Discount Threshold and the Cost of Consideration

Bart J. Bronnenberg and Yufeng Huang*

May 16, 2015

Preliminary. Please do not cite or post without contacting the authors.

Abstract

Why are consumers unresponsive to small price discounts? This paper proposes that it is costly to consider the purchase of each product, even at full information. As a result, a consumer will not consider a product (and thus will not purchase) until the price becomes favorable, at which point she will buy enough quantity that justifies the fixed cost of consideration. In this paper, we identify consideration cost through discontinuities in the purchase quantity at these "marginal" prices. We then construct and estimate a model of multiple product-quantity choice, explicitly modeling the choice of consideration set. We find that considering each Yogurt product costs \$3.2-\$5.2 in a trip, which is higher than the expenditure on the product. Such mental cost discourages allocating expenditure across many products, and a love-for-variety consumer would otherwise purchase 3 times as many varieties, if consideration is free. Finally, we decompose price elasticities and find that the effect on consideration set formation constitutes most of the price response.

*Tilburg School of Economics and Management, Tilburg University, Warandelaan 2, 5037 AB, Tilburg, The Netherlands. E-mail: bart.bronnenberg@tilburguniversity.edu. We thank Jaap Abbring, Marnik Dekimpe, Els Gijsbrechts, Tobias Klein and Thomas Otter for comments and suggestions. Bronnenberg and Huang thank the Netherlands Organization for Scientific Research (NWO Vici Grant) for financial support.

1 Motivation

As a widely-held belief among practitioners, consumers have a “promotion threshold”, and small price discounts do not persuade them into purchasing. As a result, sales do not respond to small discounts (Van Heerde et al., 2001). In Figure 1, we reproduce this pattern using consumer level data, conditional on the provision of price information (feature and display). It is puzzling that consumers show little to no response to small price discounts, but are very responsive to large ones; this implies a partly convex demand curve.[1]

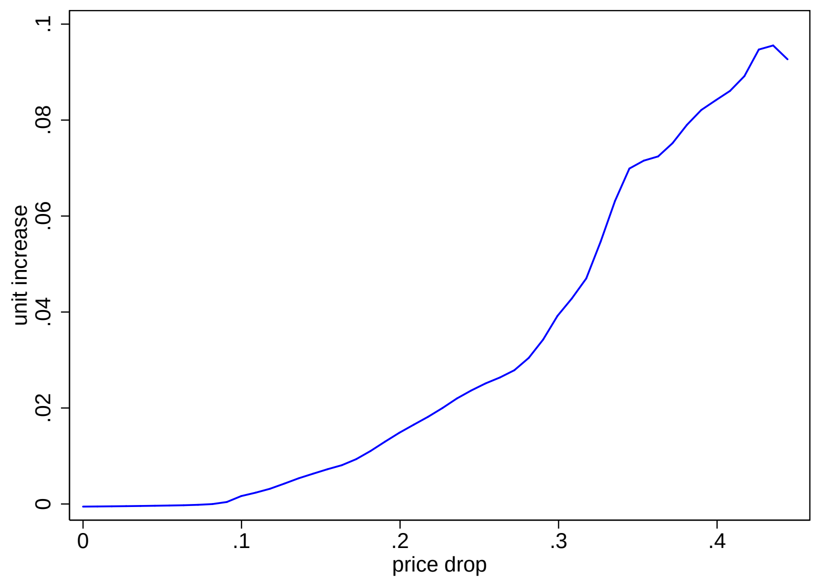

Figure 1: Convex response to discount

Notes: This figure shows the observed relationship between price and quantity, adjusted for consumer-product fixed effects. It is produced by consumer level scanner panel data in yogurt category, from Information Resource Inc. (IRI). Vertical axis: number of equivalent units, normalized by the consumer’s averaged purchase quantity under the regular price. Horizontal axis: discount from the regular price (in US dollar).

In this paper, we provide an explanation to a consumer’s unresponsiveness to small discounts. We propose that, even at full information, it is costly to think about the purchase of each product. This concept is related to Shugan (1980), who characterizes such thinking costs as costs of interpreting each product characteristics, and analytically derives the implied decision rule under these

[1]The 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).

costs. Think of yogurt for concreteness: an example of this mental cost is the effort in processing the nutrition content; compared to price, this characteristic is difficult for a regular consumer to understand. As a result, the consumer would *choose* not to consider a product, unless the low price and the associated quantity (i.e. total utility) justifies the fixed cost of consideration. Thus, if price of the product drops to a point where she decides to purchase, her purchase quantity will exhibit discontinuities around this point. The convex price response curve in Figure 1 is therefore a "smoothed out" version of such quantity jumps, across consumers with different price acceptance thresholds.

On the other hand, these quantity discontinuities also serve as identifying variations, for the researcher to distinguish between consideration and preference. As a classical question in demand estimation, when we observe no purchase to a product, it is difficult but important to distinguish a consumer's lack of consideration, from her low preference. Knowing how much (lack of) demand is due to (lack of) consideration is important, because it dictates how much information a firm should optimally provide. Yet the task is difficult given data on concentration of expenditure in subsets of products, since high concentration in one product might either imply consumers prefer it much more than others, or that consideration costs for the other products are too high. However, with price variations in panel data, expenditure concentration due to preference or due to costly consideration will generate different quantity choice patterns, when a "marginal" price discount is *just enough* to convince a consumer to start purchasing a product. If she previously does not choose the product because of low preference, then, at the slightly more favorable price, she will "try out" the product at a small quantity, due to decreasing marginal utility. However, if she previously does not buy because of high fixed cost, she will purchase a large amount at the new price, and take advantage of the economy of scale. In other words, discontinuities in quantity choice, at the price where a consumer just begins the purchase, is a unique prediction of a fixed acquisition cost – and this serves as identifying variation for limited consideration.

In the paper, we first present a simple model with a consumer making decisions of consideration

and quantity, holding full information about price. We analytically and numerically show that the price threshold depends on how much quantity a consumer *would* buy if consideration cost is expended. The model can then derive testable implications, in particular on how much consumer purchase quantity would change at this price threshold; that is, at the maximum price where we observe a purchase. Without imposing any structural models, we test for this using consumer scanner panel data in the yogurt category, from Information Resource Inc. (IRI). We find very large quantity jumps when the price just becomes acceptable for *each* consumer. The magnitude of the quantity jump is not explained by quantity discount (that is, unit price is non-increasing in the quantity one purchases) and indivisible quantity (i.e. there is a minimal quantity that is available), and therefore, strongly speaks for the existence of consumer consideration cost. Our reduced-form evidence can be replicated in some other categories, such as milk, coffee, and salty snacks.

Next, we structurally quantify the magnitude of consumer consideration cost, by estimating a model of product-quantity choice on the same scanner data. To accommodate our identification strategy, our model characterizes consumer choices over multiple products, as well as their quantity choice for each product they purchase. In particular, our model needs to accommodate quantity discontinuities, as it is key to identification. Consequently, standard models for multiple discrete choice (Hendel, 1999; Kim et al., 2002; Dubé, 2004) do not apply in this context, as they rely on the property that optimal quantity choice is everywhere continuous in price. Also, different from existing models of limited consideration – for example Goeree (2008), Van Nierop et al. (2010) and Dehmamy and Otter (2014) – consideration decision in our case is endogenous to the expected utility from purchase, which varies with price. Therefore, we explicitly model a two-stage decision problem, in which a consumer first chooses a subset of product to costly consider, and then chooses the quantity within this consideration set. Specifically, we model choices over bundles of products in the first stage, similar to Gentzkow (2007). In the second stage, the consumer naturally has a tendency to choose over multiple products, due to decreasing marginal utility for each product. And we characterize this by quantity choices of multiple products, conditional on

membership of a consideration set. Because consideration sets are unobserved by the researcher, we then integrate quantity choices over all potential consideration sets. Although this approach suffers from heavier computational burden (but in our case it is manageable), it allows for choices over multiple products, as well as quantity discontinuity at the price thresholds, both of which are key to our problem. In addition, we can also flexibly allow for nonlinear prices (quantity discounts) and discrete quantity sets (Allenby et al., 2004). Finally, we do not consider price search as it is non-central to our research question (and price search does not explain the Figure 1). To accommodate this simplification, we condition consideration and quantity choice on category purchase decisions, both in the model and in the data.[2]

We estimate structural parameters in the model, using IRI data in yogurt category, and find that in this context, consideration cost is between \$3.2 and \$5.2 *per product-trip*. This means that consideration cost is 1.2-1.9 times the magnitude of a consumer's per-trip expenditure on a product. As the first implication, such high cost means that the consumer cannot afford the effort of considering all products. Despite having love-for-variety preferences, we find that they would have purchased 3 times as many varieties on average, without costly consideration. This suggests large potential increase in welfare from policies that lower consideration cost.

Under costly consideration, an important role of price discount is to incentivize consumers to *think about* a product. As the second implication, we decompose price elasticities into an effect on inclusion of the consideration set, and another effect on quantity choice given consideration-set membership. We find that elasticities given consideration-set membership is about 1/3 of the overall price elasticities, which implies that (in this context) price discount is ineffective on quantity choice once a consumer starts to think about the product. On the contrary, price elasticities on consideration set formation is large, which implies that the primary role of price discount is to incentivize consideration. This means that prices are acting as important drivers of consumer heuristics and attention, and an important motive for planning price discounts should be to pen-

[2]Likely, the consumer has traveled to the focal product shelf. We take the argument in Seiler (2013) that these consumers make little effort in searching for price, compared to the consumers who did not buy products in the refrigerated section.

etrate the consideration cost barrier. In addition, higher consideration cost makes thinking about multiple products more *un*attractive, hence intensifies price competition. This implication is opposite to that of price-search cost, *a la* Diamond (1971).

Also, we allow for and estimate preference heterogeneity in the consumption utility (taste). As the third implication of this paper, consumers who purchase *despite* high consideration cost are those with high taste. When feature advertising reduces consumer consideration barrier, it attracts consumers who would otherwise *not* purchase – i.e., those with lower taste. Hence, feature advertising "downward-selects" customers with lower taste, who are also more elastic to price discounts. Therefore, setting a product on feature increases the price elasticities for conditional (on purchase) quantity choice, and this explains why price discounts are usually synchronous to feature advertising. In contrast, in conventional models that treat feature and display as persuasive, products under feature have lower elasticity and should be (without further complicating the model) complemented with a price increase.

