Revealed preference, rational inattention, and costly information acquisition', *The American Economic Review* **105**(7), 2183–2203.

Chan, Tat Y. (2006), 'Estimating a continuous hedonic-choice model with an application to demand for soft drinks', *The Rand Journal of Economics* **37**(2), 466–482.

Clerides, Sofronis and Pascal Courty (2015), 'Sales, quantity surcharge, and consumer inattention', *Review of Economics and Statistics* (0).

Dehmamy, Keyvan and Thomas Otter (2014), 'Utility and attention-a structural model of consideration', *Working Paper* .

Della Bitta, Albert J and Kent B Monroe (1981), 'A multivariate analysis of the perception of value from retail price advertisements', *NA-Advances in Consumer Research Volume 08* .

Diamond, Peter A. (1971), 'A model of price adjustment', *Journal of Economic Theory* **3**(2), 156–168.

Draganska, Michaela and Daniel Klapper (2011), 'Choice set heterogeneity and the role of advertising: An analysis with micro and macro data', *Journal of Marketing Research* **48**(4), 653–669.

Dubé, Jean-Pierre (2004), 'Multiple discreteness and product differentiation: Demand for carbonated soft drinks', *Marketing Science* **23**(1), 66–81.

Dubé, Jean-Pierre, Günter J. Hitsch and Puneet Manchanda (2005), 'An empirical model of advertising dynamics', *Quantitative Marketing and Economics* **3**(2), 107–144.

Dzyabura, Daria and John R Hauser (2011), 'Active machine learning for consideration heuristics', *Marketing Science* **30**(5), 801–819.

Gentzkow, Matthew A. (2007), 'Valuing new goods in a model with complementarity: Online newspapers', *American Economic Review* **97**(3), 713–744.

Gilbride, Timothy J. and Greg M. Allenby (2004), 'A choice model with conjunctive, disjunctive, and compensatory screening rules', *Marketing Science* **23**(3), 391–406.

Goeree, Michelle (2008), 'Limited information and advertising in the us personal computer industry', *Econometrica* **76**(5), 1017–1074.

Gupta, Sunil and Lee G. Cooper (1992), 'The discounting of discounts and promotion thresholds', *Journal of Consumer Research* pp. 401–411.

Haviv, Avery (2015), 'Does purchase without search explain counter-cyclic pricing?'.

Hendel, Igal (1999), 'Estimating multiple-discrete choice models: An application to computerization returns', *The Review of Economic Studies* **66**(2), 423–446.

Hendel, Igal and Aviv Nevo (2006), 'Sales and consumer inventory', *The RAND Journal of Economics* **37**(3), 543–561.

Kahn, Barbara E., Manohar U. Kalwani and Donald G. Morrison (1986), 'Measuring variety-seeking and reinforcement behaviors using panel data', *Journal of Marketing* **23**(2), 89–100.

Kawaguchi, Kohei, Kosuke Uetake and Yasutora Watanabe (2016), 'Identifying consumer inattention: A product-availability approach', *Working Paper* .

Kim, Jaehwan, Greg M. Allenby and Peter E. Rossi (2002), 'Modeling consumer demand for variety', *Marketing Science* **21**(3), 229–250.

Manski, Charles F. and Steven R. Lerman (1977), 'The estimation of choice probabilities from choice based samples', *Econometrica: Journal of the Econometric Society* pp. 1977–1988.

McAlister, Leigh (1982), 'A dynamic attribute satiation model of variety seeking behavior', *Journal of Consumer Research* **9**, 141–150.

Nevo, Aviv and Konstantinos Hatzitaskos (2006), 'Why does the average price paid fall during high demand periods?'.

Pinna, Fabio and Stephan Seiler (2014), 'Consumer search: Evidence from path-tracking data'.

Ranjan, Bhoomija, Paul Ellickson and Mitch Lovett (2016), 'The effect of location and assortment on category consideration, learning, and choice'.

Roberts, John H. and James M. Lattin (1991), 'Development and testing of a model of consideration set composition', *Journal of Marketing Research* pp. 429–440.

Seiler, Stephan (2013), 'The impact of search costs on consumer behavior: A dynamic approach', *Quantitative Marketing and Economics* **11**(2), 155–203.

Shugan, Steven M. (1980), 'The cost of thinking', *Journal of consumer Research* pp. 99–111.

Van Heerde, Harald J., Peter S. H. Leeflang and Dick R. Wittink (2001), 'Semiparametric analysis to estimate the deal effect curve', *Journal of Marketing Research* **38**(2), 197–215.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel and Philip Hans Franses (2010), 'Retrieving unobserved consideration sets from household panel data', *Journal of Marketing Research* **47**(1), 63–74.

Villas-Boas, Sofia Berto (2007), 'Vertical relationships between manufacturers and retailers: Inference with limited data', *The Review of Economic Studies* **74**(2), 625–652.

Walker, AM (1968), 'A note on the asymptotic distribution of sample quantiles', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 570–575.

