A Pure Characteristics Approach for Evaluating Welfare Effects from Introducing New Products with Options*

Link to the Latest Version

Yiran Gong†

October 7, 2023

Abstract

Structural estimation of welfare effects from new goods largely relies on demand models that include a logit error and the simulation of a counterfactual that removes the new goods. In markets where consumers have both brand preferences and product-type (option) preferences, nested demand models come into play. However, the logit errors in nested logit models not only lead to an inaccurate prediction of shares in the counterfactual but also make welfare estimates sensitive to the number of nests. To deal with these two problems which give rise to implausibly large welfare estimates, I develop an empirical framework to estimate a pure characteristics demand model that allows for option (nest) choices and eliminates the need to rely on logit errors. I then provide an application using scanner data to quantify the welfare effects of four new products simultaneously introduced to the U.S. shampoo market where consumers have strong preferences over options. Results show that consumer welfare increases the most from the shampoo product that offers a niche option. Compared with models featuring logit errors, my approach reduces welfare overestimation by at least 73%. It also provides a better fit and more accurate welfare estimates than the pure characteristics model ignoring options.

Keywords: New product introduction, Demand estimation, Consumer welfare.

*I thank my advisers Robert Clark and Nahim Zahur for their comments and support, and Jean-François Houde, James G. MacKinnon, Eric Richert, and Oleksandr Shcherbakov for helpful advice. I would especially like to thank Christopher Conlon and Scott Orr for their discussions. I am also grateful for comments from conference participants in the Rising Stars Session at IIOC 2023 and in the IO session at CEA 2023.

†Department of Economics, Queen's University, Kingston, Canada, yiran.gong@queensu.ca

1 Introduction

New goods significantly contribute to economic growth by creating new industries or by improving product quality systematically. In many mature markets, firms constantly launch new products as well to stimulate consumers' interests. There may also be large welfare gains from new goods in these mature industries due to greater product variety. Hence, a careful welfare evaluation of new products is crucial to understanding their variety effect and more importantly whether they substantially raise standards of living.

The workhorse approach to quantify the welfare change is to first estimate a flexible demand model and then predict a counterfactual scenario in which the new products are not present to compute the difference in welfare with and without the new products. For modeling demand, most papers assume a logit error (e.g., Petrin (2002), Nevo (2003), etc.) for computational ease because it makes the model tractable.[1] Such a consumer-product specific error, however, is problematic when predicting the counterfactual without the new products because, with each new product, some consumer always gets a logit draw that predicts a higher utility than other existing products. Even if the new product is identical to some existing product, consumer welfare still increases after the new product is added. Consequently, welfare estimates tend to be implausibly large. Petrin (2002), for example, finds that the welfare increases from minivans are dominated by the logit error but not by the characteristics of minivans. In an attempt to address this, Ackerberg and Rysman (2005) penalize the logit error by adding a function of the number of products to the model, although their approach does not entirely remove the effect from the error. Another approach is to eliminate the error altogether by estimating a pure characteristics model, as proposed by Berry and Pakes (2007). Song (2007), for instance, uses this approach to evaluate new-product introduction and finds that consumer welfare from new PC products is ten times

[1]The AIDS (Almost Ideal Demand System) model is also used in previous literature on demand estimation (e.g. Deaton and Muellbauer (1980) and Hausman (1996)). In contrast to the characteristic-space approach, a product is considered an integrated entity in the AIDS model. Consequently, welfare estimates from such a model also tend to be too large because it fails to generate realistic substitution patterns.

higher when using the random coefficients logit model than the pure characteristics model.

An additional problem when using models with logit errors arises in markets where consumers have preferences for both brands and product types (or *options*). Examples include many consumer packaged goods (CPG) categories. For instance, when shopping for laundry detergent, consumers can choose brands such as Tide and Purex, and also choose detergent options such as liquid and powder detergent. Data on households' shopping trips, which I introduce later, show that consumers tend to repeatedly choose not only the same brands but also the same options. Some consumers even care more about options than they do about brands, such that for detergent their preference ordering could be (1) Tide liquid, (2) Purex liquid, (3) Tide powder, then (4) Purex powder. While brand preferences are well understood in the literature (e.g., Nevo (2001), Erdem et al. (2008), Borkovsky et al. (2017), etc), tastes for options are also key to characterizing consumers' product choices and to thinking about the impact of new product introduction. The existing approach treats options as nests (e.g., the nested logit (NL) model in Berry (1994) and Cardell (1997) or the random coefficients nested logit (RCNL) model in Brenkers and Verboven (2006)), and includes another logit-type error for the nests. This leads to an additional problem: there can be many nests and researchers have to decide on the number of nests to use in the demand system when the true number is unknown. When new products are introduced and product variety increases, it becomes harder to determine the correct number of nests. Due to the additional logit error for the nests, choosing different numbers of nests leads to different welfare estimates.

The objective of this paper is to develop an approach to simultaneously address these two problems that arise when evaluating the introduction of new products in a world where consumers have preferences over brands and options. The approach utilizes the pure characteristics demand model that removes logit errors, but unlike existing models, incorporates a nesting structure, allowing consumers to simultaneously choose both a brand and an option offered by the brand. Although the model must still specify the options (the nests), welfare estimates are no longer sensitive to the number of options since the nested logit error is removed. Welfare estimates from this model directly reflect consumers' valuations of brand and option choices, which are treated as discrete characteristics. Even without the consumer-product specific logit error, the pure characteristics model is still flexible enough to capture heterogeneity in consumer choices. An unobserved characteristic is assumed in the model to reflect the product-level shock. The taste parameters are allowed to depend on observed and unobserved consumer demographics and they vary over market/time.

My approach involves a two-step identification and estimation procedure, which is similar to Bajari and Benkard (2005). The main difference between this paper and theirs is that while they only propose the option case in theory and never implement it due to practical difficulty, I develop a full empirical framework with new identification and estimation strategies, which makes this paper the first empirical application of pure characteristics models capturing option preferences. The new estimation strategies can improve estimation efficiency and provide a better model fit.

The key to the first step is to recover the unobserved product characteristic, which is challenging with the presence of many discrete characteristics as covariates in the estimation of a conditional distribution. My approach achieves identification using the average price of all options offered by a brand. In contrast, Bajari and Benkard (2005) rely on the assignment of a baseline option, but this is not necessary using my approach. This is appealing since in many markets it is unclear which option one should choose as the baseline (e.g., for detergent it is not clear whether it should be liquid or powder). It also avoids having to predict the price of a baseline product in cases where one is not offered by the brand at all. I also provide new results on an over-identification test of the unobserved characteristic. This test can justify (or falsify) the application of the model developed in this paper to other markets where consumers have preferences over options. In order to deal with estimation challenges arising from the many discrete covariates, I adopt a new method developed in the machine learning literature by Athey et al. (2019), the Generalized Random Forests. Bajari and Benkard (2005) ignore all discrete characteristics since their first-step estimation would be very time-consuming when incorporating these characteristics. But dummy variables like brand, option, and market/time fixed effects that are crucial to capture consumer preferences are indeed discrete characteristics.

The second step of my approach is to back out the taste parameters. Parameters for continuous characteristics are calculated using the first-order conditions of consumers' utility maximization problems. For discrete characteristics, I estimate the parameters through multivariate probit models with consumer demographics. Using these structural estimates, I then simulate a counterfactual where the new products are removed from the choice set. The welfare change is measured by compensating variation when consumers face the simulated choice set instead of the actual choice set that includes the new products.

The application utilizes scanner data from Information Resources, Inc. (IRI). I estimate demand in the U.S. shampoo market and quantify welfare changes from four new products introduced concurrently in June 2006 by L'Oréal USA, one of the major firms. The shampoo market is well suited for studying the question because (i) firms are very active in introducing new products to the market and (ii) consumers usually choose specific shampoo options designed for certain hair types. I use both the market-level data on sales, prices, and promotional variables and the micro-level data on households' shopping trips and demographics.

It is worth highlighting several results in terms of model fit. First, pure characteristics models with option choices provide a better fit than those without, e.g., Bajari and Benkard (2005) provide an application of the model without options to the personal computer market. This is consistent with the descriptive pattern of the data that option choices matter to consumers. Models ignoring options fail to predict the substitution pattern at the option level. Second, for pure characteristics models with options, the approach developed in this paper can further improve the model fit from the approach proposed in theory by Bajari and Benkard (2005). Such an improvement comes from the first-step identification and estimation, where price information on all options is incorporated and a fast and efficient method is used to handle the discrete characteristics. Finally, when predicting market shares in the counterfactual without the new products, my model can fit the pre-introduction data better than models with logit errors, implying that the welfare estimate from my approach is more reasonable.

Regarding the welfare estimates, I find that the four new products create a total welfare gain of $86 thousand within six months of their introduction from total sales of around 155 thousand bottles (1 bottle = 16oz) in two cities. The average welfare gain is $0.56 per bottle or 11.64% of the average price paid by consumers. The product with the largest sales volume offers a relative niche option. It creates the largest welfare gain of $0.62 per bottle, which is equivalent to 12.83% of the average price.

