Bundle Choice with Limited Consideration: An Application to Yogurt

Xiaoyu(Fisher) Yu*

November 22, 2022

Click here for the latest version.

Abstract

I combine complementarity in bundle choice and consideration set of products into a demand model for differentiated products. I show identification of consideration probabilities from the asymmetric demand even when only marginal market shares of products, not bundle options, are observed. I apply the model to the yogurt market with consumer-level and store-level data to quantify the complementarity and the degree of limited consideration. I tailor a novel estimation approach that features efficiency gain via combining the loglikelihood of consumer choices and market shares with moment conditions to my empirical application. I find a considerable demand synergy in the bundle of different products and a significant proportion of consumer inertia, defined here as choosing from the last purchases, in consumer demand for yogurt. Compared to the standard discrete choice model, my estimation results suggest that accounting for complementarity between products and consumers' limited consideration set can substantially affect price competition analysis.

Keywords: Discrete Choice, Bundle Demand, Limited Consideration

---

*Department of Economics, Rice University. Email: fisheryu@rice.edu. I would like to thank Xun Tang, Maura Coughlin, Robin Sickles, Isabelle Perrigne, Arun Gopalakrishnan, Matthew Thirkettle and Jeremy Fox for their valuable advice and guidance. I also thank Qinyou Hu, Sen Lu, Dibya Mishra, and all the participants at Rice Economics Department Seminar for useful comments and discussions. I gratefully acknowledge the support from all faculty members of Rice University, Department of Economics. I would like to thank IRI. for making the data available. All estimates and analysis in this paper, based on data provided by IRI. are by the author and not by IRI.

1 Introduction

The workhorse model in demand estimation is the random coefficients discrete choice demand model, introduced by Berry, Levinsohn, and Pakes (1995) (BLP henceforth), who provide a tractable framework to estimate flexible substitution patterns between many differentiated products in the presence of price endogeneity. Like the standard discrete choice model, the BLP framework presumes that consumers choose one single option and consider all available products. However, it has long been recognized in consumption literature that consumers can choose different products at one time (see an early survey by Houthakker (1961)), and they may not be fully aware of all products available in the market during the decision process, e.g., Roberts and Lattin (1991). Abstracting away from more realistic behavioral assumptions could produce misleading results in demand estimation, which is always crucial for understanding a firm's pricing and marketing strategy, and even for conducting competition analysis in the industry.

Motivated by these two limitations, in this paper, I develop a new discrete choice model that extends the existing BLP framework to simultaneously account for (i) extra utility(disutility) from choosing multiple products as a bundle and (ii) a possibility of the limited consideration set. In the model setup, I treat the deterministic utility change as a model primitive when the consumer chooses multiple products jointly rather than separately. More specifically, it is the difference between the average utility of any bundle and the sum of the average utility of all products in the bundle. I refer to this utility discrepancy as "demand synergy" following Iaria and Wang (2020) and Wang (2021). Complementarity in terms of compensated demand can arise when the demand synergy in the bundle of products is positive (as shown in Gentzkow (2007)), as opposed to the standard discrete choice model where alternatives are strictly substitutes by nature. As for the consideration stage, I parameterize the consideration set formation as the hybrid consideration specification introduced by Abaluck and Adams-Prassl (2021). This specification combines the classic default specific consideration (DSC) and alternative specific consideration (ASC) to allow inertia, which often arises from previous purchases, and attention shifters like advertising to affect the consideration set format. The consumer can choose one product as a single option or a bundle option of multiple products from the consideration set. I discuss the identification conditions for my empirical specification. Given the available household purchase data, the observed default consideration set, and excluded instruments for the price, the model parameters are point identified. I also show that when the bundle option is allowed, the hybrid consideration model can be point identified with only market share data of products, not bundles. However, the presence of bundle options, in general, fails the nominal illusion property in the latent choice probabilities; hence primitives on preference are not nonparametrically identified alone.

I apply the framework above to consumers' demand for yogurt. As mentioned by Dubé (2019), complementarity may be an important part of choices for product categories like yogurt, where consumers purchase large assortments of flavors or variants. In my empirical application, I use product-level data from retailers' scanners (macro data) in combination with the consumer panel on household yogurt purchases (micro data). From the consumer panel, I find that nearly 30% of yogurt purchase incidences involve multiple products, and around 60% of choices stick to what consumers chose last time. My empirical model treats the multiple products in one shopping trip as a bundle option. I attribute the pattern of choice persistence partly to the default consideration set from household last yogurt purchases. My estimation strategy is adapted from a novel method proposed by Grieco, Murry, Pinkse, and Sagl (2022) who use macro and micro data in a mixed likelihood approach, combined with the standard BLP moment conditions. The main advantage of this approach is the efficiency gain from the maximum likelihood estimator, compared to other classic approaches such as micro moments in Berry, Levinsohn, and Pakes (2004) and Petrin (2002). I adjust their method by (i) allowing for the consideration stage in forming individual choice probability and (ii) aggregating the model-predicted bundle shares into marginal product market shares to match the store sales data. My estimation results indicate a sizeable positive demand synergy on average in bundled yogurt products, and the synergy increases with the household size. There exists a substantial degree of inertia, up to 80% chance that the household considers only last-time purchases and the outside option. I do not find much individual heterogeneity in price sensitivity, and the advertising effect becomes smaller in my model. I also conduct post estimation analysis of price effects with the estimated parameters and find quite different diversion ratios and cross-price elasticities, compared to those obtained from a standard BLP model. I find larger magnitudes in my model and opposite signs for products in a bundle. In particular, for two products often bundled, the complementarity caused by considerable demand synergy dominates the substitution effect when we evaluate how the marginal market share of one product responds to the price change in the other.

To the best of my knowledge, this paper is among the first combining the complementary options and limited consideration in a discrete choice demand model. Hence it mainly contributes to two threads of literature in modeling consumer demand.

Complementary goods purchased as a bundle are commonplace in the consumer's shopping basket. Early papers such as Hendel (1999) and Dubé (2004) extend the discrete choice model to multiple discreteness in different products demand. The major difference compared to the demand synergy in my model is that their complementarity comes from a specific nonlinear function form in utility, rather than the affine relationship in mean utility between bundle and bundle components as Gentzkow (2007). There are also models that capture cross-category complementarity in a continuous demand framework, e.g., Mehta and Ma (2012); Thomassen, Smith, Seiler, and Schiraldi (2017). However, the complementary across products has yet to be well adapted into the demand model for differentiated products, mainly due to the complexity of identifying the complementarities. Gentzkow (2007) proposes an exclusive restriction and a panel structure as two identification strategies to distinguish the complementarity in the deterministic part of the utility from the correlation in error terms. Fox and Lazzati (2017) and Allen and Rehbeck (2019) discuss the nonparametric identification of additive utility in bundles in a simple discrete choice model. My setup for the complementarity between alternatives in the BLP framework is closest to Iaria and Wang (2020) and Wang (2021). Both papers propose the same bundle utility specification as the affine relationship, and their identification results rely on the same demand inverse that maps the product-level market shares to the market-product mean utility. Iaria and Wang (2020) propose a new instrument-free identification argument and a maximum likelihood estimation approach when market shares of bundles are observed. On the contrary, Wang (2021) requires excluded instruments for identification when only marginal product-level market shares are observed, leading to a GMM estimation approach. Nonetheless, existing papers on complementarity in demand have always assumed the full choice set, yet to take into account the limited consideration set.

On the thread of literature on limited consideration in consumer's choice, a simplified consideration set model assumes away the structural model for consideration formation, such as the searching process. Without auxiliary data available to supplement the observed choices, most empirical papers consider two probabilistic specifications on the consideration set formation. One is the default specific consideration (DSC), assuming a probability that the decision-maker always adheres to the default option or a default subset of options such as the chosen one(s) in the status quo (Hortaçsu, Madanizadeh, & Puller, 2017; Ho, Hogan, & Scott Morton, 2017; Heiss, McFadden, Winter, Wuppermann, & Zhou, 2021). The second one is the alternative specific consideration (ASC), in which the agent is supposed to consider each option with a probability independent from the consideration probabilities of other options (Ben-Akiva & Boccara, 1995; Goeree, 2008; Van Nierop, Bronnenberg, Paap, Wedel, & Franses, 2010). My consideration model adapts the hybrid model of DSC and ASC introduced in Abaluck and Adams-Prassl (2021) to simultaneously capture the two different types of consideration formation process.[1]

Most existing empirical research with probabilistic specification on the consideration stage often relies on the two-way exclusion restrictions on the consideration stage and utility shifters to separately identify the parameters in each part, e.g., Goeree (2008). Van Nierop et al. (2010) validates the exclusion restrictions in a grocery shopping context from online experimental data tracking participants' consideration sets. A noticeable alternative identification strategy is developed by Abaluck and Adams-Prassl (2021). Without imposing exclusion restrictions, they explore the identification of consideration probability and latent choice probabilities conditional on the consideration set by exploiting the cross-partial derivatives of the demand function. They assume there exists an exogenous covariate that can shift both utility and consideration probability. Under full consideration, the cross-partial derivatives of market shares with the covariate are symmetric, while asymmetric under limited consideration. So asymmetry property helps identify the consideration process once the cross-partial derivatives are observed from the data. I extend their identification results by allowing the bundling option while keeping the same data structure, so that the observed market shares in their case becomes the observed product-level market shares. Regarding previous studies using the same IRI academic datasets (Zhang, 2006; Van Nierop et al., 2010), they estimate the consideration set formation with only the consumer panel. This paper allows for choosing a bundle option from the consideration set and applying both macro and micro data in estimation.