Online Appendix

A Seasonality in quantity and prices

We examine seasonality in sales quantity and prices at the store-product-year-month level. Specifically, we estimate

$$\log\left(Q_{jst}\right) = \alpha_{\tau} \log\left(P_{jst}\right) + \theta_{js} + \lambda_{\tau} + \iota_{y} + \varepsilon_{jst} \tag{16}$$

where j is a product, s is a store and t is time in year-month. We denote τ as quarter and y as year, and control for quarter (of the year) and year fixed effects in addition to product-store fixed effects. In one specification, we allow for price elasticity to differ across quarter of the year.

Appendix Table 1: Seasonality in quantity and prices

	log units	log units	log price
log price	-1.63***	-1.64***	
	(0.03)	(0.04)	
log price (Apr-Jun)		0.02	
		(0.03)	
log price (Jul-Sep)		0.02	
		(0.03)	
log price (Oct-Dec)		0.02	
		(0.03)	
Apr-Jun	0.06***	0.06***	0.00
	(0.02)	(0.02)	(0.01)
Jul-Sep	0.07***	0.07***	0.01*
	(0.02)	(0.02)	(0.01)
Oct-Dec	-0.13***	-0.13***	0.01***
	(0.02)	(0.02)	(0.01)
Year 2002	0.01	0.01	0.01
	(0.02)	(0.02)	(0.01)
Year 2003	0.14***	0.14***	-0.03***
	(0.02)	(0.02)	(0.01)
product-store FE	Yes	Yes	Yes

Note: Regression on year and quarter dummies for log quantity and log prices, on the level of store-product-month.

Table 1 Column 1 shows that there is significant seasonality in the demand level. Stores sell 20% less yogurt in winter compared to summer. Column 2 shows that despite these differences in levels, price elasticity do not show significant differences both statistically and economically, suggesting that consumers with different elasticities do not sort into buying in different times in a year. In other versions, we also control for year-month fixed effects and the price elasticities do not change.

Further, we examine seasonal variation in prices. We estimate

$$\log\left(P_{jst}\right) = \theta_{js} + \lambda_{\tau} + \iota_{y} + \varepsilon_{jst} \tag{17}$$

and find that prices do not vary across seasons in an economically significant way (despite that prices are statistically significant higher by 1% in winter). We conclude that the extent of seasonality in our context is limited to parallel shifts of the demand curve.

B Test for log-convexity within products and consumer groups

The increasing elastic demand that we find could be due to differences in characteristics or pricing patterns between products. For example, some products use deep discounts more frequently and sell more, and those observations could concentrate on the right end of Figure 3. Similarly, there might be heterogeneity in consumer price sensitivity correlated with their shopping patterns. For example, if price sensitive consumers travel to the category only when discounts are deep, they will be on the right end of the figure and thus explain the convexity.

To address these issues, we control for heterogeneity and test whether demand is increasingly elastic in discount depth among similar individuals and products. To this end, we group data by product/retailer, by households of the same size, and condition on price discounts within a certain range. We estimate a linear-log specification controlling for a time trend:

$$q_{ijt} = \beta_g + \delta_t + \alpha_g \log\left(price_{jt}\right) + \varepsilon_{ijt} \tag{18}$$

where g is a "group", i.e. a unique combination of product, retailer and household size, given price discounts falling into given 15% grids.[38] Note that q_{ijt} can take value 0. After estimating α_g, we then compute an estimate for elasticity $\mathscr{E}_{gt} = \frac{\partial \bar{q}_{gt}}{\partial price_{jt}} \frac{price_{jt}}{\bar{q}_{gt}} = \alpha_g / \bar{q}_{gt}$ where $\bar{q}_{gt}$ is the average quantity for the consumer-retailer-product group.

[38] We use 0-15%, 15-30%, 30-45% and 45-60%, each range including the lower bound.

Appendix Table 2: Ratio of price elasticity at different price range

	median	std err of median
15-30% discount	1.3	0.1
30-45% discount	1.8	0.2
45-60% discount	1.4	0.2

Notes: Estimates of price elasticity at different price range, for the same product and for the same "group" of consumers with similar observables. This is executed according to Equation (18).

Within each unique combination of product, retailer and household size, we take the ratio between elasticity at discount grids above 15% and elasticity when discounts are between 0 and 15%. Table 2 presents the median elasticity ratio because the mean is driven by extreme values. We find that elasticities at lower prices (larger discounts) are generally larger than elasticities around the regular price, in particular the median ratios between elasticities at larger discounts compared to a 0-15% discount are all significantly larger than 1 within a group of similar consumers.[39] Cases with larger discounts suffer from limited power issue because we cut the data into very thin subsamples, although the magnitude of elasticities at 30-45% does seem to be considerably larger.[40]

These results show that within a consumer segment elasticities increase with discount depth, and that our results are not simply due to aggregation over heterogeneous consumers.