There are two key findings concerning welfare comparison. First, I find that my model lowers the welfare estimates by at least 73% compared with models with logit errors, which is in line with Song (2007). Even the sophisticated RCNL with a penalty term estimates a welfare gain equal to 42.64% of the average price. It is implausible that these new products generate such a large welfare impact when the market already had a large number of brands before the introduction. Second, using brand-level data that ignore option choices generally exaggerates the welfare gain from the new products. The pure characteristics model with options, for instance, generates a 36% smaller estimate than the pure characteristics model without options. Intuitively, this is because models without options do not take into account the fact that there are already close substitutes for the new products at the option level.

This paper is closely related to the literature on the welfare evaluation of new goods, especially from a nested demand system. Existing papers either utilize models with logit errors (e.g., Petrin (2002), Nevo (2003), Berry et al. (2004), Giacomo (2008), Choi et al. (2013), etc.) or follow the pure characteristics model approach without modeling the nesting structure (e.g., Song (2007)). I contribute to the literature by developing a new empirical framework to estimate the pure characteristics model with options. My approach generates a careful welfare evaluation, which can be important for assessing firms' competitive responses though this paper omits the supply side for simplicity. Once incorporating the supply side, my approach can be useful to see whether firms provide innovation that improves consumers' standard of living or they just offer similar products to crowd out rivals.

Another related literature is the estimation of pure characteristics models. Berry and Pakes (2007) propose an algorithm similar to that of Berry et al. (1995), and their framework has been applied in empirical settings such as Song (2007), Song (2010) and Agarwal (2015). Bajari and Benkard (2005) develop an alternative approach with a hedonic pricing function, and Bajari and Kahn (2005) use this approach to estimate housing demand. Yet another approach is to rewrite the model as mathematical programming with complementarity constraints (e.g., Pang et al. (2015), Sun et al. (2017), Jiang and Chen (2021), etc.). This paper adds option choices to the pure characteristics model, which can be applied to a wide range of markets where consumers care about both brands and options. A potential drawback of pure characteristics models is that price elasticities can be large when products are assumed to be close substitutes and a large number of products are included in the demand system. This issue can be tackled by modeling consumers' consideration sets or search frictions (e.g., Goeree (2008), Barseghyan et al. (2021), Joo (2023), Morozov (2023), etc.). Although this paper does not directly deal with consideration sets, adding options to the pure characteristics models can mitigate the problem. Products that differ in options are not so close substitutes as those from the same option. And consumers only need information on the different brands and options, which allows their consideration set to have a moderate size.

The rest of the paper is organized as follows. Section 2 presents the data and descriptive evidence of consumer choice. In Section 3, I outline the model and identification. Section 4 presents the estimation strategy. Section 5 applies the model to the shampoo market and compares the welfare estimates. Section 6 concludes the paper.

2 Data and Descriptive Evidence of Consumer Choice

2.1 Data

The data come from the IRI scanner data set (see Bronnenberg et al. (2008) for a detailed description). I use both market-level and micro-level data. The market-level data contain sales volume, dollar sales, and promotional variables from a sample of over 1,500 grocery stores in 50 U.S. geographic units during 2006. Each geographic unit is defined by IRI as an agglomeration of counties, usually covering a major metropolitan area. The micro-level data include households' shopping trips in two geographic units, Eau Claire and Pittsfield. Each city has a sample of around 1,500 households with information on their demographics.

The original data are collected by UPC (Universal Product Code), store, and week. Focusing on women's shampoo products, I aggregate the data from the UPC level to the product level. Each product is defined as a brand-size-option combination. Brands are defined by product lines, which often have different brand names, e.g., L'Oréal Vive Pro, Pantene, Suave Naturals, etc. I follow the literature (e.g., Hendel and Nevo (2006), Miller and Weinberg (2017)) to treat different sizes of the same brand-option combination differently because of the nonlinear pricing in sizes. Options are based on the benefits and designated hair types of shampoo products in the data set. For example, option *Color* represents shampoo products designed for colored hair types and offers the benefit of color defense. I choose five options with the most UPCs, namely *Regular*, *Volumizing*, *Moisturizing*, *Clarifying*, and *Color*. All remaining types are grouped into option *Other*. Thus, one example product in the data can be L'Oréal Vive Pro-13oz-*Color*.

I aggregate observations from the store-week level to the market-quarter level, following Miller and Weinberg (2017). Each market is a geographic unit defined by IRI. Aggregation to the quarter level mainly reduces data frequency and thereby mitigates the effect of consumers' stockpiling behavior. I define prices as the ratio of total dollar sales over total

Table 1: Allocation of Inside Products by Firms and Options

	Number of products					
	Clarifying	*Color*	*Moisturizing*	*Other*	*Regular*	*Volumizing*
Firms						
Alberto Culver	2	1	2	2	2	3
Diamond Products Inc.	1	0	0	1	2	2
L'Oréal USA	0	2	0	3	3	2
Procter & Gamble	3	4	4	4	2	5
Unilever	4	2	6	3	3	2
Total	10	9	12	13	12	14

Table 2: Summary Statistics of Inside Shampoo Products in 2006

Variable	Product level		Brand level	
	Mean	S.D.	Mean	S.D.
Share (%)	0.70	0.85	2.80	2.85
Price ($/16oz)	3.09	1.66	3.19	1.79
Display (units)	11	22	46	70
Price-reduction (units)	73	107	295	442

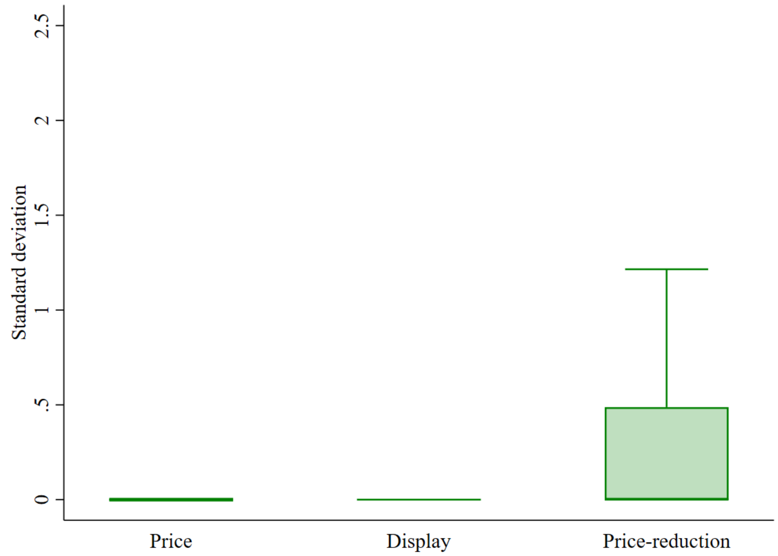
Note: This table reports the inside products' mean and standard deviation of shares and marketing variables across all market and quarter combinations in 2006.

sales volume measured in normalized bottles (1 bottle = 16oz). I also define promotional variables such as display and price-reduction as the number of times that the product has a particular promotion during the aggregated period in each market. To maintain a manageable number of inside products, I include major products that have at least 0.2% of total sales volume. This threshold is a bit smaller than previous literature such as Dubé (2005) because my analysis is at the brand-option level rather than the brand level.

In total, I have 70 inside products (including the four new products) from 17 brands owned by five major firms.[2] Each brand has on average 3.29 options, with a minimum of one option and a maximum of five. The allocation of inside products by firms and options is shown in Table 1. The most common option is *Volumizing*, which has 14 products from the five firms. *Color* is the least common option with 9 products. For each brand, the variation in sizes is much smaller than that in options. The average number of different sizes per brand is 1.71. Table 2 displays summary statistics of the inside products. At the product level, the average product occupies 0.70% share of the market, is priced at $3.09 per bottle,

[2] A list of all inside products is provided in Appendix A.

Figure 1: Standard Deviations of Marketing Variables for Brand-sizes



Note: For each brand-size combination, the standard deviation of each marketing variable is calculated across all options in each store/week. Outside values are excluded from the box plots.

and offers 11 units of display and 73 units of price-reduction. Aggregated to the brand level, the averages are all higher. More importantly, the standard deviations are higher at the brand level than at the product level. This is not surprising because, within each brand, different options of the same size often have the same price and are on promotion at the same time. This can be seen more clearly in Figure 1, which shows the box plots of the standard deviations calculated for each brand-size combination in each store/week. At least 75% of the inside brand-size combinations have zero standard deviation in price and display, and at least 50% of them have zero standard deviation in price-reduction.