Lastly, the yogurt industry has been actively analyzed to study important questions in empir-

[1] A growing trend in literature is to avoid the probabilistic consideration model via imposing restrictions on the consideration set formation (Cattaneo, Ma, Masatlioglu, & Suleymanov, 2020; Crawford, Griffith, & Iaria, 2021; Lu, 2021; Barseghyan, Coughlin, Molinari, & Teitelbaum, 2021). It is interesting to explore the possibility of combining weak restrictions on consideration set and a BLP-type demand model.

ical industrial organizations like distinguishing informative and prestige effects of advertisement by Ackerberg (2001, 2003), and the vertical relationships in Villas-Boas (2007). Some industry-specific studies are also appearing lately (e.g., Rossetti, 2018; Liu, 2019; Triggs, 2021). My paper distinguishes itself from others through its behavioral assumptions of the demand model.

The applicability of my framework definitely goes beyond yogurt consumption and allows to address more general empirical questions. From the firm's perspective, does more brand advertisements in category A have a spillover effect on brand awareness in category B, when two categories have positive demand synergy? How would the promotion of one product affect the demand for its complements, especially when the promotion itself can shift consumers' attention? From the antitrust perspective, what is the correct diversion ratio between the target and acquirer in a horizontal merger? My model can provide valuable insights into these questions and highlight the importance of consumer purchase data across multiple products and categories. Companies usually learn a full picture of the consumer behaviors from their retailers. Thus, my analytical strategy helps provide fundamental business logic for the fast-growing retail media market where the retailer helps the brand better target potential customers via the retailer's first-party consumer data.

The remainder of this paper is structured as follows: Section 2 introduces the general setup of my demand model and discusses some identification results. In Section 3, I describe both product-level and consumer-level data, and provide suggestive evidence on the consumer behavioral patterns that motivate my empirical specification. The empirical model, estimation strategy, estimation results, and price effects analysis are presented in Section 4. Section 5 concludes.

2 Model

In this section, I first formally set up my model. My model starts from the individual utility of alternatives (including bundling options) and the consideration stage to the individual choice probabilities. Lastly, market shares of each option are aggregated from individual choice probabilities by the distributions of individual preferences. They can further sum up the marginal market shares at the product level. I discuss the identification result with different data settings. I can use individual choice data from the consumer panel to identify the parameters of demand synergy and the consideration probabilities in my empirical application. When only product-level market shares are available, I show how the consideration probability formation is identified.

2.1 General Setup

The discrete choice model consists of two steps: individual i in market t first considers a subset of all feasible products as his/her consideration, and then i selects either a single option or a bundle of items from the consideration set in the first step.

Assume that there are $t = 1, \ldots, T$ independent markets, and $i = 1, \ldots, I_t$ individuals in each market t. The feasible choice set $\mathbf{J}_t$ collects all "inside" products and the outside option in market t. Consumer i[2] considers a non-empty subset $\mathbf{C}_{it} \subseteq \mathbf{J}_t$ of products from which he/she can choose one product as a single option or multiple products as a bundle option. I first model the consideration set formation following the hybrid model of default specific consideration and alternative specific consideration introduced in Abaluck and Adams-Prassl (2021), namely

$$\Pr(CS = \boldsymbol{C}_{it}) = \mu_0 \mathbb{1}(\boldsymbol{C}_{it} = \boldsymbol{C}_{i0}) + (1 - \mu_0) \prod_{j \in \boldsymbol{C}_{it}} \phi_{ijt} \prod_{j \notin \boldsymbol{C}_{it}} (1 - \phi_{ijt}), \tag{1}$$

where $\mathbf{C}_{0it}$ is the default consideration set representing the inertia status, and I allow it to vary across individuals and markets. In my empirical application, I construct the default consideration set as products purchased during their previous purchase trip plus the outside option. It is common in the literature to degenerate $\mathbf{C}_{0it}$ to the default option (Hortaçsu et al., 2017; Abaluck & Adams-Prassl, 2021). The term ϕ_{itj} is the consideration probability for alternative j in the ACS part. Following Goeree (2008) and Van Nierop et al. (2010), I model $\phi_{it}(z_{jt}, v_{it})$ as a function of observables z_{jt} that shift the consideration probability of product j, and individual unobservables v_{it}. The distribution of v_{it} is parameterized by $F_v(\cdot)$, and is often assumed to be independent from the random coefficient β_{it} in the consumer'as utility. The standard identification argument from exclusion restrictions requires non-overlapping covariates in consideration and choice stages, hence at least one of the observables needs to be excluded from the utility specification.

I adapt the bundle utility specification in Iaria and Wang (2020) and assume that consumers choose a single or bundle option from individual-specific consideration set $\mathbf{C}_{it}$, instead of the feasible

[2]I may use individual consumer, household interchangeably.

choice set. Let $\mathbf{C}_{ikt}, k = 1, 2, \ldots$ denote the collection of all options from individual i's consideration set $\mathbf{C}_{it}$ and k denotes the number of products in the option. We always include the outside option in $\mathbf{C}_{it}$ and denote a set of all single options as $\mathbf{C}_{i1t}$. The indirect utility for choosing single product is specified as

$$U_{ijt} = u_{ijt} + \varepsilon_{ijt} = \delta_{jt} + \eta_{ijt} + \varepsilon_{ijt},\ j \in \boldsymbol{C}_{it}, \quad U_{i0t} = \varepsilon_{i0t}, \tag{2}$$

where δ_{jt} is the market t-specific average utility of product j, η_{ijt} an unobserved individual-specific utility deviation from δ_{jt}, and ε_{ijt} and ε_{i0t} are idiosyncratic taste shocks for inside single option and outside option, respectively. The market-specific average utility of product j often follows the linear specification as

$$\delta_{jt} = -\alpha p_{jt} + \beta' x_{jt} + \xi_{jt},$$

where p_{jt} is the price of product j in market t and x_{jt} are product j's characteristics that may change across markets. In my empirical application, x_{jt} includes the advertisement, the flavor variety and also product j's fixed effect. ξ_{jt} denotes the market level demand shock to product j, and price endogeneity implies $E(p_{jt}\xi_{jt}) \neq 0$.

Utility for bundle $\boldsymbol{b} \in \mathbf{C}_{ikt}$ is specified as the additive sum of single options' deterministic utility (the total utility δ_{jt} subtracted from the idiosyncratic shocks ε_{ijt}), plus the additional demand synergy $\Gamma_{i\boldsymbol{b}t}$ with another bundle-specific idiosyncratic shock $\varepsilon_{i\boldsymbol{b}t}$. We can decompose $\Gamma_{i\boldsymbol{b}t}$ into market average synergy for bundle $\boldsymbol{b}$ $\Gamma_{\boldsymbol{b}t}$ and the individual deviation from the average synergy $\zeta_{i\boldsymbol{b}t}$ ($\Gamma_{i\boldsymbol{b}t} = \Gamma_{\boldsymbol{b}t} + \zeta_{i\boldsymbol{b}t}$). Finally, the total utility of individual i choosing bundle $\boldsymbol{b}$ (whether it is a single or bundle option) in market t is additively separable in market-average utility level of the option $\delta_{\boldsymbol{b}t}$, individual deviation to the average bundle utility $\eta_{i\boldsymbol{b}t}$ and the idiosyncratic shock $\varepsilon_{i\boldsymbol{b}t}$:

$$\begin{aligned} U_{i\boldsymbol{b}t} &= \sum_{j \in \boldsymbol{b}} u_{ijt} + \Gamma_{i\boldsymbol{b}t} + \varepsilon_{i\boldsymbol{b}t} \\ &= \underbrace{\sum_{j \in \boldsymbol{b}} \delta_{jt} + \Gamma_{\boldsymbol{b}t}}_{\delta_{\boldsymbol{b}t}} + \underbrace{\sum_{j \in \boldsymbol{b}} \eta_{ijt} + \zeta_{i\boldsymbol{b}t}}_{\eta_{i\boldsymbol{b}t}} + \varepsilon_{i\boldsymbol{b}t}. \end{aligned} \tag{3}$$

Under some distributional assumptions, we can obtain individual choice probabilities and the market shares of bundle by aggregating over individuals' choice probabilities. We first assume i.i.d Type-I extreme value distribution for all idiosyncratic taste shocks ε. Hence the individual choice probability of any option $\boldsymbol{b}$ conditional on the consideration set C_{it} has a multinomial logit expression. Then we average conditional choice probabilities by the probability of considering the bundle $\boldsymbol{b}$ to obtain the unconditional choice probabilities as follows

$$s_{i\boldsymbol{b}t} = \sum_{\mathbf{C}_{it} \in \mathbb{C}(\boldsymbol{b})} \Pr(CS = \mathbf{C}_{it}) \frac{\exp\left(\delta_{\boldsymbol{b}t} + \eta_{i\boldsymbol{b}t}\right)}{\sum_{k=1}^{K} \sum_{\boldsymbol{b}' \in \boldsymbol{C}_{ikt}} \exp\left(\delta_{\boldsymbol{b}'t} + \eta_{i\boldsymbol{b}'t}\right)}, \tag{4}$$

where $\mathbb{C}(\boldsymbol{b})$ collects all possible consideration sets covering products in bundle $\boldsymbol{b}$. The individual choice probabilities of the single product j have the similar multinomial logit form as Eq.(4) by replacing $\boldsymbol{b}$ with j.