C Additional testable implication: gap between quantity and incidence

In addition to Section 4.3, we present another way to test the implication of the model. When fixed cost shock $\Delta\varepsilon_{it}$ is unobserved, consideration is a threshold-crossing decision with a random price

[39] We compute the standard error of the sample median by

$$se(\hat{m}(x)) = \frac{1}{2\hat{f}(x)\cdot\sqrt{N}}$$

where x is the ratio of elasticity, $\hat{f}(x)$ is the Kernel density of x evaluated at the median, and N is the sample size of x, i.e., the number of consumer groups. This formula follows the asymptotic distribution for sample quantiles (Walker, 1968, page 570).

[40] We alternatively estimate (18) within individual-product-store. We find qualitatively similar results but the sample is too thin to draw statistical inference.

threshold. At given parameters, we numerically show that the model implies high quantity demand given consideration, at both low and high prices. In particular, at high prices where consideration is unlikely, quantity given consideration remains to be large. This implies that purchase decision is price sensitive and quantity given purchase is less price sensitive, and can be tested in data.

We numerically simulate purchase and quantity decisions, and calculate the implied purchase probability as a function of price, as well as average quantity given purchase as a function of price. We maintain the parameters in the previous example, $\beta = 3$, $\alpha = 1$, $f = 6$, $\sigma_\mu = 2$ and choice set $Q = \{0, 1, ..., 12\}$, but we now set $\Delta\varepsilon$ to be logistic random variable. The average quantities and purchase probabilities are summarized in Figure 7. Compare the left panel with Figure 4 and one finds that the un-censored part of the average quantity is exactly the same – because only fixed cost shock $\Delta\varepsilon$ is different between the two figures and it is sunk after consideration. However, while average quantity given purchase remains away from 1 at all prices, purchase probability is very low at high prices. This is because expected purchase of around 2 units will unlikely justify spending large fixed costs and the consumer will choose not to consider the product as a result.[41]

The gap between quantity and incidence can be tested with consumer purchase data. One potential test is to plot quantity given purchase and purchase probability *across* consumers, but it does not rule out the possibility that consumers who choose to purchase at high prices are selected with favorable preferences. When selection is involved, we will find sizable average purchase quantity at high prices without costly consideration. To control for selection, we estimate simple parametric models of quantity demand given purchase and purchase incidence, separately for each consumer and for each product. Specifically, we take a sub-sample of consumers who purchase a given product for more than 10 times in the data, and estimate a constant-elastic demand function:

$$\log\left(q_{ijt} | q_{ijt} > 0\right) = \gamma_{0ij} + \gamma_{1ij} \log\left(p_{jt}\right) + \eta_{ijt} \tag{19}$$

in order to capture quantity given purchase. We also estimate a simple logit model for each consumer,

$$\mathbf{1}\left(q_{ijt} > 0\right) = \mathbf{1}\left(\delta_{0ij} + \delta_{1ij} p_{jt} + \Delta\varepsilon_{ijt} > 0\right) \tag{20}$$

[41] If we set average fixed costs to zero in this model, the gap between quantity and incidence disappears. When β is large or the scale of consumption utility shock σ_μ is large, both quantity and incidence remain large at high prices; when these parameters are small, both are low at high prices. This illustrates that sunk cost is the only factor that explains the difference between the two figures.

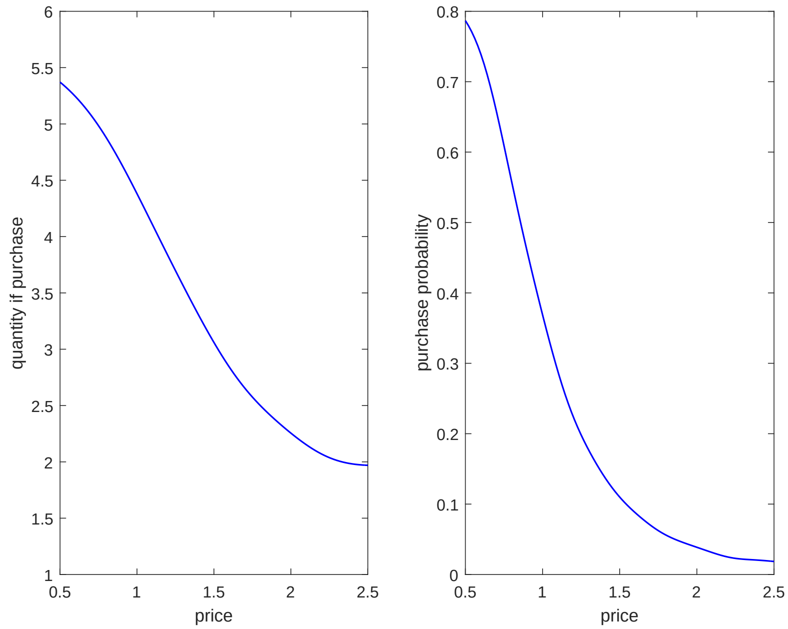

Figure 7: Average quantity given purchase and average purchase probability

Notes: Numerical illustration of model-predicted quantity given purchase (left panel) and purchase probability (right panel). These two figures are simulated from choices across 100,000 trips.

in order to capture purchase incidence. We focus on trips where the focal product is not on feature or display, and we also control for year dummies because there is a clear time trend in price. Next, given estimates of γ and δ, we predict expected quantity given purchase and purchase probability for each consumer at all prices. Finally, average across predicted quantity and incidence at given prices between \$0.4 and \$1.2. This procedure ensures that we do not summarize quantity across different *sets* of consumers at different prices.