2.2 Descriptive Evidence of Consumer Choice

To characterize patterns in consumer preferences for brands and options, I present two sets of descriptive evidence. The first set aims to show whether consumers care more about options or brands. A simple nested logit model might be too restricted because it requires a pre-determined nesting characteristic, either option or brand. Hence, I adopt a generalized nested logit model proposed by Fosgerau et al. (2022), the inverse product differentiation logit (IPDL) model, which allows for more than one nesting characteristic. The model has a linear estimating equation which is given by

$$\log(s_{jt}) - \log(s_{0t}) = \alpha p_{jt} + \boldsymbol{X}_{jt}\boldsymbol{\beta} + \sum_{d=1}^{2} \mu_d \log(s_{jdt}) + \xi_{jt}, \tag{1}$$

where s_{jt}, p_{jt}, $\boldsymbol{X}_{jt}$, and ξ_{jt} denote product j's market share, price, observed characteristics, and unobserved characteristic in market/time t, respectively, and s_{0t} denotes the outside option's share in t. There are two nesting characteristics (indexed by d), namely brand and option, and s_{jdt} denotes product j's within-nest share in t according to the nesting characteristic d. μ_1 and μ_2 are the two nesting parameters, with $\mu_1 + \mu_2 < 1$ and $\mu_d \geq 0, d = 1, 2$. Interpretation is similar to the nested logit model: a higher μ_d means that products in the same nest according to d are more similar. I use the market-level data to estimate the model. Display, price-reduction, option, brand, and market/time fixed effects are included as observed characteristics, and the endogenous price and within-nest shares are instrumented.[3]

With estimated parameters, I then calculate the diversion ratios for two groups of products. The first group represents two products from the same option but different brands; the second represents two products from the same brand but different options. The diversion ratio from product j to product k measures the fraction of consumers leaving product j and switching to product k following an increase in product j's price, which offers a better description of substitution patterns than cross-product price elasticities (Conlon & Mortimer, 2021). This exercise is also reproduced using data from other CPG categories such as laundry detergent, milk, and sugar substitute.[4] Table 3 reports the average diversion ratios across all pairs of products and all markets for each group. For all four categories, the averages are always larger in the first group than in the second. A two-sample t-test confirms this by rejecting the null that the two group averages are equal in the sample markets. Hence, following price changes, consumers are more likely to switch to a product from the same

[3]For the set of instruments used, see Fosgerau et al. (2022) for details.

[4]Details on the numbers of products, options, brands, and firms for these categories are provided in Appendix B.

Table 3: Average Diversion Ratios from Estimated IPDL Models

	Diversion ratio ($-\frac{\partial s_k/\partial p_j}{\partial s_j/\partial p_j}$)			
	Shampoo	Laundry detergent	Milk	Sugar substitute
Same option, different brands	8.82%	4.31%	10.30%	5.94%
Same brand, different options	5.53%	2.13%	6.87%	5.25%
t-test	110.496***	48.177***	33.860***	8.656***

Notes: This table reports the average diversion ratios across all markets. The t-test denotes the two-sample t statistic for which the null is that the two groups have equal means in the sample markets. *** denotes the 1% significance level.

Table 4: Tests on Average Numbers of Different Choices Made by Households

	Option	Brand
Expected number of different choices	3.11	3.54
Mean observed number of different choices	2.02	2.22
t-test	-50.470***	-42.023***

Note: The sample size of the t-test is 3036. *** denotes the 1% significance level.

option than from the same brand. They exhibit strong preferences for options, which would have been ignored if we were to model the demand only at the brand level.

Second, focusing on the shampoo market, I utilize the micro-level data to show households' product choices and switching patterns. Regarding the product choices, I calculate the number of different brands or options purchased by each household in 2006. The data show that on average, each household has four shopping trips over the course of the year and purchases one bottle of shampoo each time. There are on average 12 brands that are always available in each store/week and six available options. If each household randomly chooses one brand or one option at a time, the expected numbers of different brands and options, when it makes the choices four times, are 3.54 and 3.11, respectively.[5] I then test whether the average observed numbers of brands (2.22) and options (2.02) are lower than the expected values. In Table 4, the t statistics are both negative for option and brand choices, with p-values less than 1%. Thus, the sample households purchase significantly smaller numbers of different options and brands than they would do from randomization. This implies that

[5] Let $x(\geq 4)$ denote the number of available choices. The consumer is asked to randomly choose one choice each time and repeat this four times. Then the expected number of different choices selected in the four tries, y, is given by $y = \frac{(2x-1)(2x^2-2x+1)}{x^3}$.

Table 5: Conditional Probabilities of Maintaining Product Choices

	Option	Brand
Estimated mean conditional probability	38.26%	22.60%
Mean probability from independent choice	16.67%	8.33%

Note: The estimated mean conditional probability is the diagonal mean of the estimated transition matrix for households' product choices. The mean probability from independent choice assumes that households' product choices from the six available options or 12 available brands are independent across periods.

households do not randomize their option and brand choices and that they have tastes for certain options and brands.

In addition, I estimate the transition matrix of households' brand and option choices across different periods in 2006. Table 5 reports the average probabilities of choosing an option or a brand in the current period conditional on choosing the same option or brand in the previous period. On average, the sample households maintain their option choices with a probability of 38.26% and brand choices with a probability of 22.60%. These numbers are much larger than the probabilities of making independent option or brand choices across periods. This again confirms that households tend to stick to their chosen options or brands over time. Therefore, the two sets of evidence suggest that consumers have preferences over options and brands, which motivates modeling demand at the brand-option level.

3 Model and Identification

3.1 The Pure Characteristics Model with Options

The pure characteristics model I use follows Bajari and Benkard (2005). Suppose there are J products available on the market. Each product $j(=1,...,J \in \mathcal{J})$ is a combination of brand $b(=1,...,B)$, size $l(=1,...,L)$, and option $o(=1,...,O)$, and can be characterized by a K-dimensional vector of observed characteristics $\boldsymbol{X}_j$ and a scalar of unobserved (to the econometrician) characteristic ξ_j. $\boldsymbol{X}_j$ can be further divided into a K^c-dimensional vector of continuous characteristics, $(x^c_{1j},...,x^c_{K^c j})$, and a K^d-dimensional vector of discrete characteristics, $(x^d_{1j},...,x^d_{K^d j})$, where $K^c + K^d = K$. The continuous characteristics can be, for example, promotional variables such as display and price-reduction. The discrete characteristics can be option, size, brand, and market/time dummies. The unobserved (by econometricians) characteristic ξ_j represents some measure of product quality that is known to consumers themselves. Let p_{jt} denote the price of product j in market/time $t \in T$. There are I consumers and each is indexed by i. Consumer i has income y_{it} in t. It is further assumed that the utility function is log-linear in continuous characteristics and linear in discrete characteristics. Given the prices and the income level, each consumer's utility maximization problem is written as

$$\max_{j,m} u(\boldsymbol{X}_{jt}, \xi_{jt}, m; \boldsymbol{\beta}_{it}) = \sum_{k=1}^{K^c} \beta_{ikt} \log(x^c_{kjt}) + \beta_{i\xi t} \log(\xi_{jt}) + \sum_{k=1}^{K^d} \beta_{ikt} x^d_{kjt} + m \\ \text{s.t. } p_{jt} + m \leq y_{it}, \tag{2}$$

where m denotes the numeraire good and $\boldsymbol{\beta}_{it} = (\boldsymbol{\beta}_{ikt}, \beta_{i\xi t})$ denotes the vector of consumer i's taste parameters.

Theorem 1. *For any two products j and j' with positive demand in t,*

a. if $\boldsymbol{X}_{jt} = \boldsymbol{X}_{j't}$ and $\xi_{jt} = \xi_{j't}$, then $p_{jt} = p_{j't}$;

b. if $\boldsymbol{X}_{jt} = \boldsymbol{X}_{j't}$ and $\xi_{jt} > \xi_{j't}$, then $p_{jt} > p_{j't}$;

c. $|p_{jt} - p_{j't}| \leq M(|\boldsymbol{X}_{jt} - \boldsymbol{X}_{j't}| + |\xi_{jt} - \xi_{j't}|)$ for some $M < \infty$.

Proof. See Bajari and Benkard (2005).

Theorem 1 means that there exists a unique equilibrium price function, denoted as $\boldsymbol{p}_t(\boldsymbol{X}_{jt}, \xi_{jt})$, in each market. This price function is an equilibrium object as a result of firms' interactions or product choices and is market-specific (hence the t subscript). For each product j, which is described by a bundle of characteristics $\boldsymbol{X}_{jt}$ and ξ_{jt}, there is an equilibrium price p_{jt}. Implicitly, the price function already captures the interactions among different products in equilibrium, so the characteristics of other products do not appear in $\boldsymbol{p}_t(\cdot)$. Its form depends on all market primitives, for example, firms' marginal costs of production, the ownership matrix, consumer preferences, etc. Also, it is non-separable because there are no assumptions on the market primitives. Such a general form has the advantage that there is no need to worry about imposing incorrect supply-side assumptions in practice.

3.2 Model Identification

Identification of the model proceeds in two steps. The first is to identify the unobserved product characteristic and the equilibrium price function. The second is to identify the taste parameters. For simplicity of notation, I omit the market/time subscript t in this section.

In the first step, the key assumption for identification is that ξ_j only varies at the brand level.

Assumption 1. *For all $j, j' \in \mathcal{J}$, if $b(j) = b(j')$, then $\xi_j = \xi_{j'} = \xi_b$.*

ξ_b denotes the brand-level unobserved characteristic. Assumption 1 can also be interpreted as a timing assumption: observing the set of brands and all ξ_b's, consumer i first makes the brand choice; given the chosen brand, consumer i then chooses one of the options. Intuitively, ξ_b captures consumers' overall perception of a brand. Advertising activity, for example, is mostly at the brand (or product line) level but not at the brand-option level. Also, firms usually assign managers for each brand but not for each individual product or option.