Our primary contribution is that we relate the descriptive literature on promotion effect, to the literature on limited attention. On the one hand, 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), but do not provide deep explanation that relate to rational consumer decisions.[3] Our model of costly consideration contributes to the understanding of this empirical regularity, by providing a rational explanation to it. On the other hand, we provide an alternative identification strategy that separates lack of consideration from lack of preference. In particular, our identification strategy relies on observations of "price acceptance thresholds", i.e. the highest price at which the consumer would purchase positive quantities. Our usage of modern consumer level scanner data allows us to obtain dense empirical support of price, which enables the finding of these "price thresholds". We provide strong evidence of discontinuities in purchase quantity around the threshold. With this identification strategy, we also propose direct tests for costly consideration using standard marketing data, as well as a structural model that can be used to quantify these costs.

[3] Although Gupta and Cooper (1992) do relate this phenomenon to reference point theories in psychology.

In addition, substantively, our empirical results give insights into the effect of price discount, and marketing strategies that aim at overcoming consumers' consideration barrier. On the one hand, we decompose price elasticities into responses in consideration sets, and those in consumption decisions given consideration. We find the main effect from a price change on consideration set formation, which speaks for the practitioner intuition that prices alter consumer attention and heuristics, rather than their fully informed decisions. On the other hand, we study feature advertising as a tool to penetrate the consideration barrier, and find different interpretation of it compared to the literature, which mostly treat feature as persuasive (or a complementary consumption good *a la* Becker and Murphy, 1993).

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, which is simple enough to provide analytically as well as numerical solutions. This then generates key testable implications for the existence of costly consideration. We then test for this without any structural model. Then, Section 5 parametrizes the model and discusses estimation details. Section 6, 7 and 8 discusses, respectively, parameter estimates, implied prices elasticities and its decomposition, and implications about feature advertising. Finally, Section 9 concludes.

2 Related literature

We are primarily connected to three strands of literatures. First, we are related to the literature on promotion effect. Van Heerde et al. (2001) estimate a semi-parametric model using sales and price data, conditional on information arguments such as feature and display, and find that sales is unresponsive to small price change, and is most responsive to moderately large discounts. Their results are replicated in other contexts. We complement their work by providing individual level evidence. To this end, we adjust for consumer heterogeneity, product heterogeneity and store heterogeneity, and the interaction across these. With dense individual-level scanner data, we find clear evidence that an individual is close to unresponsive to price discounts until it reaches 20 cents. In addition,

we contribute to this literature also in that we provide an explanation that rationalizes this behavior; this 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 limited consideration literature. Among earlier works, Shugan (1980) provides psychological justifications or analytical framework on the implication of consumer thinking cost. Modern structural analysis empirically tease apart lack of attention (or high thinking cost) versus lack of preference, in several ways. One way is to obtain 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 choice. For example, Goeree (2008) assumes that advertising is informative, and advertising expenditure acts as exclusion restriction in the utility function (given consideration); Kawaguchi et al. (2014) proposes that product (un)availability is a good exclusion restriction to test for (in)attention.

The most closely related work to us is Dehmamy and Otter (2014). Their paper utilizes the sunk cost property of consideration in a consumer's decision of purchase quantity: fixed cost does not enter quantity choice because it is sunk. Therefore, they propose that one can test for exclusion restrictions, by testing whether they affect quantity choice conditional on purchase. In their experimental application, they provide evidence that number of shelf facing and location on the shelf are only affect consideration, and therefore are good exclusion restrictions. However, their methodology requires data on shelf facing and location, which is not always available. In our framework, we highlight the discontinuity in (continuous) quantity choice due to fixed consideration cost, and endogenous consideration decisions to prices. Our model implies that prices (and potentially other product characteristics) affect these decisions, and generates 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.

Finally, our work is related to the literature on multiple discrete choice. Hendel (1999); Kim

et al. (2002); Dubé (2004) model a consumer's product and quantity choice, and make simplifying assumptions to isolate the choice problems of different products. While this approach eases computation burden, from 2^J to J potential options (J is the number of products), the isolation of quantity decisions across products simplifies away their intense competition for consideration set membership. In our paper, we model consideration as a separate decision stage, and assume that the consumer is rational in that she takes into 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, should a research question requires that quantity decisions cannot be separately modeled. Our model is flexible enough to allow for nonlinear prices (quantity discounts) and discrete quantities (Allenby et al., 2004).

3 Data

3.1 Construction

We use the Behavioral Scan panel data from Information Resource Inc. (IRI) Academic Data Set (Bronnenberg et al., 2008), in the years 2001 to 2003.[4] We focus on the yogurt and yogurt drink categories. A "store visit" is recorded when a household purchases yogurt or yogurt drink in a specific trip – and we assume that the household has traveled to the product shelf of interest. 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 – importantly package size.

At the SKU level, price is defined as the outlet level revenue divided by the outlet level units sold. For price changes (discounts), we define it as the percentage change relative to the maximum price in the past 4 weeks – the underlying rationale being that there are many weekly discounts

[4]All our results, structural and reduced form, are robust when we add data in year 2004-2007.

where prices drop in one week but resume in the neighboring ones.[5] The data-set also records whether the product is on feature advertising or in-store display, or both.

Next, we aggregate the SKU level data. We define a "product" as one with a specific name recorded by the data, regardless of the flavor or package size. For example, "General Mills Columbo Light" is considered as a product, where "General Mills Columbo Light in berry flavor in 8 oz" is a distinct SKU.[6]

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, 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 4 units of 10 oz, is considered to have purchased *6 equivalent units*. Since few consumers bought non-integer equivalent units,[7] we can characterize quantity choice as discrete (in integer units), and price as a step function of quantity. This captures large-quantity discounts which is frequently observed in this product category.

Finally, we average prices, feature advertising and in-store display into product-quantity level. To do so, we take purchase quantity-weighted average of each variable, among all SKUs for a given product, in a given store and week. For binary variables such as feature and display, we record the associated probability on the product level.

[5]One needs to assume that prices do not temporarily increase beyond its regular level. We find some cases with price increase, yet such events occur much less frequently compared to price decrease.

[6]As a robustness check, we alternatively defined a "product" as a product name - flavor category combination, and the essential qualitative reduced-form evidence remain robust.

[7]In the full sample, 11% of all positive quantity choices involve non-integer equivalent units; yet, only 5% of these rounding errors are within 0.2 in absolute value. Therefore, even though rounding might introduce some errors, only a small portion of the sample is affected.

Table 1: Demographics

	Mean	Median	StDev
family size	2.5	2.0	1.3
age of household head	53.7	60.0	15.7
household annual income	45114.3	40000.0	27804.1
obs.	6558	6558	6558

Notes: this table reports household-level summary statistics in the year 2003. Annual income is nominal.

3.2 Summary statistics

3.2.1 Demographics

There are 8,397 households in the sampling period. Taking a cross-section in 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 53.7 years, and an annual income of $44,114. Household characteristics in other years are very similar.

3.2.2 Trips

Table 2 summarizes the duration for each consumer-retailer combination to be in the sample periods, and the number of weeks within that duration when we observe purchase. Overall, a household is present in a duration of 56.4 weeks between the first and the last trip to the same retailer (with purchase of yogurts), within which 10.4 weeks are associated with yogurt purchases. Within those 10.4 weeks, 5.7 weeks are associated with purchase of the top 10 products, and 2.5 weeks are with the top 3 products.

We drop weeks without yogurt purchase, for a given household and store, and only focus on the product choice conditional on buying at least one product within the category. The purpose of doing so is to condition on price awareness – essentially, we assume that the household has at least "glanced through" all the price tags on the shelf, and hence the price information is free. Thereby, we avoid having to model a costly search process on top of our model.

Table 2: Trips

	Mean	Median	StDev
duration (weeks)	56.4	47.0	53.3
purchase incidence	10.4	4.0	18.2
incidence (top 10 products)	5.7	2.0	11.6
incidence (top 4 products)	3.2	1.0	8.0
obs.	19761	19761	19761

Notes: this table reports summary statistics on the number of trips, per household-retailer, conditional on purchase (of a certain range of products).

3.2.3 Products, prices and concentration

In the sampling periods, there are 84 distinct products. We compute in-sample market share, based on the shares of total consumer expenditure in yogurt, and find that concentration is only moderate: overall, the top 4 products take up 34% of the market, while the top 10 occupy 61%. Among the top 10 products, average per-volume price is 0.14 dollar/oz, with a standard deviation of 0.07.

3.2.4 Variety and quantity

In sharp contrast to market concentration, we find that within-trip expenditure is heavily concentrated. Among all shopping trips (household-week-store combination), consumers do purchase considerable quantities: 74% are involved with more than one (equivalent) unit of purchase, and 23% more than 5 units. However, consumers are not willing to spread the expenditure across different varieties: 97% purchases *one or two* products.

3.2.5 Discounts, feature and display

Feature and display are rare in the yogurt category. On the product-store-week level, we find that there are only 7.1% of all observations with *more than half* of the SKUs for a given product are on feature.[8] Displays are even fewer – 1.5% are on display.

Conditional on whether a product is on feature or display, we can then plot the distribution

[8] To be precise, the percentage of feature is calculated as a sales volume weighted average.

Table 3: Variety and quantity

	nr. prod. 1	2	3	4	5	6	Total
1 unit	24.46	0.00	0.00	0.00	0.00	0.00	24.46
2 units	14.72	2.72	0.00	0.00	0.00	0.00	17.44
3 units	8.27	2.12	0.29	0.00	0.00	0.00	10.68
4 units	9.74	2.21	0.32	0.03	0.00	0.00	12.31
5 units	6.12	1.97	0.36	0.04	0.00	0.00	8.50
6 units	6.81	1.96	0.36	0.05	0.01	0.00	9.18
7 units	1.48	1.23	0.31	0.07	0.01	0.00	3.10
8 units	1.83	0.93	0.30	0.05	0.01	0.00	3.12
9 units	0.72	0.59	0.22	0.05	0.01	0.00	1.59
10+ units	5.59	2.73	0.94	0.29	0.06	0.02	9.64
Total	79.73	16.46	3.09	0.60	0.10	0.02	100.00

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

of price changes, shown in Figure 2. Discounts are frequently aligned with feature or display: 80% of the products on feature or display are on a discount with no less than 5% price drop. On the contrary, when there is no feature or display, only 8% of the products are on a 5%-or-more discount.