The average quantity and incidence is plotted separately for the top 4 products in Figure 8. Within reasonable prices, we see choice probability drops close to zero at high prices while quantity given purchase stays bounded away from 1 – which is by construction feasible. This is in line with a theory of costly consideration set, and cannot easily be explained by rational choice theory without some form of fixed acquisition costs.

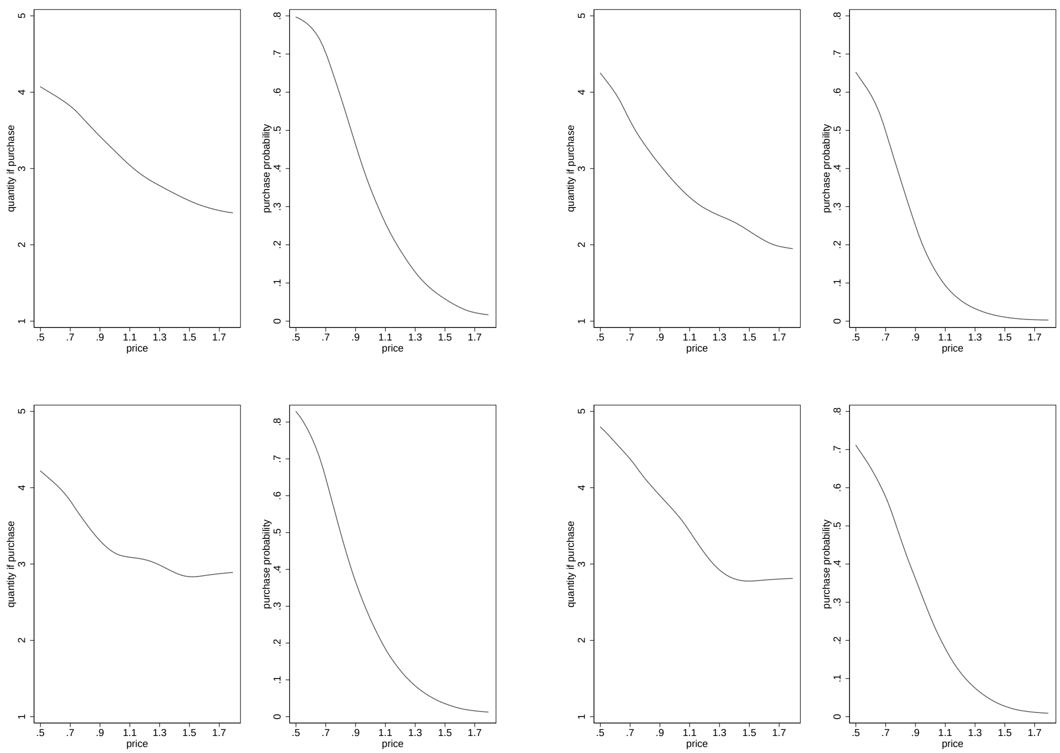

Figure 8: Within-individual predictions on quantity and incidence: top 4 products

Notes: By-individual-product predictions of quantity given purchase (left panel in each block) and purchase probability (right panel), then averaged across individuals over a fixed set of prices. The top blocks are (left to right) Dannon Light and Yoplait Original, and the bottom two are Colombo Classic and Colombo Light.

D Monte Carlo results

We specify a simple model as in Section 4.2, and try to estimate model parameters without additional exclusion restrictions. First, we assume that the analyst has correctly specified utility functional form as in the data-generating process. We take $b = 3$, $a = 1$, $, f = 6$, $\sigma_\mu = 2$ and $\sigma_\varepsilon = 1$, simulate data and estimate parameters in 100 experiments, take the difference between the true parameter and the estimates, and document its mean, median, standard deviation, and mean squared error. To make sure that our Monte Carlo result do not only work with one set of parameters, we perform another set of 100 experiments with $b = 2$, $a = 1$, $, f = 3$, $\sigma_\mu = 2$ and $\sigma_\varepsilon = 1$. The results are presented in Table 3. In both cases, we find that we can precisely recover model parameters without bias.

Appendix Table 3: Monte Carlo results with parametric utility

Panel A: $(b,a,f,\sigma_\mu,\sigma_\varepsilon) = (3,1,6,2,1)$

	mean	median	stdev	rmse
α - true α	0.00	-0.00	0.04	0.04
f - true f	-0.00	-0.00	0.01	0.01
scale of μ - true scale of μ	0.00	-0.00	0.05	0.05
β - true β	-0.00	0.00	0.04	0.04

Panel B: $(b,a,f,\sigma_\mu,\sigma_\varepsilon) = (2,1,3,2,1)$

	mean	median	stdev	rmse
α - true α	0.00	0.01	0.03	0.03
f - true f	0.00	-0.00	0.02	0.02
scale of μ - true scale of μ	0.00	0.00	0.04	0.04
β - true β	0.00	0.00	0.04	0.04

Note: Distribution of the gaps between true parameter and estimated parameters from 100 Monte Carlo experiments. We use the same parametric model – in equation (1) and (2) – in both the data-generating process and the estimation routine.