Next we calculate the market share of bundle $\boldsymbol{b}$ in market t. Here we parameterize the distribution of the individual deviation $\eta_{i\boldsymbol{b}t} = \sum_{j \in \mathbf{b}} \eta_{ijt} + \zeta_{i\mathbf{b}t}$ via random coefficients β_{it} governed by some distribution $F(\cdot)$ with parameters Σ_F. We remark that $\eta_{i\boldsymbol{b}t}(\beta_{it})$ can also be a function of demographics of individual i and product characteristics in option $\boldsymbol{b}$ in market t. Regarding consideration formation stage, given the distribution $F_v(\cdot)$ and its independence with random coefficients β_{it} in the utility following Goeree (2008), we can obtain the market share of option $\boldsymbol{b}$ as follows:

$$s_{\boldsymbol{b}t} = \iint s_{i\boldsymbol{b}t}\left(\delta_t\left(\Gamma_t\right), \beta_{it}, \phi_{it}\right) dF\left(\beta_{it}; \Sigma_F\right) dF_v(\cdot), \tag{5}$$

where $F_v(\cdot)$ is the distribution of unobservable v_{it} in $\phi_{it}(\cdot, \cdot)$. One can also obtain product-level market shares $s_{j\cdot}$ as a (weighted) sum of $s_{\boldsymbol{b}t}$ for bundles including product j: When only product-level sales are observed, the market share of product j is the weighted sum of market shares of all options including product j, denoted as $\mathbb{P}(j)$:

$$s_{j\cdot t} = \sum_{\boldsymbol{b} \in \mathbb{P}(j)} \omega_{j\boldsymbol{b}} s_{\boldsymbol{b}t}, \tag{6}$$

where weights $\omega_{j\boldsymbol{b}}$ are positive integers indicate the number of product j in the bundle $\boldsymbol{j}$, known or observed by the analyst. In my empirical specification, all weights are equal 1 when I observe a purchase incidence of two different products. Depending on the context, weights can have different values, but I assume it is fixed when calculating the product level market share $s_{j\cdot t}$.

2.2 Identification

Depending on the data availability, identification results may vary. With the consumer level data, the bundle market shares can be observed and it helps the point identification of my model. However, when only product level market shares are observed, I show that the consideration formation stage can be identified, but preferences in the utility specification are not in the absence of extra functional form assumptions.

2.2.1 Identification with Bundle Market Shares

Point identification when market shares of all options (single + bundle) is established under three key conditions: (i) observed market shares of bundle options, (ii) two-way exclusion restrictions, (ii) excluded instruments for price.

When the market shares of bundle options are either observed or identified from a large consumer sample, the bundle options can be treated as just like other options in the multinomial choice model conditioning on the consideration set. I can observe bundle options chosen by individuals from the consumer data. Hence, I am less concerned about the absence of market shares of bundle option in the product level data. Furthermore, consumer panel data also provide extra identification sources for nonlinear parameters in utility specification (Berry & Haile, 2020; Grieco et al., 2022).

The fundamental identification challenge in the consideration stage is to distinguish the two drivers of consumer's choice: either highest utility for the option compared to many others, or a small number of options considered even though the utility of the chosen one is not the highest among all feasible options. Therefore, without individual data on product awareness, identification must rely on functional form assumptions to disentangle preference and consideration. The most common identification strategy is the two-way exclusive restrictions, such as in Goeree (2008), who requires sufficient exogenous variations in excluded variables that only affect consideration but not the indirect utility, and vice versa. My empirical model of the inertia in consideration stage assumes the default specific consideration, based on household last yogurt purchases. The probability of the default consideration is mainly driven by the average level how, on average, consumers' choices respond to exogenous changes in characteristics of products not in the consideration set.

The price endogeneity in the BLP framework lies in the correlation between prices p_{jt} and market-product specific unobservables ξ_{jt}. I construct Hausman-type instruments (Hausman, 1996; Nevo, 2000) via average prices of same products in other geographic markets at the same time as the data. The intuition behind is that the price variations across other geographic markets are driven by changes in marginal costs that are supposed to be exogenous to the local demand shock ξ_{jt}. I also use some cost shifters from input data to provide additional exogenous variations for prices.

2.2.2 Identification with only Product Market Shares

Here I extend the point identification of consideration probabilities from Abaluck and Adams-Prassl (2021) to the case allowing for the choice of 2-goods bundle with linear pricing. I consider the same standard additive random utility models as the Eq.(4) in Abaluck and Adams-Prassl (2021); however, there are two main differences from their discrete choice framework, namely:

(i) The consideration set is formed at the single product level, while I allow multiple-product bundles in the choice stage. Therefore, the sum of latent product-level market share conditioning on the consideration set would exceed 1 with positive choice probabilities on bundles.

(ii) The absence of "nominal illusion" no longer holds, because a constant shift of one exogenous attribute of all products (single options) has a larger effect on the choice probabilities of bundle options than choice probabilities of single options.

In terms of (ii), the exogenous attribute is assumed to be the *price* in Abaluck and Compiani (2020), so "nominal illusion" in standard discrete choice model means that a constant price change of all options does not change the choice probabilities. With the presence of bundle option, one unit change of all prices in market t change $\delta_{\boldsymbol{b}t}$ by $\alpha \sum_{j \in \boldsymbol{b}} \omega_{j\boldsymbol{b}}$, which is greater than α, the change of δ_{jt}. Hence the individual choice probabilities $s_{i\boldsymbol{b}t}$ **do** change. Nonetheless, we can still exploit the asymmetry of cross partial derivatives of marginal market shares of products under limited condition to identify the consideration process as Abaluck and Adams-Prassl (2021). In the full consideration, the cross partial derivatives of market shares are still symmetric at the product level. Below I show the details

Symmetry in Latent Product Shares

I denote the full feasible set of inside alternatives as $\mathcal{J} = \{1, \dots, J\}$, the multi-product bundles $\boldsymbol{b} \in \{(j, j') : j \neq j', j, j' \in \mathcal{J}\}$[3], the consideration set $C \subseteq \mathcal{J}$ and all consideration sets that contain alternative j as $\mathbb{P}(j)$. I suppose outside option 0 is always considered and the default inside option is product 1. Let the continuous exogenous attribute that shifts both consideration and utility be $\mathbf{z} = (z_1, z_2, \cdot, z_J)$, which corresponds to the prices in Abaluck and Adams-Prassl (2021). In my empirical application, the advertising level works as a valid candidate for such a shifter, given the endogenous prices. Suppressing other observables, the marginal product-level market shares $\{s_{j\cdot}(\mathbf{z}), j \in \{0\} \cup \mathcal{J}\}$ and their cross partial derivatives $\{\partial s_{j'\cdot}(\mathbf{z})/\partial z_j, j, j' \in \{0\} \cup \mathcal{J}\}$ w.r.t. the exogenous shifter are assumed to be identified from the data.[4] The latent marginal product level market shares conditional on the consideration set C are denoted $s^*_{j\cdot}(\mathbf{z}|C)$.

We first show that the cross partial derivatives of latent marginal market shares are symmetric. Let θ_i denote the random coefficient for $\mathbf{z}$. I obtain the cross derivatives as follows[5]

$$
\begin{aligned}
\frac{\partial s^*_{j\cdot}}{\partial z_{j'}} &= \sum_{\boldsymbol{b} \in \mathbb{P}(j)} \frac{\partial s^*_{\boldsymbol{b}}}{\partial z_{j'}} \\
&= \sum_{\boldsymbol{b} \in \mathbb{P}(j), \boldsymbol{b} \neq (j,j')} \int \theta_i s^*_{i\boldsymbol{b}} \Big(\sum_{\mathbf{b}' \in \mathbb{P}(j')} s^*_{i\mathbf{b}'} \Big) dF_\theta - \int \theta_i s^*_{i(j,j')} \Big(1 - \sum_{\mathbf{b}' \in \mathbb{P}(j')} s^*_{i\mathbf{b}'} \Big) dF_\theta \\
&= \int \theta_i \Big[\Big(\sum_{\boldsymbol{b} \in \mathbb{P}(j)} s^*_{i\boldsymbol{b}} \Big) \Big(\sum_{\mathbf{b}' \in \mathbb{P}(j')} s^*_{i\mathbf{b}'} \Big) - s^*_{i(j,j')} \Big] dF_\theta \\
&= \int \theta_i [s^*_{ij\cdot} s^*_{ij'\cdot} - s^*_{i(j,j')}] dF_\theta
\end{aligned}
$$

The last term is exchangeable between j and j' as long as individuals have the same utility preference θ_i on the shifter $\mathbf{z}$ across products. Therefore we have the same symmetric results at the marginal product level as in Abaluck and Adams-Prassl (2021)

$$
\frac{\partial s^*_{j\cdot}(\mathbf{z})}{\partial z_{j'}} = \frac{\partial s^*_{j'\cdot}(\mathbf{z})}{\partial z_j}. \tag{7}
$$

[3]The size of bundle options are assumed to consist of two different products for simplicity.

[4]The marginal product share can be constructed either by aggregating observed bundle shares or measured directly in the way of Wang (2021).

[5]I suppress the consideration set C for expository brevity.