Further, conditional on being on feature or display, the discount distribution has a mode of 40%. This suggests an unwillingness to set shallow discounts. This is in line with practitioner's belief that consumer response to discount is convex.

4 Model setup and testable implications

4.1 Overview

In this section, we outline the general model without specifying the functional form. Then, to illustrate the key idea, we provide a numerical example using a simplified version of this model, which clearly provides testable implications. We then provide supporting evidence, directly from the data.

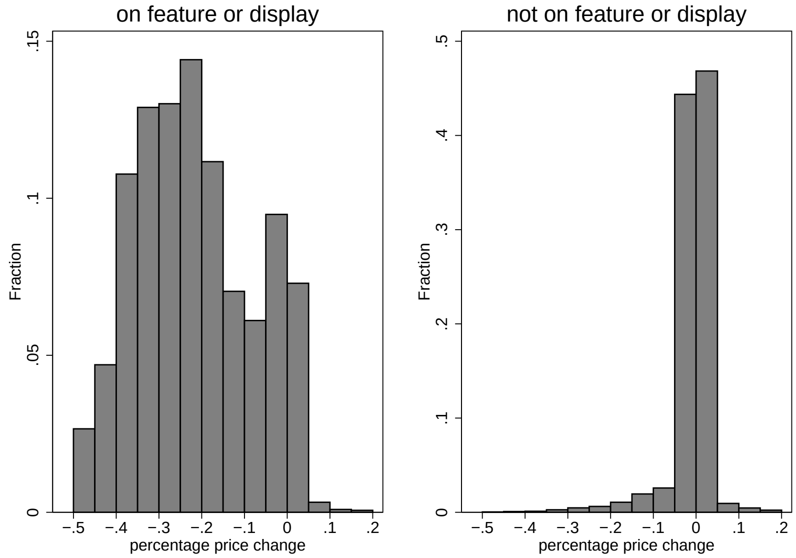

Figure 2: Distribution of price changes

Notes: this figure plots the distribution of percentage price changes. Base price is defined as the maximum price in the past 4 weeks. Feature and display are defined as at least *half* the SKUs for a given product are on feature or display, respectively.

4.2 General setup

An individual i, determined to purchase at least one unit of yogurt, travels to the refrigerated product shelf in trip t, where J different yogurt varieties, and an outside option ("other varieties"), are present. The individual knows the price of each product, but needs to incur some consideration cost, to include *each* of the J products into her consideration set $K \subset J$.[9] She then chooses quantity q_{ijt} for each product j in the consideration set.

Denote the $(J \times 1)$ vector of purchase quantity $\mathbf{q}_{it}$. The consumer's full utility can be specified in the following form:

$$u_{it}\left(\mathbf{q}_{it}, K_{it}\right) = \underbrace{c_{it}\left(\mathbf{q}_{it}\right)}_{\text{consumption utility}} - \underbrace{F_{it}\left(K_{it}\right)}_{\text{fixed cost}} + \underbrace{\lambda_i m_{it}}_{\text{money}} \tag{1}$$

[9] Depending on the context, J denotes either the number of products or the full set of products.

where $c_{it}(.)$ is a random consumption utility function, which takes as arguments purchase quantity $\mathbf{q}_{it}$, as well as consumption utility shocks from trip t, $F_{it}(.)$ is a random fixed cost function, which is dependent on consideration set K_{it} as well as t-specific shocks. Finally, m_{it} is the money left from the trip, that can be spent on the outside options, which brings utility $\lambda_i m_{it}$.

The consumer maximizes her utility (1) subject to the budget constraint

$$\sum_{j \in K} p_{jt}\left(q_{ijt}\right) \cdot q_{ijt} \leq m_{it}, \tag{2}$$

where we do not impose constant unit price, but rather specify the unit prices $p_{jt}\left(q_{ijt}\right)$ as functions of quantity. The budget constraint then incorporates prices into the consumer maximization problem.

4.3 An illustrative model

When the consumption utility is continuous and the fixed cost is strictly positive, purchasing a very small (i.e. close to zero) amount will not justify the effort spent in evaluating the product. For the consumer, she needs to purchase a minimum quantity to justify the consideration cost. This then implies that, when price of a product drops to the point when a consumer finds it optimal to *start purchasing* it – we call this price the “threshold price” – she always purchases above a minimum quantity threshold. Therefore, when the fixed cost is positive, the observed quantity choice of a given consumer is strictly positive (away from zero) at the threshold price, but (by construction) zero when the price is above the threshold price. This is to say, quantity jump at the threshold price uniquely identifies consumer fixed cost.

To illustrate this, we greatly simplify the model in Section 4.2. Despite that this simplification brings the model far from realistic, it gives an explicit solution to the individual demand function, which serves as a good concrete example. Note that the simplification in this section does not correspond to our empirical exercise.

Suppose that there is only one product of interest. In the utility specification (1), let the con-

sumption utility be a quadratic function: $c_{it}(q_{it}) = q_{it} - \frac{\beta_i}{2} q_{it}^2$, and a random fixed cost incurs whenever the consumer considers the product: $F_{it}(K_{it}) = f_{it} \cdot (||K|| = 1)$. Also, we assume that the unit price is constant in quantity; and we normalize the marginal utility to money, λ_i, to 1. Finally, note that we have assumed away randomness in the consumption utility $v_{it}(.)$, so that the consumer knows the exact quantity she will purchase. The consumer problem then reduces to

$$\max_{q_{it}} q_{it} - \frac{\beta_i}{2} q_{it}^2 - p_t \cdot q_{it} - f_{it} \cdot \mathbf{1}(q_i > 0).$$

Note that, in this case, a non-empty consideration set is equivalent to purchasing positive quantity.

The consumer solves the problem backwards: she first determines the optimal quantity subject to consideration, and then chooses whether it is optimal to consider the product. Her optimal quantity given consideration is determined by the first order condition:

$$1 - \beta_i q_{it} - p_t = 0$$

which gives $q_{it}^{foc} = \beta_i^{-1}(1 - p_t)$.

Her consideration decision then involves comparing utility from consuming q_i^{foc} and zero. This is to say, $q_i^* > 0$ if $u_i\left(q_i^{foc}\right) > 0$, which yields

$$q_{it}^* = \begin{cases} \beta_i^{-1}(1 - p_t) & \text{if } \beta_i < \frac{1 - p_t^2}{2 f_{it}} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

4.4 Detecting demand jumps at the threshold price

Note that in the above example, the "threshold price" is determined by the price at $\beta_i = \frac{1 - \bar{p}_{it}^2}{2 f_{it}}$, or $\bar{p}_{it} = \sqrt{1 - 2\beta_i f_{it}}$. Since this is the highest price that this consumer will accept, the lowest quantity

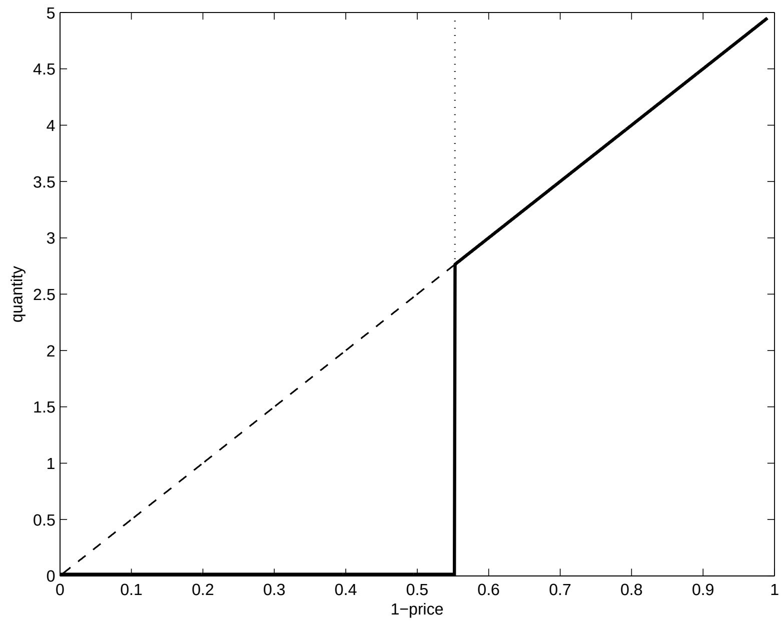

Figure 3: Quantity choice as a function of price

Notes: The solid curve illustrates the demand schedule implied by the simple toy model. The dotted line is the "threshold price" $\sqrt{1-2\beta_i f_{it}}$ (she is indifferent between purchasing or not), and the dashed line is the optimal quantity given purchase, q_{it}^{foc}, in the cases when she did not choose to purchase. Specifically, $F = 2$ and $\beta_i = 0.2$.

she will ever purchase (other than zero) is

$$\bar{q}_{it} = q_{it}^{*}\left(\bar{p}_{it}\right) = \frac{1-\sqrt{1-2\beta_i f_{it}}}{\beta_i}.$$

That is, given $f_{it} \gg 0$, when price crosses the threshold, the observed purchase quantity will jump between zero and $\bar{q}_{it} \gg 0$. In other words, the researcher can observe a quantity jump even when consumption utility itself is continuous. Figure 3 graphically illustrates this.

If the researcher knows the threshold price $\bar{p}_{it}$ for each consumer-trip, she can then test whether the consumer fixed cost f_{it} is positive, by testing whether the quantity at a price *slightly* below $\bar{p}_{it}$ – i.e. the threshold quantity $\bar{q}_{it}$ – is significantly different from zero.