To further confirm that our results are not driven by specific functional form assumptions, we simulate choices using the same model but now estimate an auxiliary model where the consumption utility function is approximated by a polynomial. Specifically, in estimation, we specify the utility function as

$$c_{it}(p_t, q_{it}) = \sum_{\tau=1,\ldots,5} b_\tau q_{it}^\tau - \alpha p_t q_{it} + \mu_{it}(q_{it}) \tag{21}$$

where b_τ are coefficients of the polynomial specification and τ is up to the 5th order. Price disutility and consumption utility shock μ are specified in the same way as (1). Note that equation (21) is a mis-specified utility function for finite order τ.

The Monte Carlo results are summarized by Table 4. In the first case, we find that we can estimate most of the parameters without any bias, but produces small bias estimates in scale parameter μ and the first-order term in the consumption utility b_1. In the second case, we find that the model (with utility function approximated with a 5th order polynomial) recovers the true parameter precisely. We try to increase the order of the polynomial approximation and find that these bias go away but at the cost of power.

Appendix Table 4: Monte Carlo results with polynomial utility

Panel A: $\left(b, a, f, \sigma_{\mu}, \sigma_{\varepsilon}\right) = (3, 1, 6, 2, 1)$

	mean	median	stdev	rmse
α - true α	-0.02	-0.03	0.02	0.03
f - true f	-0.10	-0.08	0.18	0.20
scale of μ - true scale of μ	-0.13	-0.12	0.05	0.14
b(1) - true b(1)	0.35	0.33	0.10	0.36
b(2) - true b(2)	0.00	0.00	0.01	0.01
b(3) - true b(3)	0.01	0.01	0.01	0.01
b(4) - true b(4)	-0.00	-0.00	0.00	0.00
b(5) - true b(5)	-0.00	-0.00	0.00	0.00

Panel B: $\left(b, a, f, \sigma_{\mu}, \sigma_{\varepsilon}\right) = (2, 1, 3, 2, 1)$

	mean	median	stdev	rmse
α - true α	-0.01	-0.01	0.00	0.01
f - true f	0.00	0.00	0.00	0.00
scale of μ - true scale of μ	-0.01	-0.01	0.00	0.01
b(1) - true b(1)	0.01	0.01	0.00	0.01
b(2) - true b(2)	0.00	0.00	0.00	0.00
b(3) - true b(3)	0.00	0.00	0.00	0.00
b(4) - true b(4)	0.00	0.00	0.00	0.00
b(5) - true b(5)	-0.00	-0.00	0.00	0.00

Note: Monte carlo results similar to Table 4, except that consumption utility is modelled in a flexible polynomial form as in Equation (21), with τ set at the 5th order.

E Evidence from other categories

We examine whether shoppers in other categories exhibit similar behavior as shoppers of yogurt. Specifically, we examine their price response curves and the distribution of their purchase quantity at the price thresholds (i.e. maximum accepted prices), in the same way as we produce Figure 3 and 5. Figure 9 reproduces our descriptive evidence in the yogurt category for ground coffee (in coffee), potato chips (in salty snacks) and soup. The shopping behavior due to fixed costs extend to some other categories beyond yogurt.

F Additional tables and figures

Appendix Table 5: Variety and quantity per trip

	nr. products 1	2	3	4	5	6	Total
1 unit	17.03	0.00	0.00	0.00	0.00	0.00	17.03
2 units	14.87	1.40	0.00	0.00	0.00	0.00	16.27
3 units	9.85	1.72	0.11	0.00	0.00	0.00	11.68
4 units	11.82	2.22	0.19	0.01	0.00	0.00	14.24
5 units	7.59	1.96	0.27	0.02	0.00	0.00	9.85
6 units	8.68	2.15	0.29	0.03	0.00	0.00	11.15
7 units	1.85	1.14	0.23	0.04	0.00	0.00	3.26
8 units	2.23	1.02	0.21	0.03	0.01	0.00	3.50
9 units	0.86	0.62	0.17	0.03	0.00	0.00	1.68
10+ units	7.04	3.14	0.89	0.23	0.04	0.01	11.34
Total	81.81	15.37	2.36	0.39	0.05	0.01	100.00

Notes: This table reports percentage share of observations for a given variety-quantity combination.

Appendix Table 6: Discounts and promotion by pack-size

oz	discount	depth	feature	display
12	0.159	0.246	0.0930	0.0241
16	0.193	0.246	0.111	0.0238
32	0.0413	0.165	0.0152	0.000376
48	0.108	0.176	0.0473	0.00790
56	0.107	0.171	0.0629	0.00105

Notes: This table reports, by package size in oz, the share of discount (above 5%), feature, display, and the depth of discount if discounted.