Point Identification of Hybrid Consideration Model

The observed product market shares and latent shares are linked by consideration probabilities

$$s_{j\cdot}(\mathbf{z}) = \sum_{C \in \mathbb{P}(j)} \pi_C(\mathbf{z}) s^{\star}_{j\cdot}(\mathbf{z} \mid C). \tag{8}$$

Adapting the Theorem 1 in Abaluck and Adams-Prassl (2021), under similar assumptions, we can conclude that if we observe asymmetric cross partial derivatives in product-level market shares, then the limited consideration occurs with a positive probability, i.e.

$$\frac{\partial s^*_{j\cdot}(\mathbf{z})}{\partial z_{j'}} \neq \frac{\partial s^*_{j'\cdot}(\mathbf{z})}{\partial z_j} \Longrightarrow \pi_{\mathcal{J}}(\mathbf{z}) < 1.$$

Following the identification construction procedure in Abaluck and Adams-Prassl (2021), we can replace the observed market shares in their setting with the marginal product-level market shares here. Hence we can achieve point identification on consideration probability without much modification. I sketch the proof using the current hybrid consideration formation, so that the observed marginal product-level market shares follow:

$$s_{j\cdot}(\mathbf{z}) = \begin{cases} 1 - \mu(z_1) + \mu(z_1) \sum_{C \in \mathbb{P}(0,1)} \Pi_{j \in C} \phi(z_j) \Pi_{j' \notin C} (1 - \phi(z_{j'}) s^*_{j\cdot}(\mathbf{z}|C), & j = 0, 1 \\ \mu(z_1) \sum_{C \in \mathbb{P}(0,1)} \Pi_{j \in C} \phi(z_j) \Pi_{j' \notin C} (1 - \phi(z_{j'}) s^*_{j\cdot}(\mathbf{z}|C), & j = 2 \ldots J. \end{cases} \tag{9}$$

Some algebra using the symmetry of cross partial derivatives of the latent market shares shows that for non-default alternatives ($j, j' > 1$), the difference between observed cross-partial derivatives is specified as:

$$\frac{\partial s_{j\cdot}(\mathbf{z})}{\partial p_{j'}} - \frac{\partial s_{j'}(\mathbf{z})}{\partial p_j} = \frac{\partial \log \phi_{j'\cdot}}{\partial z_{j'}} \left(s_{j\cdot}(\mathbf{z}) - s_{j\cdot} \left(\mathbf{z} \mid \mathcal{J}/j' \right) \right) - \frac{\partial \log \phi_j}{\partial z_j} \left(s_{j'\cdot}(\mathbf{z}) - s_{j'\cdot}(\mathbf{z} \mid \mathcal{J}/j) \right). \tag{10}$$

where J/j denotes the market where alternative j is not available. The difference between non-default option j and the default inside alternative then becomes

$$\frac{\partial s_{j\cdot}(\mathbf{z})}{\partial z_1} - \frac{\partial s_{1\cdot}(\mathbf{z})}{\partial z_j} = \frac{\partial \log (\mu_1)}{\partial z_1} s_{j\cdot}(\mathbf{z}) - \frac{\partial \log (\phi_j)}{\partial z_j} \left(s_{1\cdot}(\mathbf{z}) - s_{1\cdot} (\overline{\mathbf{z}}_j) \right), \tag{11}$$

where $\overline{\mathbf{z}}_j$ for the case when the $z_j \to +\infty$ Equations (10) and (11) consist of $\frac{1}{2}(J-1)(J-2)+(J-1)$

equations in total with J unknowns in the derivatives of consideration functions ϕ_j and μ. One can obtain a unique solution to J unknowns from the system of equations with observed marginal product-level shares and their cross partial derivatives under a full rank condition. Once the first-order change of consideration functions is identified, their levels can be recovered through the condition that $\mu(z_1) \to 1$ as $z_1 \to +\infty$ in DSC and $\phi_j \to 1$ as $z_j(z_1) \to +\infty$ in ASC.[6] Hence, the consideration formation process is point identified with observed marginal product-level market shares, and also the existence of an exogenous and continuous shifter. It is worth noting that the proof does not presume either (i) all marginal product shares summing up to 1 or (ii) the absence of nominal illusion.

Latent Market Shares

However, latent market shares of products and bundles are not non-parametrically point identified even if we can identify the consideration process. Relaxing the key assumption of absence of nominal illusion[7] in Abaluck and Adams-Prassl (2021), one can see that, at given level of $\mathbf{z}$ in Eq.(9), there are at least 2^{J-1} possible cases of consideration sets and for each consideration set with size l ($l >= 2$). Hence we need to identify at least $l+1$ product-level latent shares or even $\frac{1}{2}l(l-1)+(l+1)$ latent market shares of bundles. In contrast, we can observe at most $\frac{1}{2}J(J-1)+1$ market shares at bundle level and only $J+1$ marginal product-level shares. Without further restrictions in functional form of the utility, it is not clear how to uniquely pin down those latent shares in the constructive manner as Abaluck and Adams-Prassl (2021). I suppose it must rely on parameteric specifications in the random utility model to achieve point identification of preference primitives.

3 Data on Yogurt Markets

I apply my discrete choice model to household demand for yogurt using both store sale data and consumer purchase panel. In this section, I introduce the yogurt store sale data and consumer

[6]Details on identification proof can be found in Abaluck and Adams-Prassl (2021) and their online appendix.

[7]The absence of nominal can still hold among inside goods when the weights in every bundle option sum up 1. I leave the fully exploration for future research.

panel data, define the market and purchase options, and describe consumers' purchase behavior for the empirical setting.

3.1 Product Level Data

I obtain yogurt sales from the retail scanner data in the IRI academic datasets. The retail scanner data cover about sales of 30 consumer package goods categories at participated grocery stores in 50 geographic markets from 2001 to 2012. A detailed description of the IRI academic datasets can be found in Bronnenberg, Kruger, and Mela (2008). I select the retail scanner data at two largest grocery stores in Pittsfield(Massachusetts) from 2008 to 2012, mainly because of the availability and quality of the consumer panel that I will discuss later.

The store scanner data record weekly sales at the UPC (Universal Product Code) level. It contains information on price, sale units and some marketing variables like advertisement, display, and temporary price discount. The store scanner data are supplemented by item's attribute information on brand, manufacturer, equivalent volume, and other characteristics like style, fat content, flavor, calorie, and whether it is organic. The market is defined at the week-store level. I igone the store choice issuein household yogurt purchase decision. Given the small proportion of yogurt in the total expense of grocery shopping basket (4% on average), it is reasonable to assume that yogurt purchase decision usually does not drive household grocery store choice. Yogurts of national brands are available in both stores around the same time in most cases, with the exception for private label yogurts. There are no inventory data, so I could not observe either price of yogurt that had never been purchased in the market, nor the out-of-stock status of yogurt products any time during the week.

Given the large number of items (1,107 unique UPCs), I aggregate UPC items into the brand-attribute level as the alternative for consumer to choose, to make tractable the size of the choice set. The brand-attribute alternative is close to the definition of the product line in Triggs (2021). Regarding brands, I select major ones that satisfy a minimum sales share threshold: 10% of sale volume share in at least one market. For attributes, I use three categorical ones that do not vary across markets: whether it is Greek, whether it is organic, and one of three levels in fat content: fat-free, reduced-fat, and whole fat.[8] I further rule out brand-attribute alternatives that fail to reach 0.1% sales volume in any market. I will refer to the brand-attribute alternative as the "product" thereafter. Finally, I have 22 inside products in the store data at the brand-attribute level, and I combine the remaining yogurt items and no purchases into the "outside" option when calculating market shares. The discrete choice model abstracts away from the choice of quantity of yogurt in each purchase incidence. Hence market size is the total number of shopping trips at the grocery store in each week, and I approximate the market share of no purchases using the consumer panel data. As Table 1 shows, national major brands like Yoplait and Dannon offer most non-organic and non-Greek yogurts, while other popular brands like Chobani and Stonyfield concentrate on market segments of Greek and organic yogurt, respectively. Most popular products are either fat free or reduced fat in our sample, as opposed to a small fraction of whole fat yogurt. Table 1 also provides the entry and exit information

I also aggregate market-specific characteristics of products like prices, flavor variety, advertising, display and discount from the item level. Given the equivalent volume of each item, I aggregate the items' prices into price per pint by volume-based weights. The variety of flavors is calculated by the number of items' flavors within each product and normalized in each market between 0 and 1. The marketing variables like advertisement, display and discounts are first converted into numerical values and weighted by the number of items. The advertisement feature codes the size of promotion text and the retailer coupon or rebate information. I convert levels of features into one value from $(0, 0.25, 0.5, 0.75, 1)$ in which 1 means coupon, 0.75 for large ad size, 0.5 for medium size, 0.25 for small and 0 if none. In-store display information coded as 2 for major place like lobby or end-aisle, 1 for minor display and 0 if none, and I rescale it to $(1, 0.5, 0)$ respectively. Discount is a binary variable for whether a temporary price reduction is greater than 5% the regular price. Product availability varies moderately across markets, mainly because of the trend of Greek yogurt led by Chobani and followed by other brands later. I also observe brands like Breyers and Colombo

[8]With over 2% fat is considered whole fat, reduced fat is between 0.5% and 2%, and below 0.5% is fat free. Source: https://www.livestrong.com/article/444387-low-fat-yogurt-vs-no-fat-yogurt/.