We construct the threshold price $\bar{p}_{it}$ as the maximum price that we observe individual i making a purchase, *around* period t. Specifically, we focus on 4 weeks before the focal period t, and divide price reductions relative to the 4-week maximum price into 5-cent grids. However, because

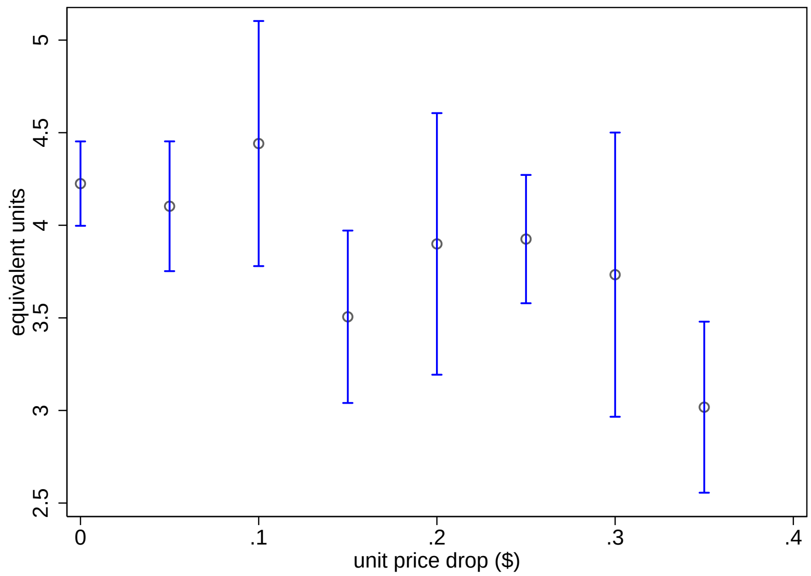

Figure 4: Quantity response to price for the marginal consumer

Notes: X-axis is the price reduction (the difference between regular price and current price), at which the consumer would *not* purchase the product, but would purchase at a slightly larger discount. Y-axis is the total number of equivalent units the consumer purchases for the given product.

yogurt units are indivisible, given that the consumer purchases at $\bar{p}_{it}$, her quantity choice should by construction be no smaller than 1. Therefore, we measure quantity in multiples of the minimum *available* package size, so that quantity 1 is always feasible for the consumer. Then, we instead test whether quantity choice at the threshold price is greater than 1.

This results in Figure 4, which shows that the average purchase quantity at the threshold price, given that the consumer purchases, is between 3 and 4 times the minimal *available* package size. This shows that a consumer will purchase a non-negligible amount even at marginal changes in price, which implies some discontinuity in the demand function. Under our model specification, this jump in quantity identifies the magnitude of consumer fixed cost.

In addition, for the consumers who do not respond to sizable price reductions, their purchase quantity is slightly lower. This reflects heterogeneity in consumer tastes, which causes the marginal consumers at a lower price threshold to purchase lower quantity.

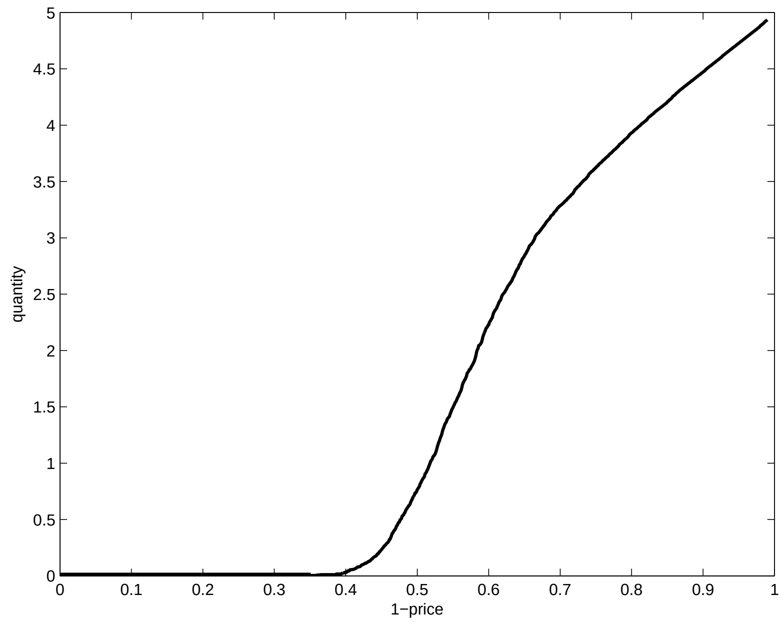

Figure 5: Averaged quantity response

Notes: This figure is generated by averaging across different curves in Figure 3, but with different β_i. Specifically, we assume $\log(\beta_i)$ to be Normal with mean $\log(0.2)$ and variance 0.1. Other parameters follow the previous figure.

4.5 Aggregate price response across consumers

When the researcher does not observe $\bar{p}_{it}$, quantity response discontinuity is smoothed away because of heterogeneity in the threshold price. However, since the threshold price follows a distribution generated by preference heterogeneity, the market average demand function is the average of individual demand, such as the ones defined in (3), but with heterogeneity in β_i. We numerically generate one possible average demand schedule in Figure 5. Note that the average quantity response to price change displays an S-shape, in which the steepest part reflects aggregation across different thresholds.

We replicate this shape using our data. Specifically, conditional on that the consumer faces a discount, we look at the *deviation* from her average purchase quantity among her trips in the same store, among the past four weeks when there is no discount.[10] As shown in Figure 1 at the beginning of this paper, the result is an S-shaped average quantity response function, which indicates lack of

[10] We focus on deviation to take out individual and product fixed effects.

response in demand, to price reductions that is less than 20 cents. This provides further support to the model without having to assume the location of price thresholds. Related to our evidence, Van Heerde et al. (2001) finds similar evidence on the aggregate sales data.

4.6 Downward-selection of marginal consumers

Finally, with heterogeneity in β_i, consumers who *only* purchase under lower fixed cost – presumably during feature advertising – are consumers with lower consumption utility. Therefore, a further implication by the model is that purchase quantities for consumers who only purchase under feature is lower than those of the regular customers.

We test for this implication by plotting the observed density function of purchase quantity, given that it is positive, and conditional on whether the product is on feature or not. Shown in Figure 6, for products sold on feature, the mode quantity choice for consumers *who purchase* is 3 equivalent units, compared to the 4 units for products not on feature. This suggests that, despite selling more in total, products on feature sell less *per consumer*, reflecting selection of the group with lower consumption utility than a regular customer, who would purchase less.

5 Full structural model and implementation

5.1 Overview

Recall that in the general model in Section 4.2, the consumer i maximizes her utility subject to the budget constraint in 2:

$$u_{it}(\mathbf{q}_{it}, K_{it}) = \underbrace{c_{it}(\mathbf{q}_{it})}_{\text{consumption utility}} - \underbrace{F_{it}(K_{it})}_{\text{fixed cost}} + \underbrace{\lambda_i m_{it}}_{\text{money}},$$

which is defined on a vector of quantity $\mathbf{q}_{it}$, and fixed cost as a function of consideration set $K_{it} \subset J$. In our empirical implementation, we limit dimensionality by restricting the total number of

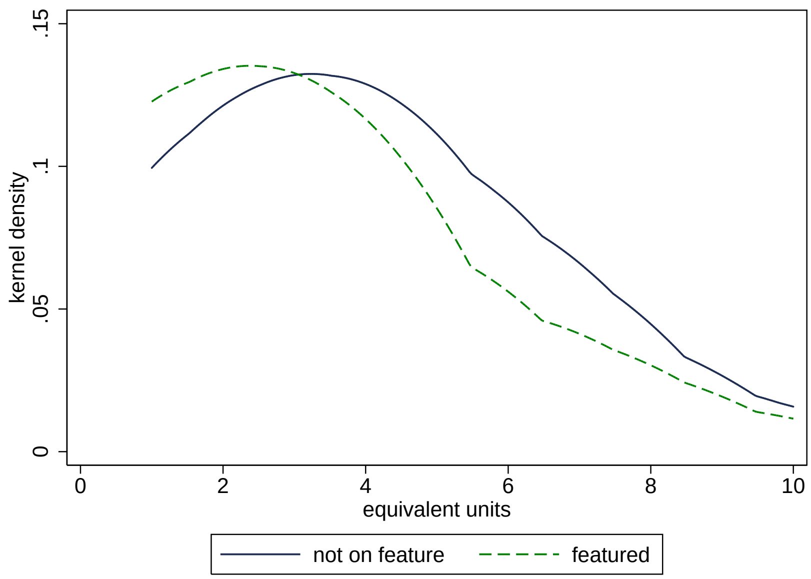

Figure 6: Conditional quantity distributions: featured or not

Notes: Kernel density plots of observed quantity distributions for products sold not on feature, and for products sold on feature (dashed). This is the distribution across all consumers, stores, dates and products. Bandwidth is set to 2 for smoothness. the only difference is that in this figure, we focus on the consumers who have never purchased without feature before.

products in the consideration set to be at most 2, i.e. $||K|| \leq 2$. In fact, as shown in Table 4, there are only 18 out of more than 16,000 observations, where an individual violates this restriction and purchased more than 3. Hence, imposing the consideration set size at 2 is not far from being realistic.

With this restriction, we can specify consumption utility and consideration cost, both as functions of membership and quantities of two products. This section provides details on parametrization of each part of the model, solution of the optimal choice rules, and implementation in estimation. Finally, this section also documents our choice of a sub-sample in estimation, to alleviate computation burden that would be prohibitive in the full sample.

5.2 Parametrization

5.2.1 Consumption utility

Denote the *purchase set* as $L_{it} = \left\{ j | q_{ijt} > 0 \right\}$. L_{it} is a subset of K_{it}, which is in turn a subset of all products J. Because we restrict the size of consideration set, we can denote $L_{it} = \{k, l\}$.[11]

We specify the (random) consumption utility as

$$c_{it}\left(\mathbf{q}_{it}\right) = \sum_{j \in L_{it}} \tilde{\beta}_{ij} \log\left(q_{ijt} + 1\right) + \gamma_i \prod_{j \in L_{it}} \log\left(q_{ijt} + 1\right) \cdot \mathbf{1}\left(||L_{it}|| = 2\right) + \mu_{it}\left(q_{ikt}, q_{ilt}\right)$$

where the consumption sub-utility is specified in log, and defined on *discrete* quantities $q_{ijt} \in \{0, 1, ..., \bar{q}\}$.[12] Coefficients $\tilde{\beta}_{ij}$ and γ_i capture the shape of marginal utility in consumption: the $\tilde{\beta}_{ij}$'s capture marginal utility to *percentage* increase in quantity for each product j, while γ_i capture the interaction between percentage increases in quantities between two products in the choice set. When the choice set is singleton, the interaction effect is set to zero.