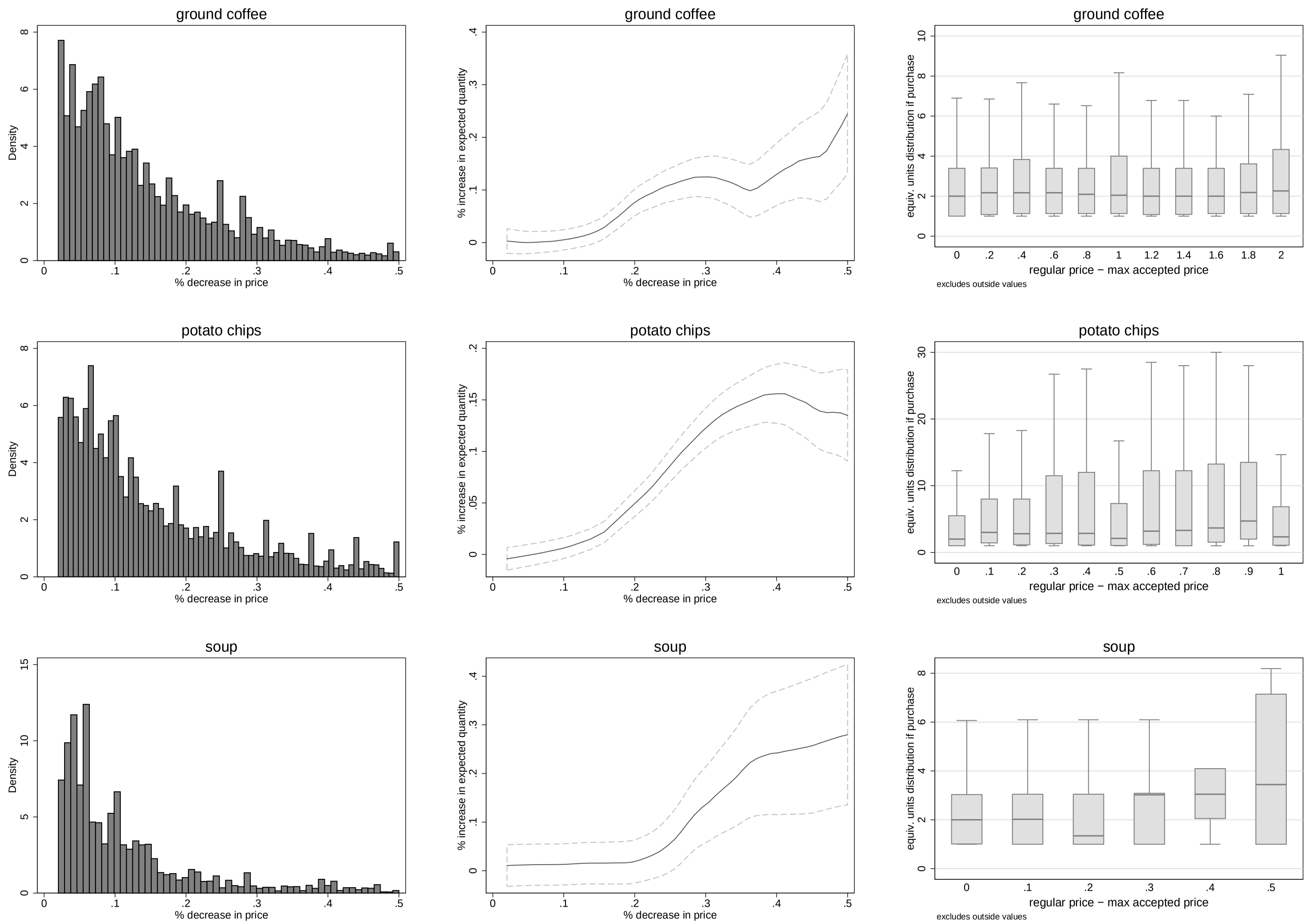

Figure 9: Discount, discount response and minimal purchase quantity at discount threshold

Notes: These figures show replications of Figure 3 and 5 for ground coffee, potato chips and soup. We also present discount distributions given 1) that discount exceeds 2% and 2) that the product is not on feature or display.

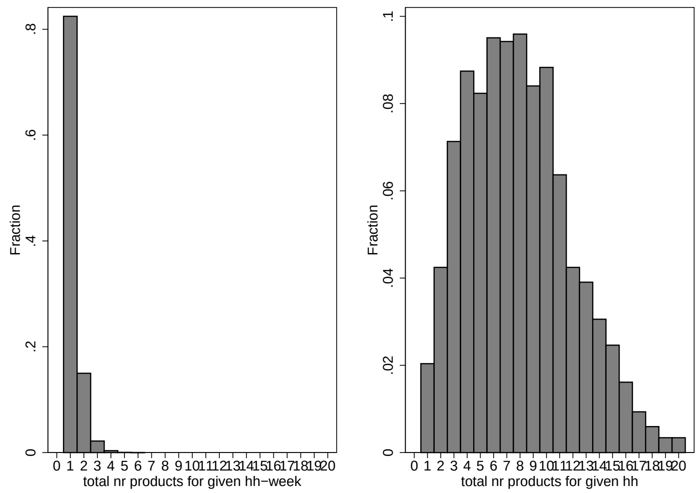

Figure 10: Per-trip and total variety for a given household

Notes: Left: distribution of total number of products purchased in a given trip, for households who had between 20 and 40 trips involving yogurt purchase in the sample. Right: distribution of total number of products purchased in the entire sample duration for the same set of households.