Table 1: Market Shares of all (Inside) Yogurt Products

Brand	**Style**	**Fat**	**Organic**	**First week**	**Last week**	**Median market share**
Yoplait	Non-greek	Reduced	Non-organic	1/6/08	12/30/12	2.308%
Dannon	Non-greek	Free	Non-organic	1/6/08	12/30/12	1.883%
Dannon	Non-greek	Reduced	Non-organic	1/6/08	12/30/12	1.231%
Chobani	Greek	Free	Non-organic	2/3/08	12/30/12	1.130%
Private1	Non-greek	Free	Non-organic	1/6/08	12/30/12	0.742%
Private1	Non-greek	Reduced	Non-organic	1/6/08	12/30/12	0.703%
Yoplait	Non-greek	Free	Non-organic	1/6/08	12/30/12	0.680%
Stonyfield	Non-greek	Reduced	Organic	1/6/08	12/30/12	0.574%
Private1	Greek	Free	Non-organic	6/26/11	12/30/12	0.404%
Breyers	Non-greek	Reduced	Non-organic	1/6/08	11/13/11	0.398%
Stonyfield	Non-greek	Free	Organic	1/6/08	12/30/12	0.385%
Colombo	Non-greek	Reduced	Non-organic	1/6/08	1/24/10	0.330%
Colombo	Non-greek	Free	Non-organic	1/6/08	1/24/10	0.318%
Chobani	Greek	Reduced	Non-organic	5/25/08	12/30/12	0.295%
Stonyfield	Non-greek	Whole	Organic	1/6/08	12/30/12	0.261%
Dannon	Greek	Free	Non-organic	2/6/11	12/30/12	0.194%
Private2	Greek	Free	Non-organic	8/7/11	12/30/12	0.156%
Yoplait	Greek	Free	Non-organic	2/14/10	12/30/12	0.137%
Weight	Non-greek	Free	Non-organic	1/6/08	4/15/22	0.136%
Breyers	Non-greek	Free	Non-organic	1/6/08	11/20/11	0.105%
Dannon	Greek	Reduced	Non-organic	7/31/11	12/30/12	0.051%
Dannon	Non-greek	Whole	Non-organic	1/22/12	12/30/12	0.041%
Stonyfield	Greek	Free	Organic	6/26/11	12/30/12	0.031%

Note: Private1 denotes the private-label yogurt in Store 1, and Private2 denotes the private-label yogurt in Store 2. Median market shares are based on sales volume within the weeks with positive volume and adjusted for the no-purchase option.

exited the yogurt market during the sample period. Table 2 summarizes the store level data.

Table 2: Summary Statistics from Store Sales Data

	Mean	Median	Std.	1st Qt	3rd Qt	N
Markets						
Num. available products	13.97	14	1.42	13	15	492
Share of outside option	0.88	0.88	0.03	0.85	0.95	492
Products						
Price per pint	2.41	2.28	0.74	2	2.64	6,887
Advertisement	0.11	0	0.24	0	0	6,887
Display	0.01	0	0.07	0	0	6,887
Discount	0.17	0	0.28	0	0.23	6,887

Note: Share of outside option includes both no purchase and purchase of other brands.

To handle the price endogeneity, I need to excluded instrumental variables. I collect price information for the same yogurt products sold in the same grocery chain in other geographic markets from the IRI retail data. Then I construct Hausman-type instruments for prices via averaging yogurt prices in other markets. In addition, I collect data on the cost shifters like the hourly wage in the food production industry in the state of the brand's nearest manufacturer, industrial electricity price in the region of the corresponding manufacturer, and nationwide price on raw milk.[9] These cost shifters are recorded monthly or quarterly, and do not vary much across products in the same period. Therefore, most identification power comes from the variation in the yogurt prices in other geographic markets. I assume the price variable as the only endogenous product characteristics, so that other product attributes including advertising, variety of flavors, and product fixed effects are valid included instruments into the moment conditions.

3.2 Consumer Panel Data

Consumer panel is obtained from the household survey data in IRI's "behavioral" markets of two small cities called Pittsfield and Eau Claire. Pittsfield has more participating households over years

[9]Data on cost shifters are collected from https://fred.stlouisfed.org.

and reporting error was reduced after 2007. It contains information on dates, stores, purchased items and quantity, and the total expenditure of each grocery trip. The sample of households does change by year, but around 70% households stay in the survey across years. I also have demographic information on households in the consumer panel, and I use household income and family size as key demographic variables in household utility specification. Table 3 presents the summary statistics from the consumer panel. Household demographics include all participants who have ever buy yogurt once in the 492 markets.

Table 3: Summary Statistics on Consumer Panel

	Mean	Median	Std.	1st Qt	3rd Qt	N
Markets						
Num. of grocery trips	670	649	105	607	736	492
Num. of yogurt incidences	120	110	41	87	154	492
Demographics						
Annual income ($10,000)	7.03	7	3.27	5	10	2,519
Family size	2.41	2	1.25	2	3	2,519

Note: Household income was reported in income brackets in the unit of ten thousand dollars in the original data, and I uniformly draw a number from the household income bracket as household actual annual income to make it a continuous measure in the model estimation.

Size of Purchase Incidences

About 30% of purchase incidences involve multiple yogurt products which I define as the bundle option. I restrict the bundle composition to be 2 different products in the estimation, given that less than 5% shopping trips involving more than 2 yogurt products. Table 4 shows the distribution of purchase incidences of yogurt products from the consumer panel. The reduced fat and nonfat Yoplait non-Greek, non-organic yogurt products are the most popular bundle choice. I find that most co-purchases take place between reduced fat and fat free products that are the same otherwise. For major brands like Yoplait and Chobani, consumers tend to buy fat-free and reduced-fat yogurt together more often than fat-free yogurt of the same brand only.

I also find a positive correlation between the family size and the tendency to purchase the bundling option in Table 5. It is intuitive that the diversity in yogurt products in household shopping basket can satisfy the different preferences across family members. More household members

Table 4: Distribution of Top Bundle Options

Yogurt Incidence	Frequency	Percent	Cum. Pct.
Yoplait-NG-NO-RF	10,334	17.46%	17.46%
Dannon-NG-NO-FF	10,231	17.28%	34.74%
Dannon-NG-NO-RF	5,805	9.81%	44.54%
Yoplait-NG-NO-FF&RF	5,105	8.62%	53.16%
Chobani-G-NO-RF	3,944	6.66%	59.83%
Yoplait-NG-NO-FF	2,527	4.27%	64.09%
Private1-NG-NO-FF	2,111	3.57%	67.66%
Breyers-NG-NO-RF	1,851	3.13%	70.79%
Chobani-G-NO-FF&RF	1,688	2.85%	73.64%
Stonyfield-NG-O-FF	1,520	2.57%	76.21%
Private1-NG-NO-RF	1,459	2.46%	78.67%
Stonyfield-NG-O-RF	1,197	2.02%	80.69%
Dannon-NG-NO-FF&RF	1,175	1.98%	82.68%
Total	59,203	100%	100%

Note: I rank the purchase options based on the frequency from the consumer panel. I include options with more than 1,000 trips. For example, incidence "Yoplait NG-NO-FF&RF" means the bundle of Yoplait non-Greek non-organic fat-free yogurt and Yoplait non-Greek non-organic reduced-fat yogurt.

often imply a stronger preference for the variety in yogurt consumption at the household level. Hence, I consider the family size as an important demographic factor in determining the complementarity level in bundles.

Inertia and Product Switch

Although the consumer panel does not ask for what household considered before the purchase, my empirical specification for the consideration stage models the default consideration set to be last-purchased yogurt products plus the outside option. I refer the status of choosing from the default consideration as inertia. A substantial level of inertia would imply some degree of persistence in individual purchase history. I choose households with at least 2 shopping records in the consumer panel and find that around 60% of yogurt product choices come from household last-time purchases, and choices in nearly 40% of yogurt purchase incidences belong to the default set of their last purchased ones. The persistence in yogurt choices supports the general pattern found in Keane

Table 5: Bundle Size and Demographics

	Household average multiple purchase percent
Family size	0.029***
	(0.005)
Household income	-0.001
	(0.002)
(Intercept)	1.217***
	(0.015)
Number Obs.	2,519
R^2	0.015

*Note: This table reports the OLS regression results. The outcome variable is the average share of trips with multiple yogurt purchases for each household. Two regressors are household size and annual income. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.*

(2015) where the author finds substantial persistence in brand choice of consumer package goods[10] from most consumer panel data.

I also find in Table 6 that consumers are more likely to switch to a yogurt product different from last purchases when advertising, promotion, and price discount are active. In my empirical specification, I add product's advertising level into the product-market average utility.[11]

Representativeness of the Consumer Panel

According to the summary statistics of the consumer panel, the sample size of consumers with yogurt purchases varies across markets. On average, the yogurt purchases in the consumer panel represent about 6% store sales by total weighted volume. Since the consumer panel is crucial for identifying and estimating the demand synergy in bundle and the degree of inertia, markets with fewer purchase incidences are more prone to sampling errors. Therefore, I choose markets with enough consumers in the model estimation. The consumer sample needs to be also representative of the whole population in terms of the distribution of income and household size to ensure that there is no selection bias in panel participation based on demographics. Otherwise, the estimation

[10]Other factors like state dependence and individual preference heterogeneity can also contribute to the individual persistence, but those factors go beyond my current model.

[11]Admittedly, advertising can also affect the consideration formation. I leave for future research the estimation of a hybrid consideration model with bundle, allowing advertising in both utility and consideration.

Table 6: Linear Probability Model on Product Switch

	Whether current choice differs from last purchase
Advertising	0.024*
	(0.012)
Display	0.167***
	(0.020)
Discount	0.059***
	(0.011)
First week	0.523*
	(0.203)
Number Obs.	72,082

Note: The table reports the regression result of the linear probability model. The sample includes all yogurt products households purchase after their first yogurt shopping trips in the consumer panel. The outcome variable is binary and equals 1 if the product is the same as the last purchase, and 0 otherwise. First Week indicates whether the product was launched in the store that particular week. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

for nonlinear parameters on demographics would be biased.