[11] For notational simplicity, we use the same notation to denote singleton sets – in which case one can think of product l is "0", and the singleton set can be denoted as $L_{it} = \{k\} = \{k, 0\}$.

[12] The consumption utility function takes the form similar to Kim et al. (2002) and Dehmamy and Otter (2014), but imposes curvature of the consumption utility. Yet, it allows for consumption utility to interact between products. One other benefit of using the log specification, is that the interpretation each unit increase in the sub-utility is clear.

Also, the consumer derives random consumption utility shocks, $\mu_{it}(q_{ikt}, q_{ilt})$, unobserved by the researcher. One could consider $\mu_{it}(q_{ikt}, q_{ilt})$ as information (or utility shock) that favors, or opposes, purchasing a specific quantity combination. We assume, for a particular combination of quantity (q_1, q_2), $\mu_{it}(q_1, q_2)/\kappa$ is type-1 extreme value distributed, , where κ is a scale coefficient.

Finally, we allow correlation in individual preferences across brands, by imposing $\tilde{\beta}_{ij}$ as a function of observed product characteristics:

$$\tilde{\beta}_{ij} = x_j^{'}\beta_i$$

where β_i is a vector of random coefficients on characteristics, such as brand dummies and low sugar ("light"). Although each dimension of β_i is independent of each other, this introduces dependence within an individual, in $\tilde{\beta}_{ij}$ across product j's.

5.2.2 Budget constraint

m_{it} characterizes the attractiveness of the outside option, or "money". Prices affect decisions via a budget constraint, characterized in Equation 2. Note that we allow the per-product expenditure $p_{jt}(q_{ijt}) \cdot q_{ijt}$ to be non-linear in quantity, to capture the potential quantity discounts that consumers could benefit from, by buying large quantities. Operationally, $p_{jt}(q_{ijt})$ is the *lowest*, per-unit price one could get when choosing quantity q_{ijt}. λ_i is the price coefficient when substituting the budget constraint to the direct utility function.

5.2.3 Fixed cost

The consumer also "pays for" the consideration cost, $F_{it}(K_{it})$, as a function of her consideration set. We parametrize it as

$$F_i(K_{it}, \mathbf{A}_{it}) = \sum_{j \in K_{it}} \left(f_{ij} + \Delta f_A \cdot \mathbf{1}\left(A_{ijt} = 1\right)\right) + \Delta f_2 \cdot \mathbf{1}\left(||K_{it}|| = 2\right) - \varepsilon_{iKt}$$

where $\mathbf{A}_{it}$ is a vector that indicates whether each product is on feature advertising, f_{ij} denotes the baseline consideration cost for each product j for individual i, Δf_A is the increase (negative means reduction) in fixed cost when a product is on feature, and Δf_2 is the additional total consideration cost when considering two products.

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

5.3 Solution of optimal choice rules

5.3.1 Second stage decisions

Both the researcher and the consumer solve the decision process backward. In the second stage, conditional on the consideration set K_{it}, the individual maximizes utility given consumption utility shock $\mu_{it}(.)$, and chooses the quantity combination. In other words, she chooses (q_{ikt}, q_{ilt}) given $K_{it} = \{k,l\}$ ($l = 0$ in case of a single-product consideration set), to maximize utility (1) subject to the budget constraint (2).

Substitute (2) into (1), and we have the indirect utility at the second stage:

$$\begin{aligned} w_{it}\left(q_{ijt}, q_{ikt}; \mathbf{A}_{it}\right) &= \sum_{j=k,l} \left(\mathbf{x}_j^{'}\beta_i \cdot \log\left(q_{ijt}+1\right) - \lambda_i \cdot p_{jt}\left(q_{ijt}\right) \cdot q_{ijt}\right) \\ &\quad + \gamma_i \prod_{j=k,l} \log\left(q_{ijt}+1\right) \cdot \mathbf{1}\left(||L_{it}||=2\right) + \mu_{it}\left(q_{ikt}, q_{ilt}\right) - F\left(\{k,l\}, \mathbf{A}_{it}\right) + \varepsilon_{iKt}. \end{aligned}$$

Denote $\bar{w}_{it}(q_{ikt}, q_{ilt}) = \sum_{j=k,l}\left(\mathbf{x}_j^{'}\beta_i \cdot \log\left(q_{ijt}+1\right) - \lambda_i \cdot \log\left(q_{ijt}+1\right) \cdot q_{ijt}\right) + \gamma_i \prod_{j=k,l} \log\left(q_{ijt}+1\right) \cdot \mathbf{1}\left(||L_{it}||=2\right)$, and because the consideration cost with shock ε are irrelevant for the second stage decisions, we have the standard Logit choice probability

$$\Pr\left(q_{ijt}, q_{ikt} | K_{it}; \theta_i\right) = \frac{\exp\left(\bar{w}_{it}\left(q_{ikt}, q_{ilt}\right)/\kappa\right)}{\sum_{\left(q_k^{'}, q_l^{'}\right)\in Q^2} \exp\left(\bar{w}_{it}\left(q_k^{'}, q_l^{'}\right)/\kappa\right)},$$

Where θ_i denotes all relevant parameters. Note that the possible choice set combinations $Q^2 = \{0,1,...,\bar{q}\} \times \{0,1,...,\bar{q}\}$ includes buying *nothing*, or buying from only one product.[13]

5.3.2 First stage decision

From the second stage, the inclusive value – expected maximum utility – given the first stage consideration set decision, can be obtained as

$$\bar{v}_{it}(K_{it}, \mathbf{A}_{it}) = \Gamma + \kappa \cdot \log \left(\sum_{\left(q_k^{'}, q_l^{'}\right) \in Q^2} \exp \left(\frac{\bar{w}_{it}\left(q_k^{'}, q_l^{'}\right) - F\left(K_{it}, \mathbf{A}_{it}\right)}{\kappa} \right) \right),$$

where Γ is the Euler constant. Then, the first stage decision maximizes the indirect utility

$$v_{it}(K) = \bar{v}_{it}(K) + \varepsilon_{iKt},$$

which yields the choice probability

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

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 of purchase given K. To match the data, note that what is observed are the choice probabilities of a specific

[13]For singleton consideration sets, the quantity support reduces to Q.

quantity combination, or, the empirical counterparts of

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

5.4.2 Likelihood with random coefficients

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

$$\mathscr{L}\left(\theta\right)=\prod_i\left(\int_{\theta_i}\left(\prod_t\Pr\left(q_{ikt},q_{ilt};\theta_i\right)\right)dG\left(\theta_i;\theta\right)\right),$$

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

The integral on θ_i is computed by simulation. To implement the simulated maximum likelihood method, we first take M draws of random coefficients *shocks* on β_{ih} and f_{ij},[14] denoted $\hat{\beta}_{mh}$ and $\hat{f}_{mj}$ for draw m, each independently from $\mathscr{N}\left(0,I_{J+H}\right)$. Then, given parameters $\bar{\beta}_h$ and $\bar{f}_j$ and σ_β and σ_F, the individual i's random coefficient in draw m are determined by $\beta_{ihm}=\bar{\beta}_h+\sigma_\beta\cdot\hat{\beta}_{ihm}$, $f_{ijm}=\bar{f}_j+\sigma_F\cdot\hat{f}_{ijm}$; hence the random consumption coefficients on the product level are

$$\tilde{\beta}_{ijm}=\sum_{h=1,\ldots,H}\left(\bar{\beta}_h+\sigma_\beta\cdot\hat{\beta}_{ihm}\right)\cdot\mathbf{1}\left(x_j=1\right)$$

We restrict γ and λ to be homogeneous across individuals.

We then maximize the likelihood function with respect to $\bar{\beta}_{mj}$, $\bar{F}_{mj}$, σ_β and σ_F, and other parameters, taking the draws as given. For each parameter value, the empirical counterpart of the

[14] where β_{ih} is the h'th dimension of $\mathbf{x}_j$.

likelihood function is given by

$$\mathscr{L}_N(\theta) = \prod_{i=1}^{N} \left(\frac{1}{M} \sum_{m=1}^{M} \left(\prod_t l_{it}(q_{ikt}, q_{ilt}; \theta_{im}) \right) \right).$$

5.5 Sub-sample

5.5.1 Choice of the sub-sample

To restrict computation burden at a reasonable level, we implement the structural model on a random sub-sample of 10% of individuals in the data (854 households), over all their in-sample trips. Because of dimensionality concerns, we focus on the 4 products that generate the highest overall sales (which consist of 34% of the total in-sample sales),[15] and treat the rest as outside options. Finally, as previously indicated, we only consider consideration sets of size 0, 1 or 2.

The top 4 products are, respectively, Dannon Light N' Fit, Yoplait Original, Colombo Light and Yoplait Light. In the structural model, since we need to make across-trip comparisons of utility coefficients, we re-defined units to multiples of globally-available minimum units, rather than the minimum available units we use in reduced form. We adjust for choice probabilities for the unavailable units combinations.

5.5.2 Choice of characteristics

Given the choice of product set, we choose to focus on four characteristics – 3 brand indicators and the indicator for characteristics "light". These are denoted respectively, by $h = 1, \ldots, 4$.

5.5.3 Distribution of number of products

In the sub-sample, the distribution of the number of products chosen is shown in Table 4. Only 0.1% of the sub-sample purchased more than 2 different products, which justifies limiting the cap

[15] Measured in volumes; if we rank according to sales in dollars, the first 3 products are the same, while the 4th product would be regular Dannon.

Table 4: Distribution of number of products in the sub-sample

# products	frequency	share
0	10,476	0.63
1	5,726	0.35
2	359	0.02
3	17	0.00
4	1	0.00

Note: The table presents distribution of number of different purchased products in the same trip, in the sub-sample. Note that 0.1% of the sample purchased more than two different products.

of the consideration set at 2. As a side note, zero products indicates purchase of another yogurt not in the set of interest.