Appendix Table 7: Test for consumer stockpiling: product level

	Danone light	Danone light	Yoplait original	Yoplait original	Colombo classic	Colombo classic	Columbo light	Columbo light
months since last purchase	-0.01***		-0.01***		-0.00*		-0.01***	
	(0.00)		(0.00)		(0.00)		(0.00)	
– squared	0.00***		0.00***		0.00***		0.00***	
	(0.00)		(0.00)		(0.00)		(0.00)	
log price	-0.10***	-0.07***	-0.11***	-0.07***	-0.11***	-0.07***	-0.10***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
– past week		-0.02**		-0.01		0.01		0.00
		(0.01)		(0.01)		(0.01)		(0.01)
– past 2 weeks		0.00		0.00		0.01*		-0.01
		(0.01)		(0.01)		(0.01)		(0.01)
years since 200101	-0.02***	0.02***	-0.06***	-0.02***	-0.07***	-0.01**	-0.01***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Apr-Jun	0.02***	-0.00	-0.02***	-0.01**	-0.03***	-0.01	-0.02***	0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Jul-Sep	-0.00	-0.02**	-0.02***	-0.02***	-0.02***	-0.01	-0.01	0.01**
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
Oct-Dec	-0.01**	-0.02***	-0.00	0.00	-0.01***	-0.00	-0.02***	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
individual FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
obs.	56449	17398	62591	17296	37296	13145	34667	14028

Notes: Product level evidence for extensive margin only. Intensive margin suffers power issues due to limited sample size.

Appendix Table 8: Distribution of number of products in the sub-sample

	nr. prod. 0	1	2	3	4	Total
0 unit	31.31	0.00	0.00	0.00	0.00	31.31
1 unit	0.00	8.48	0.00	0.00	0.00	8.48
2 units	0.00	9.91	0.44	0.00	0.00	10.35
3 units	0.00	7.43	0.68	0.03	0.00	8.14
4 units	0.00	9.72	0.91	0.06	0.00	10.69
5 units	0.00	5.67	0.88	0.06	0.00	6.60
6 units	0.00	7.67	1.15	0.06	0.01	8.89
7 units	0.00	1.53	0.66	0.11	0.01	2.31
8 units	0.00	2.26	0.61	0.06	0.01	2.95
9 units	0.00	0.73	0.41	0.04	0.03	1.21
10+ units	0.00	6.67	1.93	0.43	0.06	9.08
Total	31.31	60.08	7.66	0.84	0.11	100.00

Note: The table presents the joint distribution of number of different purchased products, and the total number of units, in the sub-sample used for structural estimation.

Appendix Table 9: Implied elasticities

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
(A) Dannon Light N Fit	-2.38	0.12	0.14	0.24	0.24	0.20	0.15	0.14	0.25	0.14
(B) Yoplait Original	0.14	-2.77	0.22	0.21	0.36	0.21	0.47	0.11	0.17	0.16
(C) Colombo Classic	0.06	0.10	-2.94	0.34	0.04	0.10	0.09	0.19	0.00	0.19
(D) Colombo Light	0.10	0.08	0.31	-3.15	0.08	0.06	0.08	0.11	0.02	0.12
(E) Yoplait Light	0.14	0.18	0.04	0.11	-3.07	0.04	0.20	0.03	0.16	0.05
(F) Dannon Other	0.10	0.09	0.10	0.07	0.03	-2.62	0.11	0.12	0.02	0.10
(G) Yoplait Thick	0.09	0.24	0.10	0.11	0.21	0.13	-3.17	0.06	0.05	0.10
(H) Dannon Stonyfield	0.04	0.03	0.11	0.07	0.02	0.06	0.03	-2.67	0.00	0.15
(I) Wells Blue Bunny	0.05	0.03	0.00	0.01	0.06	0.01	0.02	0.00	-2.24	0.00
(J) Coolbrands Breyers	0.04	0.04	0.11	0.08	0.02	0.05	0.05	0.14	0.00	-2.69

Note: The i, j element is the elasticity of total purchase quantity of product j, on price change of product i, or $\frac{\partial Q_j}{\partial P_i} \cdot \frac{P_i}{Q_j}$.