4 Estimation

I first specify the empirical model as a special case in the general setup. Then the estimation strategy adapted from Grieco et al. (2022) is introduced. Next I show the estimation results comparing my model with the multinomial logit model and the standard BLP model. In the post estimation analysis, I decompose the diversion ratios into substitution and complementarity effects. For products that are often bundled, the complementarity dominates the substitution, reversing the sign of the diversion ratio.

4.1 Empirical Specification

In the empirical application to yogurt, I model the consideration stage as the default specific consideration[12], assuming that the consumer's default consideration set in each market consists of the outside option and all yogurts from his/her last purchases that are currently available at the current grocery store

[12]Estimating the hybrid consideration model implies more computational burden due to the simulation of consideration set.

$$\Pr(CS_{it}) = \mu_0 \mathbb{1}(CS_{it} = \boldsymbol{C}_{0it}) + (1-\mu_0)\mathbb{1}(CS_{it} = \boldsymbol{J}_t)$$

Hence, the consideration set has a probability μ_0 for the default consideration set and a (1-μ_0) probability for the full feasible choice set. I simplify the consideration set as the full feasible choice set in the case of no yogurt purchase history. Such cases occur in 3% of yogurt purchases incidences and should not significantly affect the estimation.

Regarding the utility specification, first, the average utility δ_{jt} of the single yogurt product j in market t, is a linear function of price, the level of advertising, the relative variety of flavors, the product j's fixed effect and unobserved market-specific demand shock ξ_{jt}, namely

$$\delta_{jt} = -\alpha \text{Price}_{jt} + \beta_1 \text{Advertising}_{jt} + \beta_2 \text{Flavor}_{jt} + \text{FE}_j + \xi_{jt}$$

Then I introduce household heterogeneity in price sensitivity as follows:

$$\eta_{ijt} = (\theta_1 \nu_i + \theta_2 \text{Income}_{it}) \times \text{Price}_{jt}.$$

Each household has a random taste shock ν_i on price, and price sensitivity also depends on the household income level. I expect a positive sign on θ_2 since consumers with a higher income are probably less price sensitive.

As for demand synergy, I model it as a linear function with a constant term κ_0 for the extra utility of buying a bundle of two products and a household-specific term depending on the family size. A positive κ_1 is expected, in line with the positive correlation between household size and the number of yogurt products in each shopping trip displayed in Table 5. This gives

$$u_{i\boldsymbol{b}t} = \sum_{j \in \boldsymbol{b}} (\delta_{jt} + \eta_{ijt}) + \underbrace{\kappa_0 + \kappa_1 \text{size}_i}_{\text{synergy}}.$$

I confine the feasible set of all bundle options to the collection of all observed bundles with at least total 100 occurrences in the consumer panel. It leaves me 26 possible bundling options from 22 yogurt products. For shopping trips with more than 2 yogurt products, I randomly select two from all purchased products as the chosen bundle option.[13]

4.2 Estimation Strategy

Standard BLP estimator with available consumer level data often imposes market shares constraints in estimation. A new estimation procedure recently proposed by Grieco et al. (2022) relaxes market shares constraints and allows for sampling errors in observed market shares from the data. It turns out to be a more efficient estimator than previous estimator which imposes share constraints such like micro moments (Petrin, 2002; Berry et al., 2004) and the two-stage estimator (Chintagunta & Dube, 2005; Goeree, 2008). The intuition for the efficiency improvement is that the first-order conditions for their objective function are equivalent to some underlying GMM objective with an efficient weighting matrix. In contrast, the share constraints impose infinite weights on some parts of first order conditions, which leads to an efficiency loss. The estimator can also achieve optimal convergence rates in different asymptotic settings regarding the number of consumer sample size in micro data, the market size in macro data and the number of products in moment conditions.

I extend their estimation approach to allow for the consideration stage before utility-maximizing decision and the weighted sum of bundle market share to (marginal) product market shares to match the store sale data. The objective function combines the mixed loglikelihood from the consumer panel and retail sale data, and generalized methods of moments. I denote the linear parameters in average utility δ as β, and the rest model primitives as θ. We have

$$(\hat{\beta}, \hat{\theta}, \hat{\delta}) = \operatorname{argmin}_{\beta,\theta,\delta} - \log L(\theta, \delta) + \Pi(\beta, \delta) \tag{12}$$

The mixed loglikelihood follows:

$$\log L(\theta, \delta) =$$
$$\underbrace{\sum_{t=1}^{T} \sum_{\boldsymbol{b}=0}^{J_{2t}} \sum_{i=1}^{I_t} D_{it} y_{i\boldsymbol{b}t} \log s_{ibt}(\theta, \delta)}_{\text{micro loglikelihood}} + \underbrace{\sum_{t=1}^{T} \sum_{j=0}^{J_t} \left(I_t \tilde{s}_{j\cdot t} - \sum_{i=1}^{I_t} D_{it} y_{i\boldsymbol{b}t} \mathbb{1}(j \in \boldsymbol{b}) \right) \log s_{j\cdot t}(\theta, \delta)}_{\text{macro loglikelihood}},$$

[13] However, I do allow more than 2 products in the default consideration set in which case the household bought more than 2 yogurt products last time.

where D_{it} is a binary variable on whether individual i in the panel sample, $y_{i\boldsymbol{b}t}$ - whether i chooses option $\boldsymbol{b}$, and $\left(I_t\tilde{s}_{j\cdot t} - \sum_{i=1}^{I_t} D_{it}y_{i\boldsymbol{b}t}\mathbb{1}(j \in \boldsymbol{b})\right)$ - j's total purchase incidences excluding those in consumer panel. Excluding consumer sample in macro likelihood ensures that purchase decisions from the consumer sample are not double counted in the loglikelihood.

The GMM part objective function is based on standard moment conditions from instruments specified as follows:

$$\Pi(\beta,\delta)=\frac{1}{2}m'(\beta,\delta)\mathcal{W}m(\beta,\delta),$$
$$m(\beta,\delta)=\sum_{t=1}^{T}\sum_{j=1}^{J_t} z_{jt}(\delta_{jt} - (\beta' x_{jt} - \alpha p_{jt} + FE_j)),$$

where $\mathcal{W}$ is the scaled efficient weighting matrix such that $J\mathcal{W}$ converges to the inverse of the variance of moments: $\mathbb{V}(z_{jt}\xi_{jt})$, where $J = \sum_t J_t$ is the total number of products across all markets.

4.3 Estimation Results

First, I report the point estimates for parameters of interest in Table 7. I compare the results of the multinomial logit regression, the standard BLP model with random coefficients and income effect on price sensitivity, as well as the above empirical model that features the default specific consideration and bundle options. The former two models are estimated from the "pyblp" python package introduced in Conlon and Gortmaker (2020) where the standard BLP results are obtained with optimal instruments. I estimate my model using the Mathematical Program with Equilibrium Constraint (MPEC) techniques in the optimization (Dubé, Fox, & Su, 2012) of the objective function Eq (12). The estimation requires a richer data structure as the model becomes more complex. Only market shares and product characteristics are needed in simple logit model, while the income distribution is necessary for income effect on price sensitivity. Estimating the demand synergy in bundle and the degree of default consideration does rely on a consumer panel for observed bundle purchases and consumer's shopping history. As I mentioned previously in Section 4.2, I subsample markets with at least 100 yogurt purchase records in the consumer panel to estimate my model, which finally leaves 286 markets.[14] The standard errors for my estimates are derived from the

[14] I also exclude markets with obvious measurement errors, for example, markets where product sales in retail data are less than the total volume from the consumer panel.

inverse of the numerical Hessian matrix of the objective function in Eq (12) in Grieco et al. (2022). I also report estimation results with the full sample of 492 markets using the simple logit and standard BLP model in Table A1 in the Appendix. The estimation results of simple logit and standard BLP do not change much between the subsample and the full sample, therefore the selected subsample of 286 markets could be considered as representative.

Table 7: Estimation Results of Parameters

	Logit	Standard BLP	Bundle + Defaults
Price $-\alpha$	-0.745***	-1.127***	-1.261***
	(0.0443)	(0.0714)	(0.0011)
Flavor β_1	0.761***	0.779***	0.419***
	(0.0393)	(0.0339)	(0.0019)
Feature β_2	0.503***	0.479***	0.0783**
	(0.0458)	(0.0465)	(0.0269)
Random coefficient θ_1		0	0
		0	0
Income on price θ_2		0.0957***	0.0091***
		(0.0079)	(1.16e-05)
Ave. synergy κ_0			7.515***
			(0.619)
Family size on synergy κ_1			0.0744***
			(0.00171)
Inertia μ_0			0.794***
			(3.98e-05)
Number of markets	286	286	286
Data	market	market + income	consumer + market

Note: The table reports the estimation of parameters in interest across three demand models: multinomial logit with only market share data, BLP with market share and income distribution data, and my model with consumer panel and market share data. Product fixed effects are estimated but not reported. Price endogeneity is accounted for in all three models. The first two columns are estimated by the "PyBLP" python package introduced in Conlon and Gortmaker (2020). The standard BLP result uses the optimal instrument option in the package. Standard errors in the last column reported are derived from the diagonal entries of the inverse of Hessian of the objective function in Eq.(10). Random coefficient θ_1 reaches the lower bound of 0. Hence no standard error is reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

I find a large and significant demand synergy within bundles, and such bundling effect increases with household size. Households on average have 80% chance to choose from their default consideration set, which suggests the important role of past shopping history on the current decision. In addition, nonlinear parameters on price are negligible, implying little heterogeneity in price sensitivity across households. I also find that advertising effect, albeit statistically significant, diminishes in the presence of the default consideration set and the bundle option.