Imposing this assumption has two advantages. First, it avoids the assumption that ξ_j is independent of $\boldsymbol{X}_j$ at the product level, which is assumed in the pure characteristics models without options. Such an independence assumption is not likely to hold because in reality brand managers often jointly choose the characteristics of products, which are brand-size-option combinations. Second, it avoids the assumption that the price function $\boldsymbol{p}(\boldsymbol{X}_j, \xi_j)$ is monotonic in ξ_j, which has limited application in practice. As is shown in Figure 1, different options of the same shampoo brand often have the same price, which violates the monotonicity assumption at the product level. But if we impose that different options of the same brand have the same ξ, we can still have monotonicity at the brand level and apply the model to many markets like the shampoo market.[6] In some markets, however, options can be vertically differentiated and have different prices, in which case we need to alter this assumption, as discussed in detail below.

Under this assumption, I can identify ξ_b up to a normalization, since only the order but not the magnitude of ξ_b matters. Here I normalize ξ_b such that it follows a uniform distribution between zero and one.

Theorem 2. *Given the utility function in (2) and Assumption 1, if many products are observed in each market, then ξ_b is identified as*

$$\xi_b = F_{p|\boldsymbol{X}=\boldsymbol{X}_b}(\bar{p}_b), \tag{3}$$

where F_p denotes the cumulative distribution function (CDF) of price, $\bar{p}_b$ denotes the average price of brand b, and $\boldsymbol{X}_b$ contains all observed characteristics of the brand.

Proof. See Appendix C.

Theorem 2 implies that ξ_b can be identified as the conditional distribution of the average prices across options. The intuition is that when a brand tends to offer options that have higher average prices, one may consider this brand to be a high-end brand with good quality. The identification result makes use of price information on all available products from each brand. More importantly, it does not need to specify a baseline option as done in Bajari and Benkard (2005) and use prices of only that baseline. In many markets, it is unclear how to define the baseline. Hence, my approach can avoid predicting the price of the baseline if a brand does not offer it in reality. The identified ξ_b is then shared among all products of brand b so ξ_j is identified. And the price function is just given by the observed

[6] I reproduce the results on the standard deviations of price for other CPG categories and show them in Figure B1. Prices in these categories are also quite similar across options within brand-size combinations.

prices, i.e., for any point $(\boldsymbol{X}_0, \xi_0)$,

$$\boldsymbol{p}(\boldsymbol{X}_0, \xi_0) = p_j \text{ for } j \in \mathcal{J}, \tag{4}$$

where p_j is the observed price of a product with $\boldsymbol{X}_j = \boldsymbol{X}_0$ and $\xi_j = \xi_0$.

In addition, Assumption 1 implies over-identification of the model. If different options from a brand have the same ξ_b, one can use a subset of all options to identify ξ_b. When different subsets are used, ξ_b is over-identified. Thus, I provide an over-identification test in Section 4.1 by comparing different ξ_b estimates from two subsets of the options. Results show no statistical difference between the two sets of estimates, thereby confirming Assumption 1. I also provide simulation results on the power of this test.

It is easy to modify Assumption 1 to accommodate markets in which options are defined vertically. Assuming that ξ varies at the brand level is intuitive when prices are similar across options within each brand, i.e., options are horizontally defined. In some markets, however, prices can be very different across options. For example, in the PC market, options can be defined by CPU generations and laptops offering older generations often have lower prices than those offering newer generations. In these markets, we can assume that ξ varies at the option level instead of the brand level. And the identification follows likewise. In practice, one can test the assumptions on ξ using the data and choose the most reasonable one suggested by the over-identification test.

Identification in the second step follows Bajari and Benkard (2005) and Bajari and Kahn (2005). The strategies differ between taste parameters for continuous characteristics and those for discrete ones. For continuous characteristics (including ξ_j), their parameters can be derived trivially from the first-order conditions of the utility maximization problem:

$$\begin{aligned} \beta_{ik} &= x^c_{kj^*} \frac{\partial \boldsymbol{p}(\boldsymbol{x}_{j^*}, \xi_{j^*})}{\partial x^c_{kj}} \quad \text{for } k = 1, ..., K^c; \\ \beta_{i\xi} &= \xi_{j^*} \frac{\partial \boldsymbol{p}(\boldsymbol{x}_{j^*}, \xi_{j^*})}{\partial \xi_j}. \end{aligned} \tag{5}$$

where j^* denotes consumer i's observed product choice.[7] For discrete characteristics, identification can be achieved by assuming functional forms on consumer demographics. To see this, assume x_k^d is equal to 0 or 1. When consumer i optimally chooses product j^*, denote the characteristics vector of j^* as $\hat{\boldsymbol{X}}_{j^*}$ if $x_{kj^*}^d = 1$, or as $\bar{\boldsymbol{X}}_{j^*}$ if $x_{kj^*}^d = 0$. Then define $\Delta p = \boldsymbol{p}(\hat{\boldsymbol{X}}_{j^*}, \xi_j) - \boldsymbol{p}(\bar{\boldsymbol{X}}_{j^*}, \xi_j)$. The utility maximization problem implies that

$$\begin{aligned} x_{kj^*}^d = 1 &\implies \beta_{ik} > \Delta p; \\ x_{kj^*}^d = 0 &\implies \beta_{ik} < \Delta p. \end{aligned} \tag{6}$$

Assume the taste parameter β_{ik} is a function of consumer demographics, i.e.

$$\beta_{ik} = h(\boldsymbol{R}_i; \boldsymbol{\theta}_k) + \eta_{ik}, \tag{7}$$

where $\boldsymbol{R}_i$ denotes consumer i's observed vector of demographics and η_{ik} denotes some unobserved consumer-specific taste shock. If we assume η_{ik} follows some distribution, then we can identify $\boldsymbol{\theta}_k$ using (6)-(7) through maximum likelihood.

4 Estimation

4.1 Unobserved Characteristic and Price Function

In the first step, estimation of the brand-level unobserved characteristic ξ_{bt} requires an estimator for the conditional distribution of average prices on all option-level observed product characteristics, as is shown in equation (3). One challenge here is the dimensionality, as $\boldsymbol{X}_b$ contains all characteristics of all options from a brand. To deal with this, I smooth out all the continuous characteristics. Specifically, I use smoothing splines to estimate

$$p_{jt} = \boldsymbol{\psi}_t(\boldsymbol{X}_{jt}^c, Option_j, Brand_j, \gamma_t) + e_{jt}, \tag{8}$$

[7] As in Bajari and Benkard (2005), this implicitly assumes that the number of products is sufficiently large that the choice set is approximately continuous. I provide some evidence on the continuous characteristics used in this paper in Section 4.2 to explain why this assumption is plausible.

where $\boldsymbol{\psi}(\cdot)$ is a function of continuous characteristics $\boldsymbol{X}^c_{jt}$, option, brand, and market/time dummies. The continuous characteristics include display, price-reduction, and normalized bottle size, as introduced in Section 2.1. Note that here $\boldsymbol{\psi}(\cdot)$ is different from the equilibrium price function in Section 3.1 because it does not have ξ_{jt}. It can be considered as a preparation step before the actual first-step estimation. Thus, we do not want to restrict the functional form of $\boldsymbol{\psi}(\cdot)$, and the smoothing splines achieve this by admitting very flexible functional forms. In the estimation, I use a full tensor product smooth term for display and price-reduction to take into account the interaction between the two. I also include a smooth term in size because of the presence of non-linear pricing in size. The smoothed prices are obtained by setting display and price-reduction to zero and normalized bottle size to one in the estimated $\hat{\boldsymbol{\psi}}(\cdot)$. The average price of each brand, $\bar{p}_{bt}$, is then calculated using these smoothed prices.

Next, I adopt the Generalized Random Forests (GRF) method proposed by Athey et al. (2019) to estimate the conditional distribution of the smoothed average prices, which is given by

$$\hat{\xi}_{bt} = 1 - \hat{\Pr}[p_t > \bar{p}_{bt} | \boldsymbol{X} = \boldsymbol{X}_{bt}], \tag{9}$$

where $\boldsymbol{X}_{bt}$ includes the brand-level option dummies and market/time dummies.[8] The brand-level option dummies are defined as being equal to one if the brand offers the option and zero otherwise. The GRF method is particularly suitable for this estimation. It handles well a system with many discrete covariates, which in this case are the many dummies, by providing fast and efficient performance. Standard errors for $\hat{\xi}_{bt}$ are calculated via bootstrapping to take into account that the smoothed prices (instead of observed prices) are used in the estimation.