5.5.4 Distribution of purchase quantity

Figure 7 summarizes distribution of choice-sets, and Figure 8 summarizes conditional quantity distribution given a singleton choice set. Multiple-product choice occasions are uncommon but present, which justifies allowing for only a parsimonious interaction term in the consumption utility specification.

Conditional on choice, quantity distributions are heavily skewed to the right – often with the mode quantity larger than 1. This can be rationalized by the scale economy generated from a per-variety consideration cost.

6 Estimation results

6.1 Consumption utility estimates

Table 5 reports all parameter estimates. $\bar{\beta}_h's$ capture the marginal *consumption* utility characteristics h, which is in turn multiplied by the log-transformed purchase quantity. $\bar{f}_j's$ capture the benchmark consideration cost.

The mean of the transformed $\tilde{\beta}_{ij}$'s are, 1.31 for Dannon Light, 2.75 for Yoplait Original, 2.78

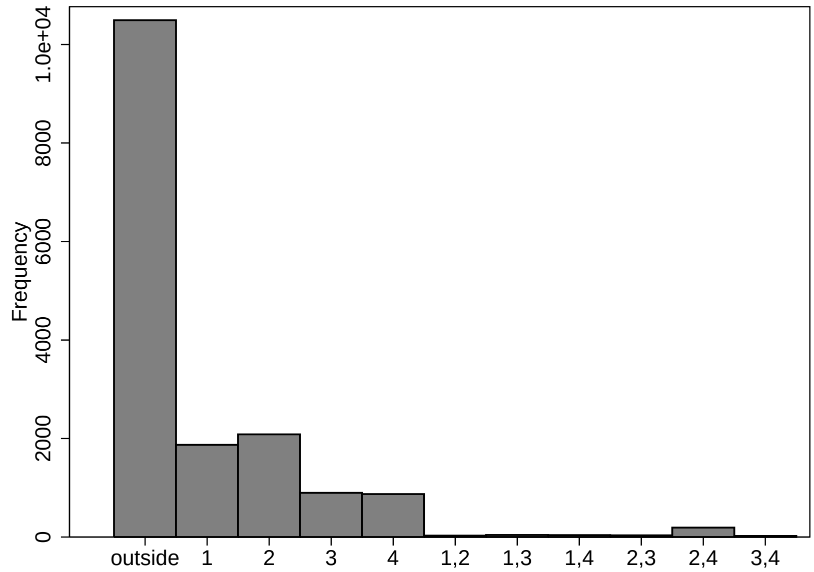

Figure 7: Quantity distribution of the sub-sample

Notes: This figure shows distribution of choices of bundles. This is creates using the full sample of households, on the choices of (combinations of) the top 4 products.

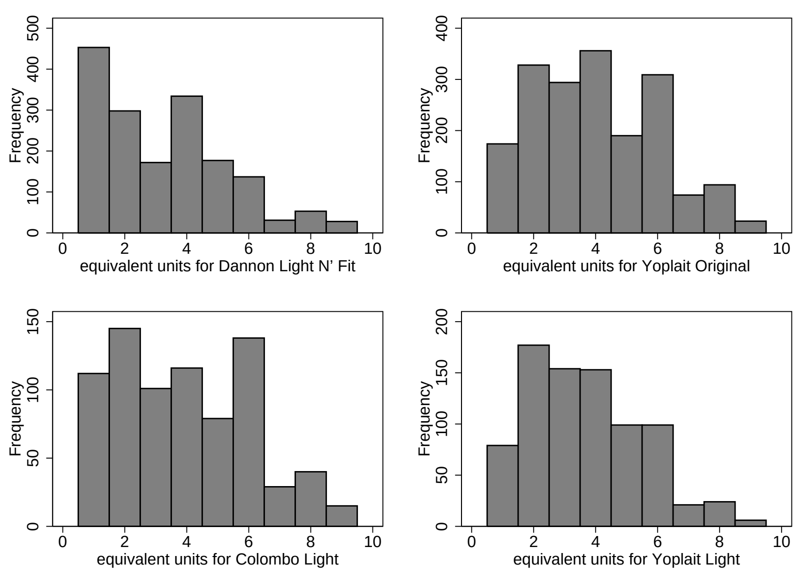

Figure 8: Quantity distribution of the sub-sample

Notes: The four figures depict quantity distributions given singleton bundle choice. As shown by Figure 7, multiple product choice is uncommon.

Table 5: Parameters estimates

	par. est.	std. err.
brand coef for Dannon ($\bar{\beta}_1$)	2.10	0.16
brand coef for Yoplait ($\bar{\beta}_2$)	2.75	0.13
brand coef for Colombo ($\bar{\beta}_3$)	3.57	0.21
characteristics coef for Light ($\bar{\beta}_4$)	-0.79	0.13
interaction of consumption utility (γ)	-1.16	0.13
price coefficient (λ)	1.44	0.06
consideration cost for Dannon Light N Fit ($\bar{f}_1$)	4.62	0.18
consideration cost for Yoplait Original ($\bar{f}_2$)	7.00	0.22
consideration cost for Colombo Light ($\bar{f}_3$)	7.49	0.24
consideration cost for Yoplait Light ($\bar{f}_4$)	6.82	0.23
changes in consid. cost for two products (Δf_2)	-0.88	0.20
changes in consid. cost under feature (Δf_A)	-0.60	0.09
scale of utility shock (κ)	7.49	0.10
std dev of brand coef ($\sigma_{\beta,1}$)	6.82	0.03
std dev of light coef ($\sigma_{\beta,4}$)	0.88	0.03
std dev of mean consid. cost (σ_f)	0.60	0.05

Note: Estimates for the all parameters. Standard errors are asymptotic (numerical).

for Colombo Light, and 1.96 for Yoplait Light. By defining random coefficients on 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.

While using log-transformed quantity restricts the curvature of consumption utility for a single product, we do estimate the decreasing marginal utility across products – as captured by γ. For example, the estimates on $\bar{\beta}_h$ and γ imply that, for an *average* consumer, consuming 1 unit of Yoplait Original brings a utility of \$0.57,[16] and consuming the second unit further increases her utility to \$0.91. If – rather than buying the second unit of Yoplait Original – she instead consumes a unit of Colombo Light together with the first unit of Yoplait, her utility would have increased to \$1.08.[17]

[16] $\tilde{\beta}_2 \cdot \log(1+1)/1.44$

[17] $\left(\tilde{\beta}_2 \cdot \log(2) + \tilde{\beta}_3 \cdot \log(2) + \gamma \cdot \log(2) \cdot \log(2)\right)/1.44$

6.2 Consideration costs

On the other hand, the money metric for *mean* consideration cost, $\bar{f}_j/\lambda$, for all four products are, respectively, \$3.21, \$4.86, \$5.20 and \$4.73. Compared to the observed expenditure on each product,[18] consideration cost is worth 1.2 - 1.9 times the per-trip expenditure. In addition, When purchasing multiple products, the total consideration cost is reduced by $\Delta f_2/\lambda = \$0.61$, suggesting a slight spillover in attention across products. Finally, when on feature advertising, a product's consideration cost is reduced by $\Delta f_A/\lambda = \$0.42$.

We also plot the distribution of F_{ij} (with simulated draws), normalized the price coefficient λ. These are shown in Figure 9. The solid lines plot the (Kernel) density for consumers who purchases, while the dashed lines plot the density of F_{ij}/λ for all consumers. This shows considerable heterogeneity in the fixed cost, and that consumers who purchase are the ones with lower consideration costs.

6.3 Varieties without consideration cost

The large F_{ij}'s suggest that thinking about each product is costly, which incentivizes a consumer to stay with fewer varieties and instead purchase large quantities. As an alternative measure of the impact of consideration cost, we simulate a consumer's choice of product and quantity, assuming that there is no consideration cost.[19] Figure 6.3 plots both the distribution of the number of products the consumers selected (under benchmark and counterfactual scenarios) and the distribution of quantity.

Since we find that consumers have strong love-for-variety preference, it is not surprising to see that their choice to spread expenditure over multiple products is limited by costly consideration. We find that if consideration is free, many consumers would choose to purchase multiple products, and very few decides to buy none. On average, the number of products a consumer purchases

[18] Conditional on product choice, the observed *average* per-trip expenditure on the 4 products are, respectively, \$2.61, \$2.92, \$3.49, \$2.46. The ratio between consideration cost and expenditure is 1.22, 1.66, 1.49 and 1.92.

[19] However, we maintain the assumption that a consumer can choose two products at most; this can be viewed as another form of consideration cost.