4.4 Price Effects

I focus on measuring the diversion ratios to analyze price effect in the post estimation. The diversion ratio with respect to price reflects the proportion of consumers substitute product j with product j' when product j's price increases by a small amount, and is defined as follows in the standard discrete choice model:

$$D_{j'j} = -\frac{\partial s_{j'}}{\partial p_j} \Big/ \frac{\partial s_j}{\partial p_j}, \quad j = 0, \ldots, J. \tag{13}$$

A higher diversion ratio between two products indicates a tougher price competition, and the calculation of diversion ratios has been important for antitrust analysis in horizontal merger. A more thorough and detailed discussion on diversion ratios can be found in Conlon and Mortimer (2021). In my model, when bundling options are allowed, I replace the market share of product j, s_j, with the marginal product level market share of j, $s_{j\cdot}$.

I consider the effect of a small amount of price change in the best-selling yogurt, Yoplait non-Greek, non-organic, reduced-fat yogurt (Yoplait NG-NO-RF) on following three other popular yogurt products:

(i) Yoplait non-Greek non-organic fat-free yogurt (Yoplait NG-NO-FF)

(ii) Dannon non-Greek non-organic reduced-fat yogurt (Dannon NG-NO-RF)

(iii) Chobani Greek non-organic reduced-fat yogurt (Chobani G-NO-RF)

These three products are ranked in a descending order of chances of buying together with Yoplait NG-NO-RF in the consumer panel. The number of incidences of Yoplait NG-NO-RF and Yoplait NG-NO-FF bundle is highest among all bundle alternatives, and even exceeds the number of shopping trips with Yoplait NG-NO-FF alone. Yoplait NG-NO-RF and Dannon NG-NO-RF are sometimes purchased together, while the frequency of the Yoplait NG-NO-RF and Chobani G-NO-RF bundle is negligible so that I exclude it from the choice set of bundle options.

Table 8 reports the median(at the top) and mean(at the bottom) of diversion ratios of three products under different models across 286 markets in the subsample. The large demand synergy reverses the sign of price effects when two products are likely to be bundled, and I also find a larger magnitude of the diversion ratios in my model than the multinomial logit and the random coefficient model.[15]

Table 8: Diversion Ratios by Models and Products

	Diversion Ratios		
	Yoplait NG-NO-FF	Dannon NG-NO-RF	Chobani G-NO-RF
Logit	0.0102	0.0143	0.0031
	0.0132	0.0161	0.0036
Standard BLP	0.0111	0.0152	0.0032
	0.0137	0.017	0.0037
Bundle + Defaults	-0.103	-0.0856	0.015
	-0.13	-0.0898	0.0169

Note: This table reports the median (top) and mean (bottom) diversion ratios from Yoplait NG-NO-RF to three products: Yoplait NG-NO-FF, Dannon NG-NO-RF, and Chobani G-NO-RF in 286 markets across three models in Table 7. I report the standard diversion ratios in Eq(11) for the first two models in the upper panel and marginal diversion ratios with product-level market shares for my model in the lower panel.

The bundle option enables me to decompose the diversion ratios of marginal product shares into the substitution and complementarity of the price effect as follows:

$$-\frac{\partial s_{j'\cdot}/\partial p_j}{\partial s_{j\cdot}/\partial p_j} = \underbrace{-\frac{\sum_{b'\in\mathbb{P}(j')}\partial s_{b'}/\partial p_{j'}}{\sum_{b\in\mathbb{P}(j)}\partial s_b/\partial p_j}}_{\textbf{Substitution}} - \underbrace{\frac{\partial s_{(j,j')}/\partial p_j}{\sum_{b\in\mathbb{P}(j)}\partial s_b/\partial p_j}}_{\textbf{Complementarity}}. \tag{14}$$

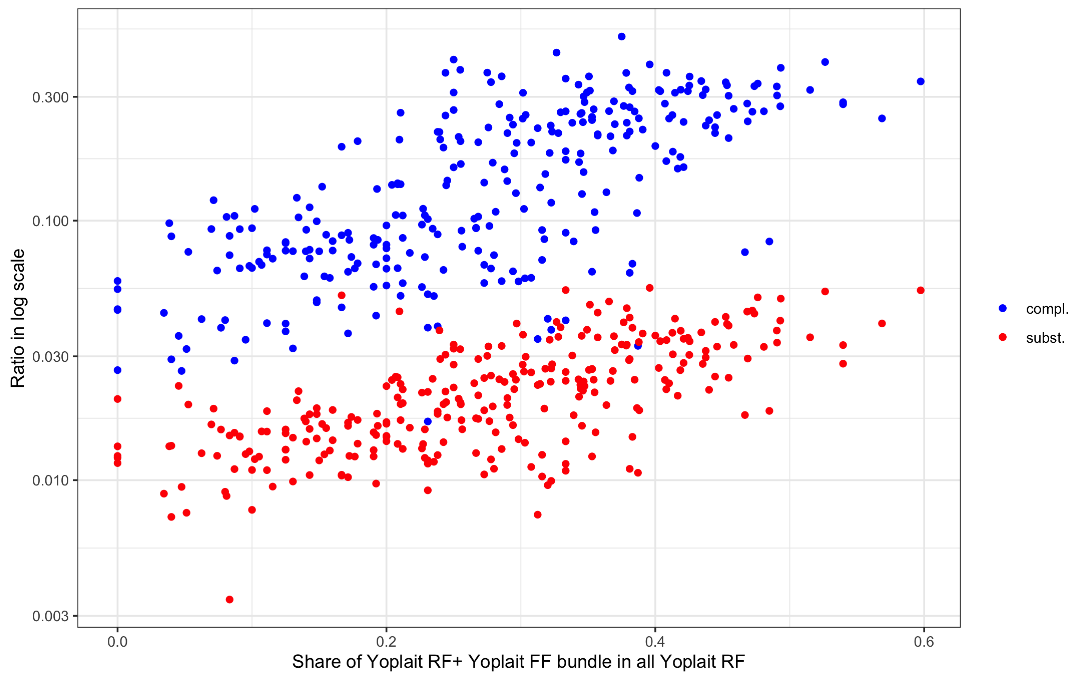
The substitution part represents the fraction of consumers who switch from any option including product j to any option including product j' but not the bundle of j and j', while the complementarity part represents the fraction of consumers who leave the bundle of j and j' as product j's price increases. With considerable demand synergy in bundle, one would expect a large

[15] I calculate the corresponding cross-price elasticities in Appendix Table A2. The pattern of cross-price elasticities remains the same with the trend of diversion ratios.

bundle market share of product j and j'. Therefore, complementarity dominates substitution when evaluating the diversion ratios in the product level.

Figure 1 displays the distribution of complementary and substitution effects against the observed share of Yoplait NG-NO-FF and Yoplait NG-NO-RF bundle among all Yoplait NG-NO-RF incidences across the sample markets.[16] I also find that in markets with relative more bundle purchases, the magnitudes of both effects increase.

Figure 1: Decompose diversion from Yoplait NG-NO-RF to Yoplait NG-NO-FF



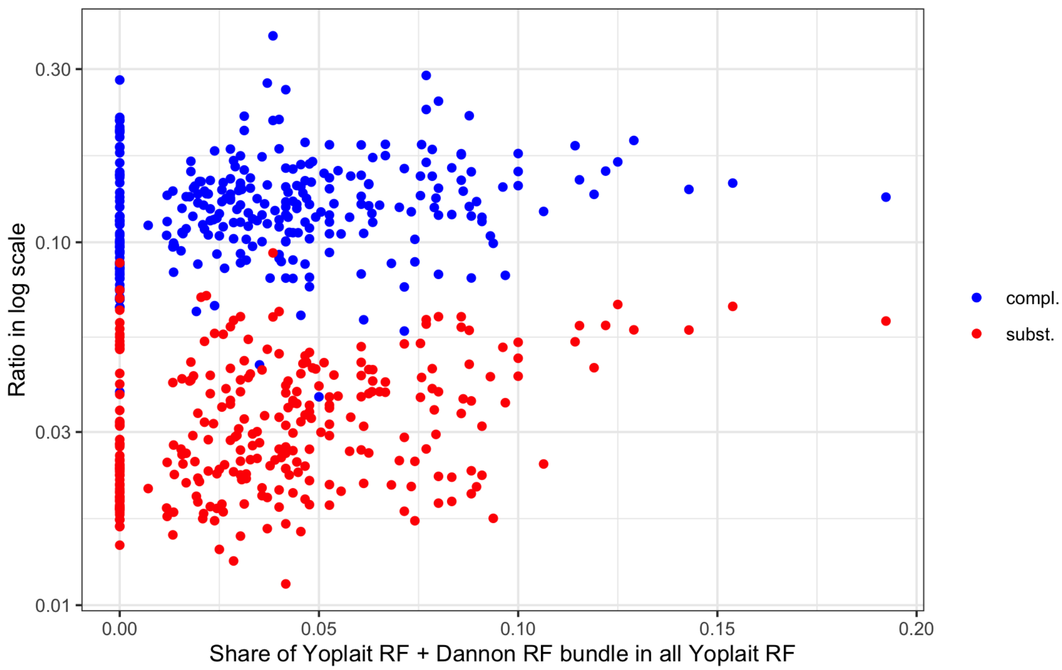
Note: This figure shows the decomposition in Eq (14) for the diversion ratio from product j Yoplait NG-NO-RF to product j' ***Yoplait NG-NO-FF*** *in each of 286 markets. The blue dot denotes the complementarity effect, and the red dot denotes the substitution effect. The x-axis measures the percentage of purchase incidences of the bundle out of total Yoplait NG-NO-RF purchase incidences in each market. I scale the y-axis by base ten logarithm.*

As for Dannon NG-NO-RF, the level of complementarity still outweighs the substitution effect, to a less degree though, as shown in Figure 2. Figure 2 also indicates that the substitution effects between Dannon NG-NO-RF and Yoplait NG-NO-RF are greater than that between Yoplait NG-NO-FF and Yoplait NG-NO-RF. It is in line with the intuition that many consumers have strong

[16] The cluster around zero y-axis reflects markets that have no observed bundle purchases. Nonetheless, my model can still predict price effects in those markets.

preference for reduced-fat yogurt, so they are more likely to switch to the reduced-fat of another brand, rather than to a yogurt product with different fat content from the same brand. In the case of Chobani G-NO-RF, only substitution effect exists due to the exclusion of its bundling option with Yoplait NG-NO-RF.