While the GRF method estimates the brand-level unobserved characteristic and the conditional CDF, it does not produce differentiable models. To get estimates of the price

[8] The estimated survival function $\hat{\Pr}[\cdot]$ is obtained using the implementation of survival forests in the R package `grf`.

function at the product level, I then use smoothing splines again because they allow for flexible functional forms and produce continuously differentiable functions from which derivatives can be estimated in the next step. This estimator of equation (4) is given by

$$\hat{p}_{jt} = \hat{\boldsymbol{\phi}}_t(\boldsymbol{X}_{jt}, \hat{\xi}_{jt}), \tag{10}$$

where $\boldsymbol{X}_{jt}$ includes display, price-reduction, option, brand, size, and market/time dummies, and $\hat{\xi}_{jt} = \hat{\xi}_{bt}$ for each product j from brand b. Like in equation (8), a full tensor product smooth term is used for display and price-reduction. All other variables are estimated via thin plate regression splines. I choose the numbers of basis functions for all the smooth terms to minimize the Akaike Information Criterion (AIC) from a grid search.[9]

As explained in Section 3.2, ξ_{bt} is over-identified if we use different subsets of options to calculate the average price. To test whether Assumption 1 is true, I use the following test procedure. First, all the options are randomly divided into two even groups G_1 and G_2, which are mutually exclusive and represent different subsets of options. Next, for each group, I calculate the average smoothed prices and then estimate $\hat{\xi}_{bt}$. Denote these estimates as $\hat{\xi}_{bt}^{G_1}$ and $\hat{\xi}_{bt}^{G_2}$, respectively. The two sets of estimates are paired since I use only those brands that appear in both groups in the estimation. Then I perform a Friedman test to detect differences in $\hat{\xi}_{bt}$ across the two groups. The null hypothesis is that $\hat{\xi}_{bt}$ are the same regardless of option groups. Rejection of the null indicates that the assumption of a brand-level ξ is violated. This test can also apply to more than two groups of estimates. In Appendix D, I simulate the power of the test and find that a sample size equal to 100 is sufficient for the test to have 80% power when the significance level is 5%. When the observations are smaller than 100, I suggest using a larger significance criterion in practice.

[9] Estimation is done using the R package `mgcv` (v1.8-42; Wood, 2017).

4.2 Taste Parameters

Estimation of the taste parameters depends on each product characteristic. When the characteristic is continuous, an estimator for derivatives of the price function is required, as shown in equation (5). From the estimated price function via smoothing splines in the first step, it is natural and easy to estimate the derivatives via finite differences. I then calculate the 95% confidence intervals for these estimates by bootstrapping to take into account the usage of the first-step estimates. The estimated taste parameters are given by

$$\begin{aligned} \hat{\beta}_{ikt} &= x^c_{kj^*t} \frac{\partial \hat{\boldsymbol{\phi}}_t}{\partial x^c_{kjt}} \quad \text{for } k = 1, ..., K_c; \\ \hat{\beta}_{i\xi t} &= \hat{\xi}_{j^*t} \frac{\partial \hat{\boldsymbol{\phi}}_t}{\partial \xi_{jt}}, \end{aligned} \tag{11}$$

where $\partial \hat{\boldsymbol{\phi}}_t / \partial x^c_{kjt}$ represents the estimator for the derivative from finite differences. As explained in Section 3.2, implementing equation (11) relies on the number of products in the choice set being large enough. In the data, display and price-reduction do have a wide range of observed values so it is reasonable to treat them as continuous.[10]

For discrete characteristics, estimation requires assumptions on the functional form of the taste parameters. I consider the following linear functional form:

$$\beta_{ikt} = \theta_{0kt} + \sum_z \theta_{zkt} r_{izt} + \eta_{ikt}, \tag{12}$$

where r_{izt} denotes the z^{th} demographic variable for household i in market/time t. I include three demographic variables: income level, family size, and renter/homeowner. η_{ikt} is assumed to follow a normal distribution. Given the inequalities in (6), I estimate equation (12) using probit models. Ideally one can estimate a full multivariate probit model to account for the correlation between all taste parameters. In practice, it is time-consuming to estimate such a model when the number of discrete characteristics is large. Hence, for feasibility, I assume the taste parameters are correlated within each type of dummy and uncorrelated across different types of dummies. Then I estimate the multivariate probit models for option,

[10] Appendix E reports histograms of these two variables.

brand, size, and time dummies, respectively.[11] Given that there are still a large number of brand dummies, I estimate the probit models separately for each firm to get taste parameters for brands.

4.3 Demand and Welfare

With all the structural parameters, demand for product j as a function of prices is given by

$$q_{jt}(\boldsymbol{p}_t) = \sum_{i=1}^{I} \mathbb{1}\{u_{ijt} \geq u_{ij't}, \forall j' \neq j\}, \tag{13}$$

where the functional form of u_{ijt} is given in the consumer's maximization problem in (2). The change in consumer welfare is estimated through a counterfactual where the new products are removed from the choice set, holding prices fixed. Such a welfare change can be interpreted as a variety effect in Hausman and Leonard (2002). There is an outside option, denoted as $j = 0$, which includes all products from the fringe brands that are not considered as the inside brands, and all fringe bottle sizes of the inside brands that are less popular. Choosing the outside option is the same as spending all the income on the numeraire good, i.e., the consumer's utility is normalized to be $u_{it} = y_{it}$ if $j^* = 0$. Then I calculate the compensating variation when the choice set changes from the actual one to the counterfactual. Let $\mathcal{J}_1$ denote the actual choice set including the new products and $\mathcal{J}_2$ denote the simulated choice set excluding the new products. Consumer i's estimated utility from the actual choice set is given by

$$\hat{u}^*_{it} = \max_{j \in \mathcal{J}_1} u(\boldsymbol{X}_{jt}, \hat{\xi}_{jt}, y_{it} - p_{jt}; \hat{\boldsymbol{\beta}}_{it}). \tag{14}$$

The compensating variation W must make the consumer equally well off when facing the simulated choice set, i.e.,

$$\max_{j \in \mathcal{J}_2} u(\boldsymbol{X}_{jt}, \hat{\xi}_{jt}, y_{it} - p_{jt} + W; \hat{\boldsymbol{\beta}}_{it}) = \hat{u}^*_{it}. \tag{15}$$

[11] I estimate time dummies instead of market/time dummies because households in the sample only appear in one of the two cities.

5 Application to New Shampoo Products

5.1 New Shampoo Introduction

I apply the pure characteristics model with options to estimate the welfare effects from the introduction of the most successful new shampoo brand in 2006 by L'Oréal USA.[12] The U.S. shampoo industry is among the most active non-food consumer packaged goods industries, where firms constantly introduce new brands to provide consumers with targeted solutions, new varieties, and/or new designs. Between 2002 and 2006, an average of 26 new shampoo brands were introduced to the U.S. market each year, according to the IRI data. The rate of introduction is the second largest in all non-food categories.

Several major firms dominate this industry. Table 6 shows volume shares of the top five firms over 2001-2007, together with the Herfindahl–Hirschman Indexes (HHIs). These firms account for over 80% of total volume sales, and there is a growing trend in their total market share. In particular, L'Oréal USA has played a more important role over the years, with its volume share growing from 4.7% to 10.3%. The HHI, on the other hand, hovers around 2,000, indicating the shampoo industry is moderately concentrated and that the level of concentration does not change much.

In June 2006, following a trend of transitioning salon products to mainstream retail, L'Oréal USA introduced a new brand called L'Oréal Vive Pro. It comes in four different options, *Color*, *Other*, *Regular*, and *Volumizing*. Figure 2 plots monthly volume shares of the four new products. For each option, I also include one of the most popular inside products. Volume shares are based on the total sales volume observed in the data. Out of the four new products, the *Color* option has the largest increase in market share and accounts for more than 1% of the total market after six months of the introduction. Overall the four new products have a quite strong market performance and obtain market shares comparable to some popular existing products. The existing products are affected differently by the

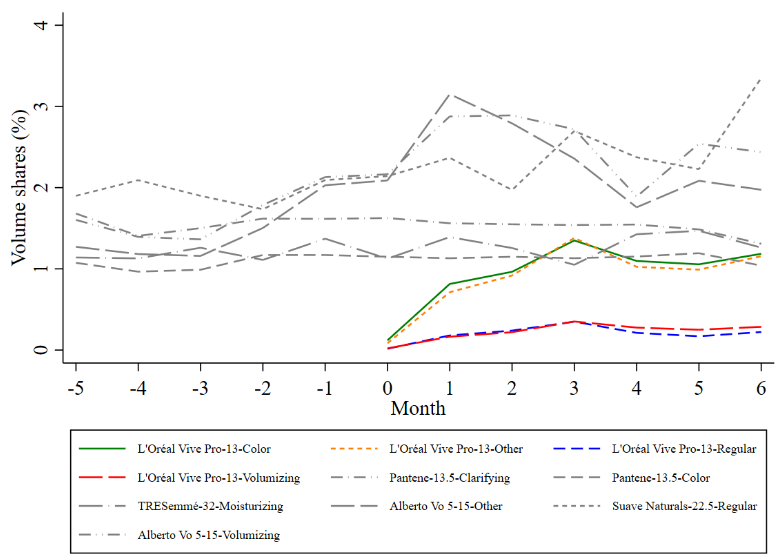
[12] Other new brands introduced in 2006 have much lower market shares, thereby considered to be fringe brands. Their effects are negligible and unlikely to contaminate the main results.

Table 6: Volume Shares (%) of Top Firms and HHI

Year	Procter & Gamble	Unilever	Alberto Culver	L'Oréal USA	Diamond Products Inc.	Total	HHI
2001	28.6	27.6	15.7	4.7	5.7	82.2	1917
2002	29.9	26.7	16.2	5.0	5.8	83.7	1967
2003	30.1	28.1	16.1	7.0	3.5	84.8	2043
2004	29.4	27.2	17.1	8.8	3.4	85.8	2003
2005	29.1	26.7	16.8	10.0	3.7	86.4	1977
2006	30.2	25.4	16.9	10.2	5.0	87.7	1990
2007	30.0	24.9	17.4	10.3	4.4	86.9	1965

Note: This table provides the firm-level volume shares for the top five firms and the HHI over 2001-2007 using the IRI data. The HHI is scaled from 0 to 10,000.