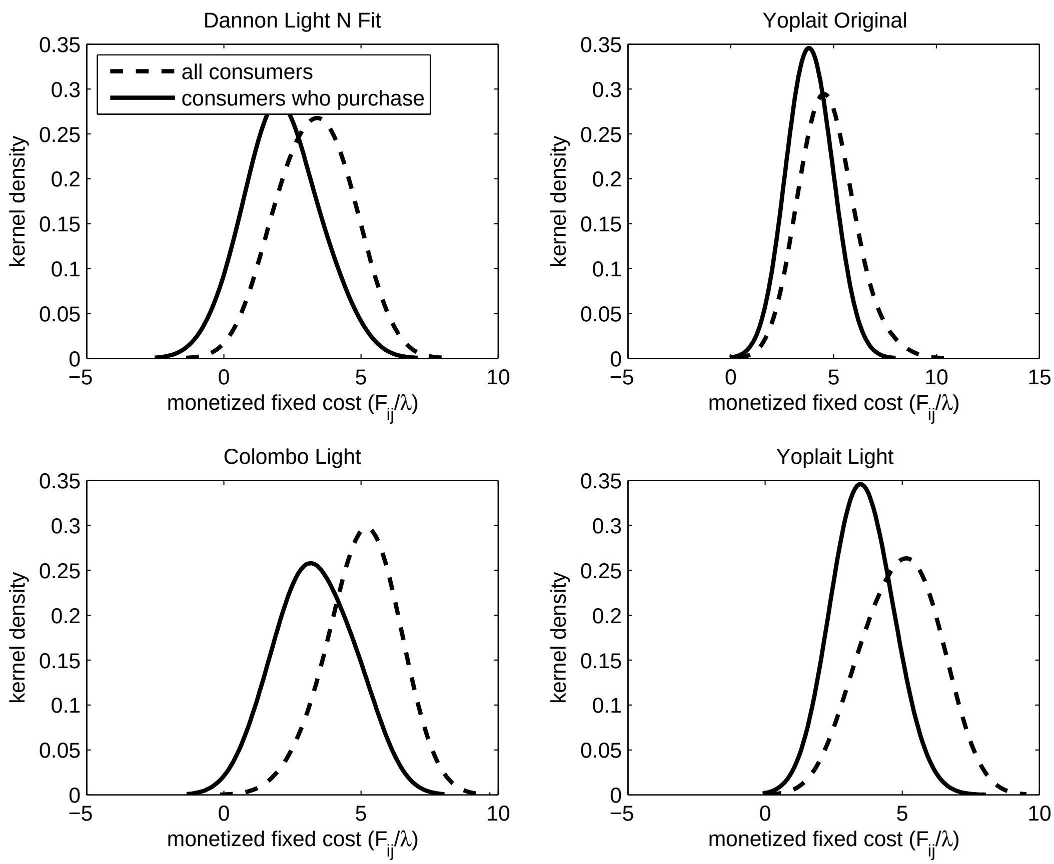

Figure 9: Distribution of consumer fixed costs

Notes: These figures present the distribution of fixed cost coefficients F_{ij}, normalized by the price coefficient λ, across consumers for each product. The normalization enables us to interpret the fixed cost in dollars. The solid lines plot the (Kernel) density for consumers who purchases, while the dashed lines plot the density of F_{ij}/λ for all consumers.

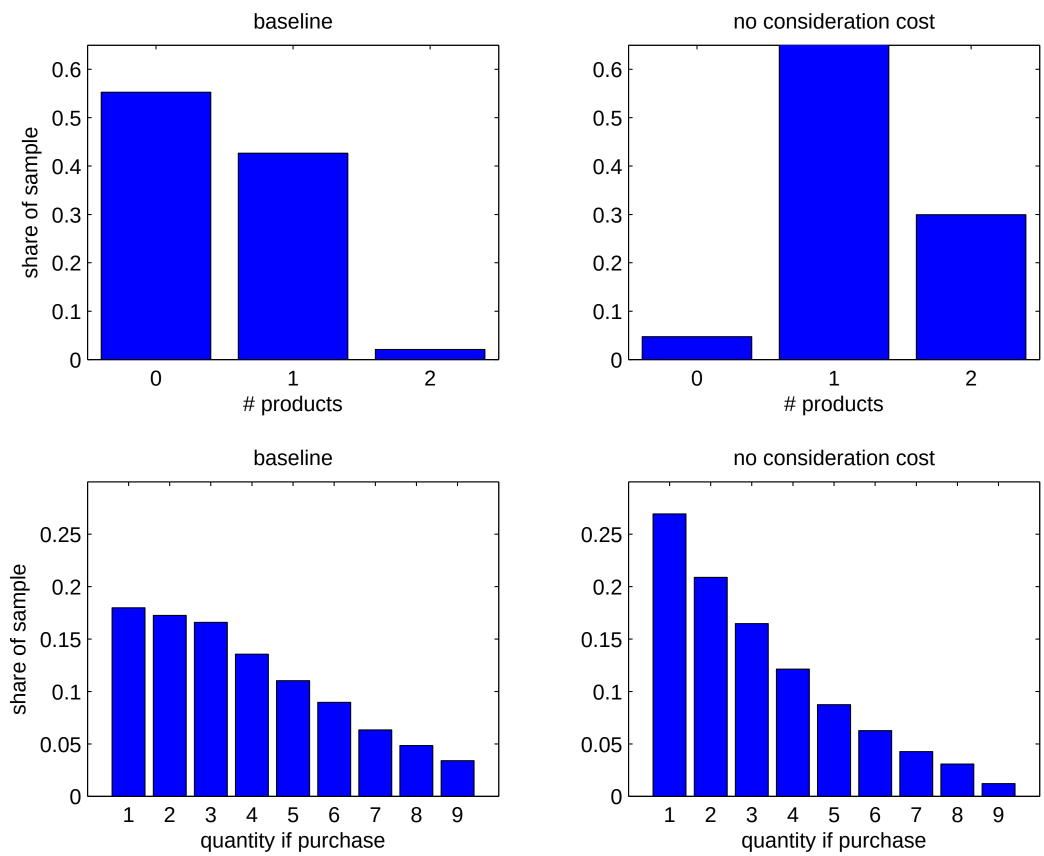

Figure 10: Distribution of the number of products and quantity per product

Notes: These figures present model-simulated distribution of quantity and number of products, in the benchmark scenario (as estimated), and under the scenario where consideration cost is set to zero. We find that consumers on average purchase 1.2 products when consideration is free, instead of the 0.4 products in the baseline.

moves from 0.4 to 1.2 (out of 4), which means that limited consideration restricts variety to 1/3 of its unconstrained optimum. We also find that quantity per-product decreases when a consumer starts to purchase more varieties.

7 Price and consideration-cost elasticities

7.1 Overall price elasticities

We compute the implied elasticities when one of the three products reduce prices by 5%. To do so, we first compute quantity choices based on each of 50 draws in the random coefficient, and

Table 6: Implied elasticities

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-1.85	0.14	0.33	0.17
Yoplait Original	0.16	-2.84	0.15	0.40
Colombo Light	0.16	0.09	-2.36	0.24
Yoplait Light	0.07	0.23	0.18	-2.84

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

then average them across draws. We then compute elasticities based on the percentage changes in the *average* quantity. We then decompose elasticities into changes in the consideration sets, and changes in quantity conditional on the consideration set.

Table 6 reports elasticities of the overall purchase quantities as response to a price change. The signs of the own- and cross- price elasticities are conventional, and the magnitude intuitive. For example, the own price elasticities for Danone Light N' Fit implies a 1.45% quantity increase, for each 1% price drop from the product.

Note that the numbers here are conditional on within-category purchases. Bell et al. (1999) find that within-category switching accounts for 75% of the total price elasticities, which hints that the within-category elasticity we find is a major part of the total elasticities.

7.2 Decomposition of price elasticities

With a model of consideration set formation and purchase choice, we can decompose the overall price elasticities into two parts. First, a reduction in price drives up quantity choice conditional on consideration; second, it also facilitates consideration set membership, in the sense that a consumer might now find it attractive to consider a product.

The decomposed elasticities are presented in Table 7 and 8. We find that consideration set membership is very responsive to price changes, as the own-price elasticities shown in Table 7 are close in magnitude compared to the overall elasticity estimates. This is not surprising, given the nature of scale economy driven by the model estimates. Because considering a product is costly

Table 7: Implied consideration-set formation elasticities

	(A)	(B)	(C)	(D)	others
Dannon Light N Fit	-1.25	0.07	0.24	0.13	0.26
Yoplait Original	0.07	-2.13	0.12	0.40	0.47
Colombo Light	0.13	0.06	-1.53	0.19	0.16
Yoplait Light	0.04	0.16	0.13	-1.92	0.13

Note: The i, j element is the elasticity of consideration probability of product j, on price change of product i, or $\frac{\partial \Pr(j \in K_{it})}{\partial P_i} \cdot \frac{P_i}{\Pr(j \in K_{it})}$.

Table 8: Implied elasticities given consideration

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-0.49	0.00	0.00	0.00
Yoplait Original	0.00	-0.97	0.00	0.00
Colombo Light	0.00	0.00	-1.06	0.00
Yoplait Light	0.00	0.00	0.00	-1.50

Note: The i, j element is the elasticity of purchase quantity of product j, given that j is the only product considered, on price change of product i; or: $\frac{\partial Q_j}{\partial P_i} \cdot \frac{P_i}{Q_j}$ given that $K_{it} = \{j\}$.

apart from price, the consumer will be likely to find it optimal to ignore an entire product when the prices are high.

On the contrary, elasticities given consideration is low – when a consumer is only considering a single product. Here, products are local monopolists in singleton sets, where the consumer can only choose a single product or the outside option.

7.3 Consideration cost elasticities and its decomposition

We then compute the elasticities of consumer purchase decisions to a change in consideration cost. To do so, we simulate choices when a product is on a hypothetical feature advertisement that reduces consumer consideration cost by 5%. Table 9 reports overall consideration cost elasticities, while Table 10 isolate the effect on consideration set formation, similar to the previous section.

It is worth noting that the overall consideration cost elasticities imply that a 1% consideration cost reduction converts to 1.70-3.75% increase in overall purchase quantities. Multiplied by the

Table 9: Implied consideration cost elasticities

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-1.70	0.19	0.44	0.30
Yoplait Original	0.32	-3.66	0.25	0.67
Colombo Light	0.26	0.12	-2.39	0.30
Yoplait Light	0.14	0.40	0.27	-3.75

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

Table 10: Implied consideration cost elasticities on consideration set formation

	(A)	(B)	(C)	(D)	others
Dannon Light N Fit	-1.88	0.03	0.26	0.14	0.34
Yoplait Original	0.15	-3.76	0.23	0.56	0.79
Colombo Light	0.20	0.10	-2.62	0.24	0.25
Yoplait Light	0.13	0.24	0.30	-3.98	0.24

Note: The i, j element is the elasticity of consideration probability of product j, on change of search cost of product i, or $\frac{\partial \Pr(j \in K_{it})}{\partial F_i} \cdot \frac{F_i}{\Pr(j \in K_{it})}$.

percentage decrease in consideration cost from feature advertising (12-16%), this means that featuring a product is associated with 27-64% sales increase.[20] Of course, given that we condition the analysis on category purchase, this calculation does not take into account the effect of feature in driving a consumer to the refrigerated section.