Figure 2: Decompose diversion from Yoplait NG-NO-RF to Dannon NG-NO-RF



Note: This figure shows the decomposition in Eq (14) for the diversion ratio from product j Yoplait NG-NO-RF to product j' ***Dannon NG-NO-RF*** *in each of 286 markets. The blue dot denotes the complementarity effect, and the red dot denotes the substitution effect. The x-axis measures the percentage of purchase incidences of the bundle out of total Yoplait NG-NO-RF purchase incidences in each market. I scale the y-axis by base ten logarithm.*

5 Conclusion

This paper develops a new discrete choice model for differentiated products that combines the bundle option and the limited consideration set. The motivation for this new model is to relax (i) the single discrete option to choose, and (ii) the full consideration of all feasible alternatives in the standard discrete choice models. My model allows both demand synergy in consumers' utility when choosing multiple products as a bundle and a consideration formation process before the decision. I discuss the identification of my model under different conditions of data availability and model restrictions. I extend the identification result of Abaluck and Adams-Prassl (2021) by showing the identification of the consideration process with product-level data only and in the absence of two-way exclusion restrictions.

I apply my model to estimate consumer demand for yogurt, using both store sales and consumer panel data. I adapt a novel estimation procedure that makes full use of both micro and macro data for a more efficient estimator. The estimation results suggest a considerable demand synergy within bundle options and a substantial effect of previous yogurt purchases on the consideration set. Taking both features into account profoundly reshapes the micro foundation for price competition analysis. For products often bundled, the complementarity driven by a large demand synergy dominates the substitution effect when the price of one product changes. It leads to the opposite sign of diversion ratios, compared to the estimation result from the standard discrete choice model that rules out the complementarity. My results emphasize the critical role of consumer-level data in demand estimation, and a better understanding of consumer behavior is always crucial for market-level analysis.

There are several directions for future research. First, some open questions remain regarding the identification of my current model under weaker data requirements or model restrictions. Does the two-way exclusion restriction help identify the demand synergy in the bundle option when the researcher can only observe product-level market shares? Even with data on individual choices or bundle shares, what identification result can be achieved if the consideration stage is not fully parameterized? Second, one needs to overcome computational challenges to estimating the hybrid model of consideration formation model. Optimizing over all parameters of interest, along with the mean utility of every product-market pair in one step, can be further complicated by the simulation procedure of individual latent consideration sets. Finally, it is necessary to complete the current framework with a supply-side model to study firms' pricing and advertising strategies. I hope my model can serve as the first step to answer important counterfactual questions in empirical settings featuring multi-product demand and limited consideration.

References

Abaluck, J., & Adams-Prassl, A. (2021). What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses*. *The Quarterly Journal of Economics*, *136*(3), 1611-1663.

Abaluck, J., & Compiani, G. (2020). *A method to estimate discrete choice models that is robust to consumer search* (Tech. Rep.). National Bureau of Economic Research.

Ackerberg, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 316–333.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, *44*(3), 1007–1040.

Allen, R., & Rehbeck, J. (2019). Identification with additively separable heterogeneity. *Econometrica*, *87*(3), 1021–1054.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, *forthcoming*.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International journal of Research in Marketing*, *12*(1), 9–24.

Berry, S., & Haile, P. (2020). Nonparametric identification of differentiated products demand using micro data. *Working Paper*.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of Political Economy*, *112*(1), 68–105.

Bronnenberg, B. J., Kruger, M., & Mela, C. F. (2008). The IRI academic dataset. *Marketing Science*, *27*(4), 745–748.

Cattaneo, M., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, *128*(7), 2796–2836.

Chintagunta, P. K., & Dube, J.-P. (2005). Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data. *Journal of Marketing Research*, *42*(3), 368–379.

Conlon, C., & Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics*, *51*(4), 1108–1161.

Conlon, C., & Mortimer, J. H. (2021). Empirical properties of diversion ratios. *The RAND Journal of Economics*, *52*(4), 693–726.

Crawford, G. S., Griffith, R., & Iaria, A. (2021). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, *222*(1), 4–43.

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

Dubé, J.-P., Fox, J. T., & Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, *80*(5), 2231–2267.

Dubé, J.-P. (2019). Chapter 1 - microeconometric models of consumer demand. In J.-P. Dubé & P. E. Rossi (Eds.), *Handbook of the economics of marketing, volume 1* (Vol. 1, p. 1-68). North-Holland. doi: https://doi.org/10.1016/bs.hem.2019.04.001

Fox, J. T., & Lazzati, N. (2017). A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, *8*(3), 1021–1036.

Gentzkow, M. (2007, June). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, *97*(3), 713-744.

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

Grieco, P. L., Murry, C., Pinkse, J., & Sagl, S. (2022). *Conformant and efficient estimation of discrete choice demand models.* (Unpublished Manuscript)

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods* (pp. 207–248). University of Chicago Press.

Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2021). Inattention and switching costs as sources of inertia in medicare part d. *American Economic Review*, *111*(9), 2737–81.

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

Ho, K., Hogan, J., & Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *The RAND Journal of Economics*, *48*(4), 877–905.

Hortaçsu, A., Madanizadeh, S. A., & Puller, S. L. (2017). Power to choose? an analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, *9*(4), 192–226.

Houthakker, H. S. (1961). The present state of consumption theory. *Econometrica: Journal of the Econometric Society*, 704–740.

Iaria, A., & Wang, A. (2020). Identification and estimation of demand for bundles. *Available at SSRN 3458543*.

Keane, M. (2015). Panel data discrete choice models of consumer demand. In B. Baltagi (Ed.), *The oxford handbook of panel data.* Oxford University Press.

Liu, C. (2019). *Essays on greek yogurt in the us market* (Unpublished doctoral dissertation). Northwestern University.

Lu, Z. (2021). Estimating multinomial choice models with unobserved choice sets. *Journal of Econometrics*.

Mehta, N., & Ma, Y. (2012). A multicategory model of consumers' purchase incidence, quantity, and brand choice decisions: Methodological issues and implications on promotional decisions. *Journal of Marketing Research*, *49*(4), 435–451.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, *9*(4), 513–548.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, *110*(4), 705–729.

Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, *28*(4), 429–440.

Rossetti, J. A. (2018). *Product variety in the us yogurt industry* (Unpublished doctoral dissertation). The Ohio State University.

Thomassen, Ø., Smith, H., Seiler, S., & Schiraldi, P. (2017). Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, *107*(8), 2308–2351.

Triggs, T. (2021). *Essays in industrial organization* (Unpublished doctoral dissertation). University of Michigan.

Van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., & Franses, P. H. (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, *47*(1), 63–74.

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

Wang, A. (2021). *A blp demand model of product-level market shares with complementarity* (Unpublished doctoral dissertation). Center for Research in Economics and Statistics (CREST).

Zhang, J. (2006). An integrated choice model incorporating alternative mechanisms for consumers' reactions to in-store display and feature advertising. *Marketing Science*, *25*(3), 278–290.

APPENDIX

A Appendix Tables and Figures

Table A1: Estimation of Parameters with all Markets

	Logit	Standard BLP
Price $-\alpha$	-0.79***	-1.2***
	(0.062)	(0.043)
Flavor β_1	0.67***	0.75***
	(0.046)	(0.028)
Feature β_2	0.38***	0.43***
	(0.064)	(0.042)
Random coefficient θ_1		0
Income on price θ_2		0.09*
		(0.0037)
Num. of markets	492	492

Note: This table reports the estimation results of multinomial logit and standard BLP. I follow the same estimation procedure as Table 7 with the full 492 markets. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A2: Cross-price Elasticities across Models and Products

	Elasticities		
	Yoplait NG-NO-FF	Dannon NG-NO-RF	Chobani G-NO-RF
Logit	0.0481	0.0481	0.0448
	0.0499	0.0499	0.0458
Standard BLP	0.055	0.054	0.0494
	0.0577	0.0572	0.0517
Bundle + Defaults	-0.692	-0.434	0.187
	-0.712	-0.464	0.191

Note: This table reports the median(top) and mean(bottom) cross-price elasticities from the price change of Yoplait NG-NO-RF to the market share of three products: Yoplait NG-NO-FF, Dannon NG-NO-RF and Chobani G-NO-RF in 286 markets across three models in Table 7. I report the standard cross-price elasticities for first two models in the upper region, and cross-price elasticities at the product level market shares for my model in the lower region.