Figure 2: Monthly Volume Shares of Some Inside Products



Note: The horizontal axis shows the number of months after the introduction. Month 0 represents the official month when the four new products were introduced. Product sizes are measured in ounces (oz).

introduction. For example, there is a downward trend in volume shares of two products from the brand Alberto Vo 5. Products from TRESemmé and Pantene seem to be unaffected while the product from Suave Naturals exhibits an upward trend.

5.2 Reduced-Form Evidence

In this section, I provide reduced-form evidence on the impacts of the four new products. Using market-level data in 2006, I run a series of regressions to quantify the effects

Table 7: Change in Overall Market Size

	Total sales volume (Bottles)
$\mathbb{1}${Post-Introduction}	482.930
	(305.829)
Constant	20895.427***
	(178.400)
Market	×
Quarter	×
R^2	0.989
Obs.	600

Note: The regression uses data at the market-month level in 2006. Standard errors are clustered at the market level and are shown in parentheses. *** denotes the 1% significance level. 1 bottle = 16oz.

Table 8: Changes in Market Sizes of Different Options

	Total sales volume (Bottles)					
	Color	*Other*	*Regular*	*Volumizing*	*Clarifying*	*Moisturizing*
$\mathbb{1}${Post-Introduction}	67.114***	408.311***	100.802	-108.569	-33.185	48.458
	(20.141)	(72.101)	(66.752)	(113.492)	(51.372)	(73.822)
Constant	1232.420***	4712.962***	4906.216***	3872.486***	2621.512***	3549.831***
	(11.749)	(42.059)	(38.939)	(66.204)	(29.967)	(43.063)
Market	×	×	×	×	×	×
Quarter	×	×	×	×	×	×
R^2	0.973	0.986	0.966	0.980	0.978	0.980
Obs.	600	600	600	600	600	600

Note: The regressions use data at the market-month level in 2006. Standard errors are clustered at the market level and shown in parentheses. *** denotes the 1% significance level. 1 bottle = 16oz.

on the overall market size and the option sub-markets. The pre-introduction period is from January 2006 to June 2006 and the post-introduction period is from June 2006 to December 2006. First, I regress the total sales volume on a post-introduction indicator, market fixed effects, and quarter fixed effects using data at the market-month level. Table 7 reports the results. The coefficient of the indicator is positive yet not significant. Therefore, the four new products do not expand the market significantly.

Next, I examine whether the total sales volume of each option changed after the introduction. For each option, the sales volume is regressed on a post-introduction indicator, market fixed effects, and quarter fixed effects. In Table 8, both *Color* and *Other* have significant increases in sales volume, which suggests that the new brand expands the sizes of these two sub-markets. The rest options do not experience significant effects. Given the fact that the overall shampoo market does not expand, there could be some consumers who switch to *Color* and *Other* options after the introduction. To explore the extent to which this is the case, I turn to the micro-level data. Among all households that purchased *Regular*, *Volumizing*, *Clarifying*, or *Moisturizing* before the introduction, 7.27% purchased *Color* products and 24.67% purchased *Other* products after the introduction. Out of the 7.27% or 24.67%, roughly one-third switched completely to *Color* or *Other*. Therefore, part of the increased market sizes of these two options could come from consumers who did not choose these two before but turned to them after.

Furthermore, regarding the option sub-markets, I perform an event-study analysis to compare the market shares of the new brand with those of existing brands in the post-introduction period. Specifically, for each option offered by the new brand, I use data aggregated to the product-month level to run a two-way fixed effects event-study specification, which is given by

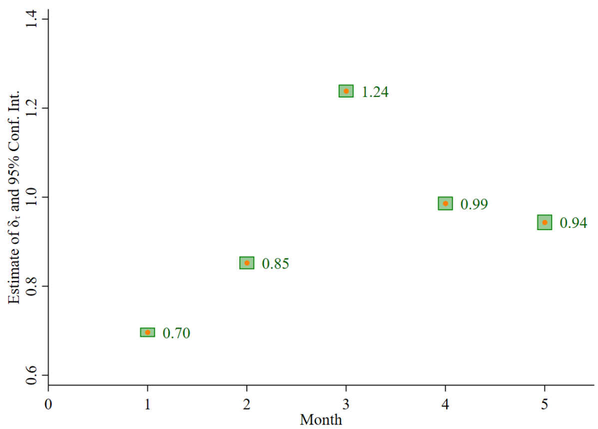
$$s_{jt} = \rho_j + \lambda_t + \sum_{\tau=1}^{5} \mathbb{1}\{\tau = t\} \times D_j \times \delta_\tau + \epsilon_{jt}, \tag{16}$$

where s_{jt} denotes product j's market share in month t, ρ_j is the product fixed effect, λ_t is the month fixed effect, and D_j is equal to 1 if j belongs to the new brand and 0 otherwise. The event-study coefficient of interest is δ_τ, where τ denotes the number of months after the introduction. I normalize δ_0 to be zero, i.e. we can interpret δ_τ as the difference in shares between the new products and other existing products in month τ relative to the difference in month 0 during which the new products are introduced. Figure 3 plots the estimates of δ_τ. For all four options, the increase in the new brand's market share is significantly larger than that in other brands following the introduction. Moreover, *Color* and *Other* experience larger changes than the other two options. For instance, the market share of option *Color* from the new brand exhibits a relative increase of 1.24% compared to existing brands three months after the introduction whereas for option *Regular* or *Volumizing* the increase is around 0.3%. Although these estimates are merely descriptive, they imply that the new brand potentially cannibalizes other brands that offer the same options.

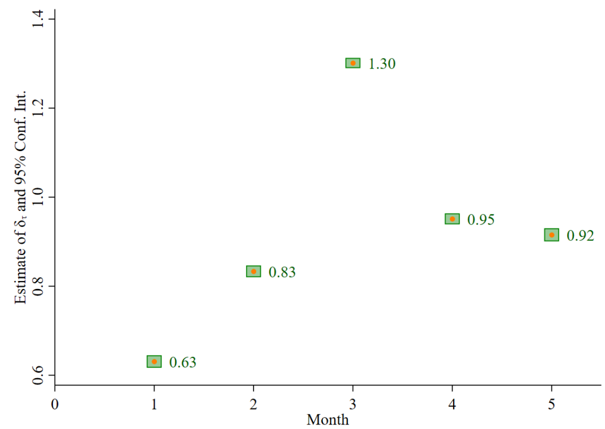
To summarize, although the four new products do not significantly expand the overall

Figure 3: Event-Study Estimates for Market Shares by Option

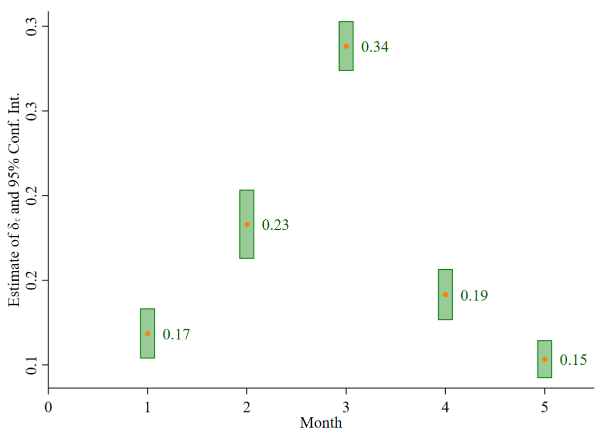
(a) *Color*



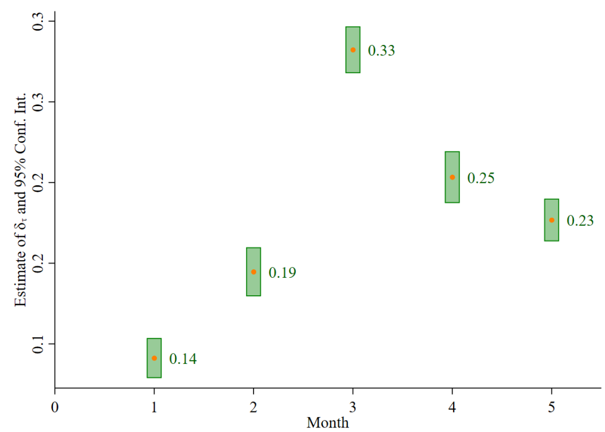
(b) *Other*



(c) *Regular*



(d) *Volumizing*



Note: This figure plots the event-study estimates of δ_τ and the 95% confidence intervals. Confidence intervals are calculated using standard errors clustered at the product level. The numbers of clusters for *Color*, *Other*, *Regular*, and *Volumizing* are 72, 245, 227, and 175, respectively.

shampoo market, they create a cannibalization effect on existing products within each option they offer. For options *Color* and *Other*, the new brand even expands the sizes of these two sub-markets. Such reduced-form evidence underlines the important role of options in capturing the substitution patterns among all inside products. Therefore, in the structural estimates shown below, I incorporate option choices in the demand model.