Appendix Table 10: Implied semi-elasticities to feature

	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
(A) Dannon Light N Fit	0.26	-0.03	-0.03	-0.06	-0.06	-0.04	-0.04	-0.03	-0.07	-0.04
(B) Yoplait Original	-0.04	0.28	-0.05	-0.05	-0.08	-0.05	-0.10	-0.03	-0.04	-0.04
(C) Colombo Classic	-0.02	-0.02	0.32	-0.07	-0.01	-0.02	-0.02	-0.04	0.00	-0.05
(D) Colombo Light	-0.02	-0.02	-0.07	0.38	-0.02	-0.01	-0.02	-0.03	-0.01	-0.03
(E) Yoplait Light	-0.04	-0.04	-0.01	-0.03	0.36	-0.01	-0.05	-0.01	-0.04	-0.01
(F) Dannon Other	-0.02	-0.02	-0.02	-0.02	-0.01	0.25	-0.02	-0.02	-0.01	-0.02
(G) Yoplait Thick	-0.02	-0.05	-0.03	-0.03	-0.05	-0.03	0.38	-0.02	-0.02	-0.03
(H) Dannon Stonyfield	-0.01	-0.01	-0.03	-0.02	-0.01	-0.01	-0.01	0.29	0.00	-0.04
(I) Wells Blue Bunny	-0.01	-0.00	0.01	0.01	-0.02	0.00	-0.00	0.01	0.25	0.01
(J) Coolbrands Breyers	-0.01	-0.01	-0.03	-0.02	-0.01	-0.01	-0.01	-0.03	0.00	0.33

Note: The i, j element is the percentage change in quantity $\Delta Q_j / Q_j$ of the column product j, when the row product i is *on* or *off* feature. We hold prices and feature of other products as given by the data.

Appendix Table 11: Expected number of variety with lower consideration costs

	no feature: choice	consideration	all feature: choice	consideration
0 product	0.13	0.03	0.11	0.01
1 product	0.58	0.13	0.58	0.08
2 products	0.28	0.84	0.31	0.91

Notes: This table reports the distribution of number of products a consumer buys and considers, separately for the case where no product is on feature (Column 1 and 2) and all products are on feature (Column 3 and 4).

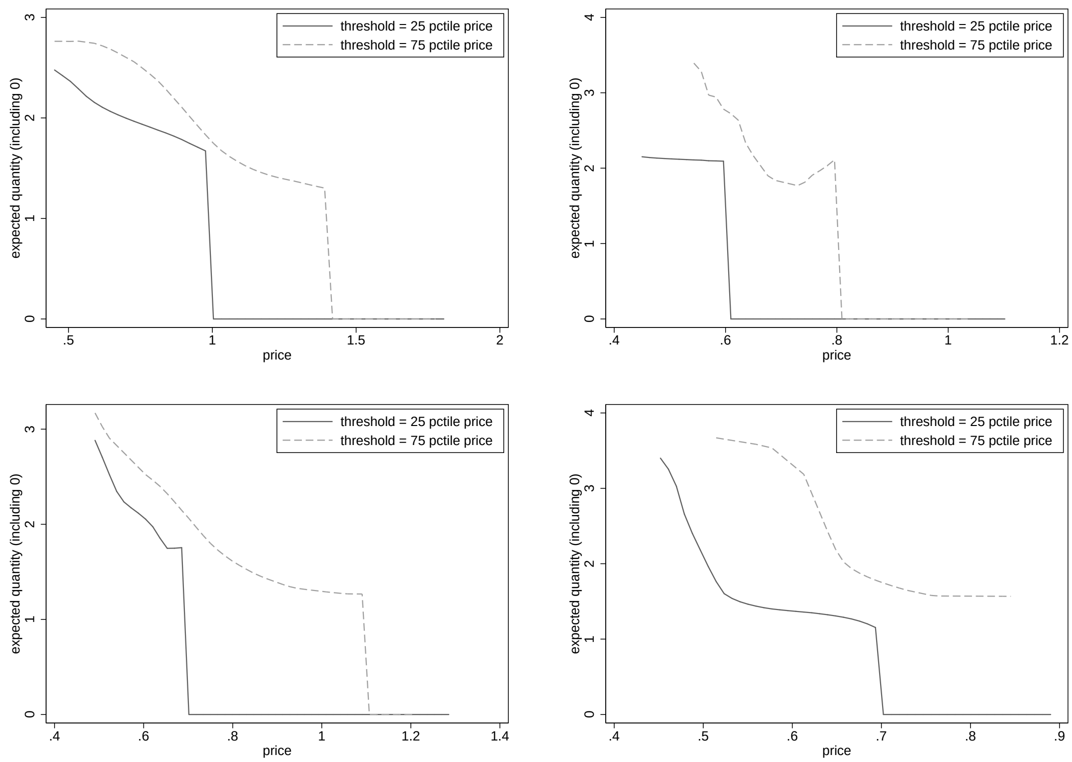

Figure 11: Individual demand by cohort: top 4 products

Notes: Demand by "cohort" defined as consumers whose observed price thresholds (max accepted price) are similar. See description for Figure 1. The top blocks are (left to right) Dannon Light and Yoplait Original, and the bottom two are Colombo Classic and Colombo Light.