8 Feature advertising and price discounts

8.1 Informative versus persuasive feature: intuition

When feature advertising is complementary to consumption (Becker and Murphy, 1993) – as commonly imposed in the empirical choice modeling literature (Chintagunta, 1993; among others) – feature advertising increases the relative importance of consumption utility, and comparatively,

[20]The highest sales increase is on Yoplait Original.

downplays the role of price.[21] Hence, featuring a products decreases its price elasticities, and this strategy should be seen as a substitute to a price discount. This implies,[22] that feature and price discounts should be asynchronous, because consumers facing feature are less price elastic. This is at odds with the observation that they are usually seen simultaneously (Figure 2).

In this framework, however, feature advertising is seen as a substitute to discounts in the consideration stage, following the previous argument. However, it is seen as a complement to price discounts in the purchase stage. To intuitively see the second point, Figure 11 shows posterior distribution of consumption utility coefficients ($\tilde{\beta}_{ij}$), separately for two groups of consumers: one group who regularly purchases the product when it is not on feature, and another group who only makes the purchase when the product is featured. The second group of consumers are the ones with lower consumption utility, or in other words more price elastic. Hence, featuring a product raises the overall quantity elasticities, which then implies that prices should also go down to "further persuade" a consumer into purchasing. This is consistent with the empirical regularity that featuring a product is associated with a price discount.

[21] This property comes from the logit structure. To see this, specify a logit model

$$u_1(f,p)+\varepsilon=\beta_1 f-\beta_2 p+\varepsilon_1,$$

and

$$u_0=\varepsilon_0.$$

Because market share is

$$s(f,p)=\frac{\exp(\beta_1 f-\beta_2 p)}{1+\exp(\beta_1 f-\beta_2 p)},$$

it is apparent that

$$\begin{aligned}\mathscr{E} &= \frac{\partial s}{\partial p}\cdot\frac{p}{s}\\ &= -\beta_2 p(1-s)\end{aligned}$$

and $|\mathscr{E}|$ should be decreasing in s, which is an increasing function of f.

[22] If we do not consider the role of feature advertising on providing price information. In this context, this is abstracted away given that we focus on choices conditional on category purchase.

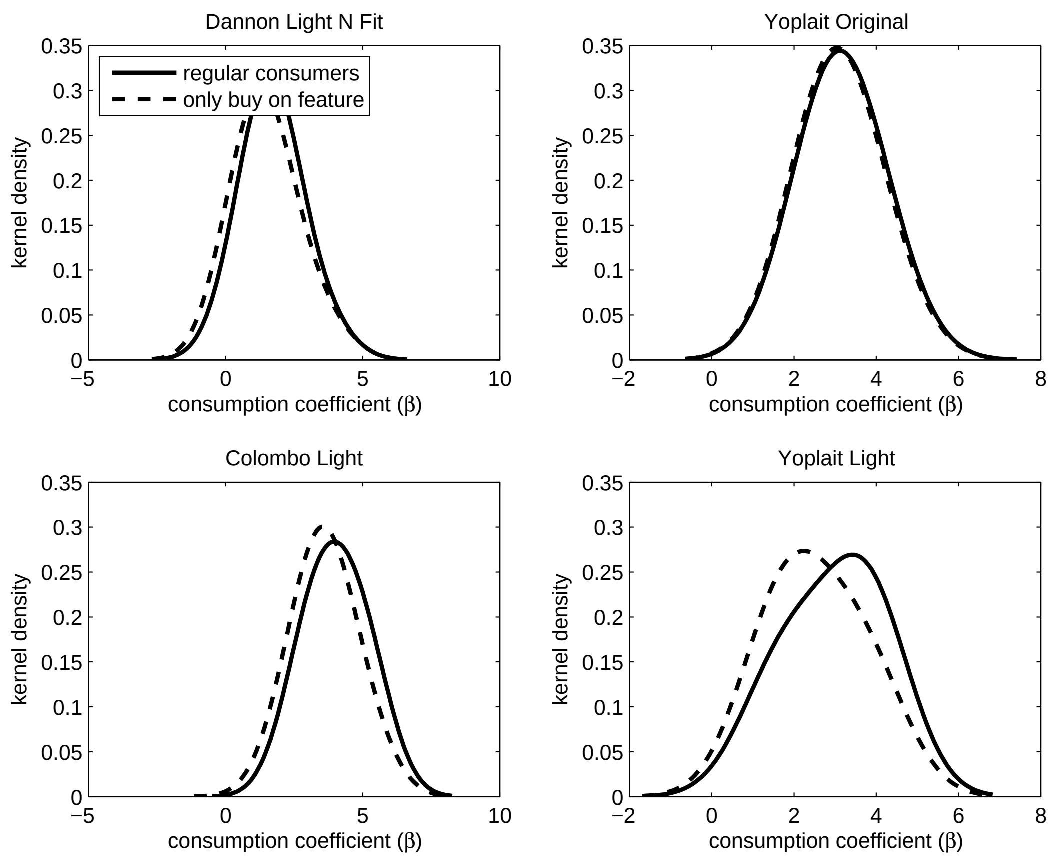

Figure 11: Distribution of utility coefficients for different consumers

Notes: These figures present the distribution of consumption utility coefficients $\tilde{\beta}_{ij}$ across consumers, for different products. The solid lines plot the posterior (Kernel) density for consumers who purchases the product without it being featured – i.e "regular consumers". On the other hand, the dashed lines plot the density of $\tilde{\beta}_{ij}$ for consumers who purchase *only when* the product is on feature. The bandwidth for Kernel estimators is set to 1.

Table 11: Changes in overall price elasticities when on feature

	(A)	(B)	(C)	(D)
Dannon Light N Fit	0.91	1.17	0.89	0.95
Yoplait Original	1.09	0.91	0.70	1.15
Colombo Light	1.02	1.16	0.97	0.79
Yoplait Light	1.00	0.61	0.94	0.91

Note: Each cell i, j presents the ratio between i's *counterfactual* elasticity to j when j is on feature with probability 1, to the benchmark elasticity presented in Table 6 (when j is featured as data shows). That is, the i, j element is $\frac{\mathscr{E}_{ij}|_{f_j=1}}{\mathscr{E}_{ij}|_{f_j}}$.

8.2 Feature advertising and price elasticities

With a model where feature reduces consideration cost (and hence is informative), we can then investigate the effect of featuring a product on its own- and cross- price elasticities. Table 11 reports the ratio between the (overall) price elasticities when the row product is on feature with probability 1, and the benchmark price elasticities when products are on/off feature as observed. We find that being on feature *reduces* the own-price elasticities by a magnitude of 7%-9% – which at a first glance suggest that feature advertising is a substitute to price discounts.

On the other hand, decomposing the elasticities, we find that quantity elasticity conditional on consideration set membership has increased as a result of being on feature, presented in Table 12.[23] This is because, when consideration costs are reduced by featuring, some consumers with lower $\tilde{\beta}_{ij}$'s are going to start considering the product – these consumers would otherwise think it is unfruitful to think about product j due to the low $\tilde{\beta}_{ij}$'s. At the same time, the low $\tilde{\beta}_{ij}$'s suggest that these are the higher-elasticity consumers, whose participation then drives up the overall price response.

[23] In this table, since we focus on quantity elasticities given singleton consideration sets, cross elasticities are by construction zero and are ignored.

Table 12: Changes in conditional quantity elasticities when on feature

	ratio between own-elasticities
Dannon Light N Fit	1.11
Yoplait Original	0.93
Colombo Light	1.07
Yoplait Light	1.18

Note: See note in Table 11, except that this table reports ratios in quantity choice elasticities conditional on consideration.

9 Concluding remarks

In this paper, we quantify the fixed mental cost a consumer has to incur when considering each product. These costs generate scale economies in consumer's quantity choice and encourages her to purchase large quantities instead of many products. Therefore, this explains why a consumer is unresponsive to small price changes.

With panel data on each consumer's purchase record, we demonstrate how each consumer's choice of purchase quantity – under price discounts that are just enough to convert them to purchase – can be used to identify (lack of) consideration from (low) preference. In fact, without costly consideration, consumers would have gradually re-allocate their expenditure under gradual price change, rather than sharply switching from one product to another. We demonstrate how one can test this in reduced form, using standard marketing scanner data.

We then construct a structural model of multiple product choice and the subsequent quantity choices, with an endogenous costly consideration decision in the first stage. Our model deviates from canonical models, which treat quantity choices of different products as separate decisions. Our model characterizes a consumer's active choice of consideration set, and can thus related to the identification strategy we propose. As by-products, the model also allows for flexible nonlinear price structures and discrete quantities.

We estimate the model using IRI behavioral scan panel data, in the yogurt category. In our case, the average (monetized) consumer consideration cost per yogurt product is between \$3.2 and \$5.2 in a given trip. This is about 1.2-1.9 times the total expenditure on each product per

trip. The magnitude of consideration cost indicates that prices are not the dominant explanation of which subset of product to choose from; and conversely, this explains why price elasticities are so small on quantity choice conditional on purchase. We also find that without costly consideration, a consumer would have chosen 3 times as many different products, to cater her love-for-variety preference.

Finally, we find that feature advertising reduces the consideration cost by \$0.42, which is on the order of 10% of the total consideration cost. This drives more consumers to purchase, but because of preference heterogeneity, mostly affects consumers with lower tastes – those consumers would otherwise not purchase. This creates a selection problem, that consumers who are attracted during a feature promotion are the ones with lower tastes and higher price elasticities. Hence, our model justifies the synchronized use of price and feature promotions.

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.

Becker, Gary S and Kevin M Murphy (1993), 'A simple theory of advertising as a good or bad', *The Quarterly Journal of Economics* pp. 941–964.

Bell, David R, Jeongwen Chiang and Venkata Padmanabhan (1999), 'The decomposition of promotional response: An empirical generalization', *Marketing Science* **18**(4), 504–526.

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

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

Chintagunta, Pradeep K (1993), ‘Investigating purchase incidence, brand choice and purchase quantity decisions of households’, *Marketing Science* **12**(2), 184–208.

Dehmamy, Keyvan and Thomas Otter (2014), ‘Utility and attention-a structural model of consideration’, *Available at SSRN 2433145* .

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.

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

Goeree, Michelle Sovinsky (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.

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

Kawaguchi, Kohei, Kosuke Uetake and Yasutora Watanabe (2014), ‘Identifying consumer inattention: A product-availability approach’, *Available at SSRN* .

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

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 SH 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.