5.3 Full Model Estimates

5.3.1 First-Step Results

Results of the first step estimation consist of estimates of the unobserved characteristic and the price function. I present the empirical distribution of the standard errors of $\hat{\xi}_{bt}$'s in Figure 4. For comparison, I also estimate $\hat{\xi}_{bt}$'s using the approach proposed in theory by Bajari and Benkard (2005), where prices of a baseline option are employed. The estimator is given by

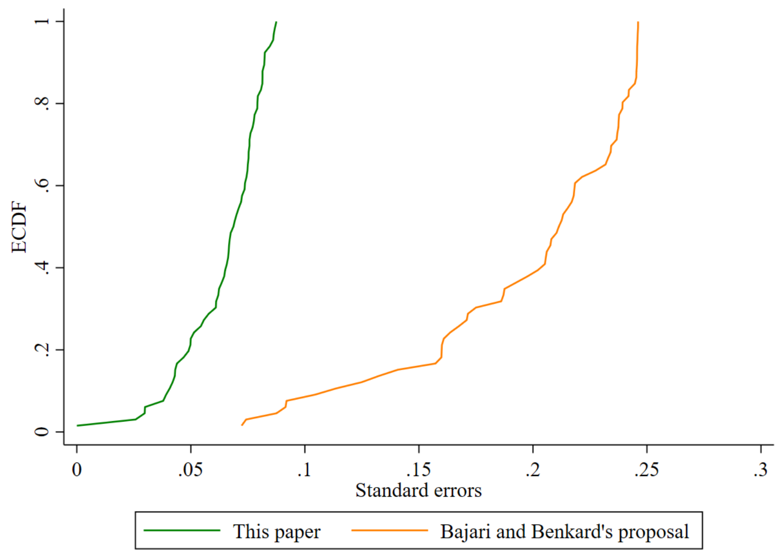
$$\hat{\xi}_{bt} = \hat{F}_{p_t|\boldsymbol{X}=\boldsymbol{X}_t}(\dot{p}_{\tilde{j}t}), \tag{17}$$

where $\hat{F}$ represents a local linear kernel estimator for the conditional CDF, and $\boldsymbol{X}_t$ includes market/time dummies. $\tilde{j}$ represents the baseline option for which I choose the *Regular* option from each brand. $\dot{p}_{\tilde{j}t}$ is the smoothed price of the baseline option $\tilde{j}$ with zero units of display and price-reduction and one normalized bottle size. Smoothed prices are obtained from smoothing splines estimates of equation (8). This procedure also predicts the price of the baseline if a brand does not offer it in reality. Figure 4 plots the distribution of the asymptotic standard errors from this approach as well.[13] It can be seen that my approach produces much smaller standard errors, with a median of 0.069 and a maximum of 0.088. Bajari and Benkard's proposed approach, by contrast, generates a median of 0.212 and a maximum of 0.246. Therefore, my identification and estimation strategies provide more precise estimates of the unobserved product characteristic.

Table 9 reports the p-values of the over-identification test on the unobserved characteristic. I consider all ten possible 3-option combinations achieved by dividing the six options into two even groups. The p-values reported are all greater than the 10% significance level, indicating that the null hypothesis cannot be rejected in the data, i.e., different options of the same brand do have the same ξ. Moreover, my sample has 235 observations in this

[13] Bajari and Benkard (2005) show that for the kernel estimator, the bootstrapped standard errors they obtain are even larger than the asymptotic ones. So I do not bootstrap the standard errors from their proposed approach given that it is time-consuming to do so.

Figure 4: Empirical Distribution of Standard Errors for $\hat{\xi}_{bt}$'s



Note: This figure plots the empirical distributions (ECDFs) of the standard errors for $\hat{\xi}_{bt}$'s. The green curve is for the bootstrapped standard errors estimated from my approach, which utilizes the GRF method and the brand-level average prices. The orange curve is for the asymptotic standard errors estimated from Bajari and Benkard's proposal, which utilizes a local linear kernel estimator and the prices of the baseline option.

Table 9: Friedman Tests of Over-Identification for the Ten Possible 3-Option Combinations

	(1)	(2)	(3)	(4)	(5)
p-value	0.327	0.674	0.500	0.336	0.199
	(6)	(7)	(8)	(9)	(10)
	0.116	0.053	0.386	0.439	0.101

Note: For the Friedman tests, the sample size is equal to 235.

step, which is larger than the reported threshold of 100 in Section 4.1 for the test to have 80% power with the 5% significance level. Therefore, the over-identification test has enough power to reject the null for my sample.

Table 10 presents the smoothing splines estimates of the price function from equation (10). The effective degrees of freedom (EDFs) are greater than one for all smooth terms except the market/time dummies, so there is nonlinearity in display, price-reduction, ξ, size, brand, and option dummies. For the smooth terms with EDFs greater than one, the F tests are mostly significant with p-values less than 1%. Thus, the smoothing splines provide proper estimates of the nonlinear equilibrium price function.

Table 10: Price Function Estimates from Smoothing Splines

	k	EDF	F
te(Display, Price-reduction)	16	5.957	20.043***
$s(\xi)$	7	5.649	56.331***
s(Size)	11	8.056	44.331***
s(Brand)	17	10.679	3.013***
s(Option)	6	2.773	1.648
s(Market/time)	4	1.000	1.508
Constant		3.212***(0.014)	
Adjusted R^2		0.986	
Obs.		235	

Note: te denotes a full tensor product smooth term. s denotes a thin plate regression spline smooth term. k denotes the number of basis functions used in the smooth term. EDF denotes the effective degrees of freedom. F denotes the ANOVA test statistic on the overall significance of the smooth term. *** denotes the 1% significance level. For the constant, the standard error is shown in parentheses.

5.3.2 Second-Step Results

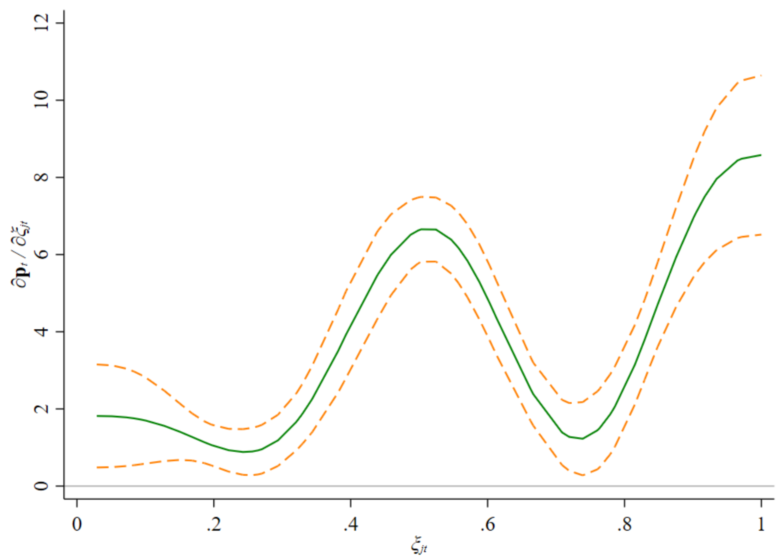
Results of the second step estimation consist of estimates of the taste parameters. I plot the derivative estimates w.r.t ξ in Figure 5. The estimates together with their 95% confidence intervals are all positive, implying the price function is strictly increasing in ξ. When ξ is close to zero or one, the confidence intervals are wider, i.e., the derivatives are less precisely estimated. This is because the data set contains fewer shampoo products with extremely low or high prices. Median estimates of the taste parameters for continuous characteristics from equation (11) are shown in Table 11. A median household has positive taste parameters for display and ξ. For price-reduction, the median estimate is negative because, in equation (11), the partial derivative w.r.t price-reduction has negative estimates, i.e. the equilibrium price is decreasing in the price-reduction promotion. In absolute values, display has smaller parameter estimates than price-reduction, indicating that price-reduction has a larger effect on consumers' choices.[14]

For discrete characteristics, Table 12 reports the median parameter estimates with 95% confidence intervals.[15] Parameters for the time dummies are both positive, with the

[14]This is consistent with the results on households' purchasing probabilities reported in Appendix F. When controlling for prices, households are more likely to purchase a product when it has a price-reduction promotion than when it has a display promotion.

[15]To save space, I omit taste parameters for brand and size dummies, but they are also estimated.

Figure 5: Partial Derivative Estimates of Price Function w.r.t ξ



Note: The solid green curve denotes the estimated partial derivative. The dash orange curves denote the 95% bootstrapped confidence interval.

Table 11: Median Estimates of Taste Parameters for Continuous Characteristics

	$\hat{\beta}_{ikt}$	95% confidence interval
Display	0.013	(0.000, 0.046)
Price-reduction	-0.137	(-0.282, -0.061)
ξ	0.916	(0.388, 1.548)

Note: This table reports the median taste parameter estimates across all households, together with 95% bootstrapped confidence intervals.

median for quarter one slightly larger than that for quarter two. The discrepancy is potentially due to lower average prices in quarter one. Parameters for the option dummies are positive as well. *Volumizing* has the largest median estimate, followed by *Other* and *Color*. *Clarifying*, *Moisturizing*, and *Regular* have relatively small parameter estimates, implying that consumers may benefit less from these options.

Table 12: Median Estimates of Taste Parameters for Discrete Characteristics

	$\hat{\beta}_{ikt}$	95% confidence interval
Time dummies		
Quarter one	2.898	(2.696, 3.102)
Quarter two	2.659	(2.453, 2.892)
Option dummies		
Clarifying	1.315	(1.057, 1.585)
Color	1.995	(1.740, 2.240)
Moisturizing	1.728	(1.515, 1.933)
Other	2.084	(1.899, 2.275)
Regular	1.800	(1.606, 2.007)
Volumizing	2.204	(1.977, 2.430)

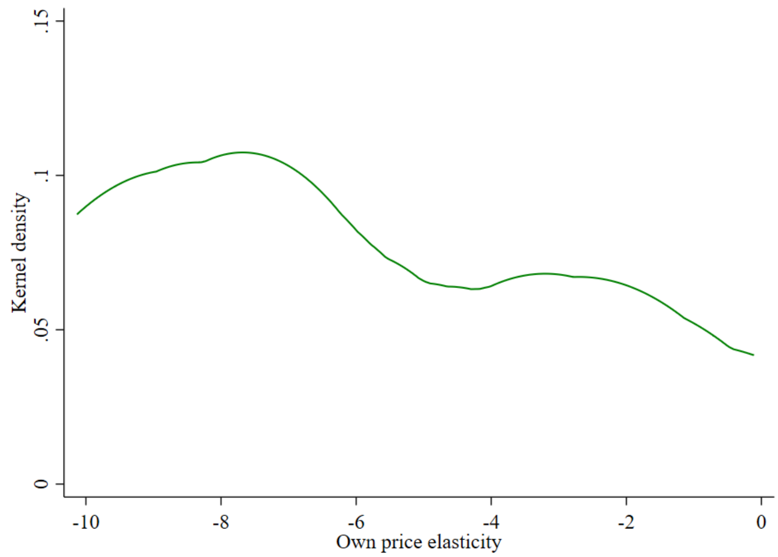
Note: This table reports the median taste parameter estimates across all households. The 95% confidence interval is calculated using standard errors that are clustered at the household level.

5.3.3 Elasticities

Next, following Bajari and Benkard (2005), I calculate the own price elasticities for all inside products using the structural parameters. The distribution of the estimated own price elasticities is shown in Figure 6. The estimates range from -0.109 to -10.032, with a mean of -6.072 and a median of -7.518. These estimates are fairly comparable with those from the random coefficients nested logit model where the average own price elasticity is -5.106.

One criticism of the pure characteristics model is that it can produce implausibly large elasticities. This problem can be mitigated by including a smaller number of inside products. For example, when estimating the demand for personal computers, Bajari and Benkard (2005) report a median elasticity of -100 for 695 inside products and a median of -11 for 24 inside products. However, reducing the number of products leads to information loss and compromises the estimation efficiency of ξ. By contrast, my approach of adding options to the model provides an alternative solution to the problem and at the same time maintains a reasonably large set of inside products to facilitate the first-step estimation. After option choices are incorporated, not all inside products are close substitutes to each other, which enables the model to produce more realistic elasticities. Also, consumers only need information on different brands and options of the inside products, so their consideration

Figure 6: Distribution of Estimated Own Price Elasticities



Note: This figure plots the kernel density of the estimated own price elasticities of all inside products.

set is not incredibly large

5.4 Model Fit

With all the structural estimates in hand, I first examine the model fit by comparing observed and model-predicted market shares in the post-introduction period, as defined in Section 5.2. When predicting market shares, I allow each household to choose among all the inside products and the outside option. Predicted quantities are calculated by aggregating all households' optimal product choices, as in equation (13). Then I calculate the predicted market shares as the ratio of predicted quantities over total volume sales. For comparison, I also predict the market shares using estimates from other pure characteristics models. At the brand-size-option level, I obtain estimates by using Bajari and Benkard's proposed identification and estimation strategies. At the brand-size level, I estimate a pure characteristics model with no option choices.[16]

Table 13 presents the results for two measures of prediction errors, the root mean

[16]This in-sample prediction of market shares does not apply to models with logit errors because the logit errors force the market shares to fit perfectly in the estimation.

Table 13: In-Sample Prediction Errors for Post-Introduction Shares

	Brand-size-option level		Brand-size level
	(1) This paper	(2) Bajari and Benkard's proposal	(3) Pure characteristics model w/o options
RMSE	0.036	0.097	0.171
MAE	0.008	0.033	0.064

Note: RMSE represents the root mean square error and MAE represents the mean absolute error. Both RMSE and MAE are averaged over all inside products.

square errors (RMSEs) and the mean absolute errors (MAEs). The predictions from my model have the smallest RMSE and MAE. Differences in prediction errors between my approach and Bajari and Benkard's proposed approach result from the first-step identification and estimation strategies. For ξ, my identification uses information on the prices of all options from a brand rather than just the price of a baseline. Also, my estimation adopts an efficient machine-learning method that produces more accurate estimates of ξ. These estimates directly affect the model fit by appearing in the utility function. In addition, they have an indirect impact on the fit by appearing in the price function, which influences the derivative estimates for the taste parameters in equation (11). My approach also improves the prediction accuracy from the pure characteristics model without options, which is less precise because consumers' option preferences are ignored. This is not surprising because previous evidence in Section 2.2 has shown that option choices matter to consumers.

Next, I compare the counterfactual predictions between my model and models with logit errors. Using the structural estimates, I predict market shares in the counterfactual where the four new products are removed and compare the predicted shares to the observed shares in the pre-introduction period, as defined in Section 5.2. The accuracy of this counterfactual prediction is important in determining how accurate the welfare estimates are. For models with logit errors, this is where they can be problematic. I obtain the prediction errors from the random coefficients nested logit (RCNL) model, the nested logit (NL) model, and the random coefficients logit (BLP) model. In each model with logit errors, I include a correction term in the spirit of Ackerberg and Rysman (2005). Specifically, the

Table 14: Out-of-Sample Prediction Errors for Pre-Introduction Shares

	(1) This paper	(2) RCNL	(3) NL	(4) BLP
RMSE	0.140	0.151	0.160	0.179
MAE	0.067	0.079	0.078	0.090

Note: RMSE represents the root mean square error and MAE represents the mean absolute error. Both RMSE and MAE are calculated using the average shares at the product-market level. For models with logit errors in Columns (2)-(4), a correction term is included in the estimation.

correction term for each product is a weighted count of the total number of products and the number of products in its own nest in each market.[17] Table 14 presents the RMSEs and MAEs calculated using the average shares at the product-market level. My approach produces smaller prediction errors than all models with logit errors, which suggests that the pure characteristics model with options developed in this paper can fit the data more closely. Intuitively, this is because, without logit errors, the new products are closer substitutes to the existing products. In the counterfactual, consumers are more likely to switch to the existing products than to choose not to purchase any inside product. Overall, this comparison of model fit suggests that the welfare estimate from my approach shown in the next section is more reasonable than those from models with logit errors.

5.5 Welfare Estimates

Using equations (14)-(15), I estimate, for each household, the change in consumer welfare from the four new products within six months of their introduction and aggregate the estimates over all households. Given the total sales volume of around 155 thousand bottles in the sample grocery stores, the total welfare gain is approximately \$86 thousand. Table 15 shows the average welfare gains scaled by sales volume. The average change is \$0.56 per bottle across the four new products. The average price is \$4.81, so the improvement in consumer welfare is equivalent to 11.64% of the prices they pay. The *Color* option provides the largest welfare gain of \$0.62 per bottle, which is equivalent to 12.83% of the price.

[17] For each model with logit errors, the prediction errors without the correction term are very close to those reported. Therefore, the cases without the correction term are omitted in Table 14.

Table 15: Consumer Welfare Estimates for the Four New Products

Option	Welfare ($/16oz)	Price ($/16oz)	Welfare/Price
Color	0.62	4.83	12.83%
Other	0.57	4.81	11.76%
Regular	0.24	4.73	5.06%
Volumizing	0.57	4.77	11.96%
Average	0.56	4.81	11.64%

Table 16: Consumer Welfare Estimates from Different Models

		Welfare ($/16oz)	Welfare/Price
Models with logit errors			
Without options	BLP-Brand	4.82	100.29%
With options	RCNL	2.49	51.88%
	RCNL with correction	2.06	42.84%
Pure characteristics models			
Without options	Pure characteristics model-Brand	0.87	18.13%
With options	This paper	0.56	11.64%

Volumizing also produces a large welfare gain of $0.57 per bottle even though it has much smaller sales. This is because the *Volumizing* option has quite a large taste parameter estimate. While fewer consumers purchased the product, those who did purchase enjoyed it a lot.

I also compare the welfare estimate from my approach with those from other existing models in Table 16. For models with logit errors, I obtain the estimates from the brand-size level BLP model, the RCNL model, and the RCNL model with a correction term.[18] To address price endogeneity in these models, I use the Hausman instrument in Hausman (1996) and the differentiation instruments proposed by Gandhi and Houde (2020). For the endogenous within-nest market share in nested models, I instrument it with the number of products per nest, a common choice from the literature on nested logit estimation. For the pure characteristics model without options, I use data aggregated to the brand-size level. All welfare estimates are scaled to the same observed sales volume in the data.

It is worth emphasizing several findings from the comparison. First, models with logit errors produce much larger welfare effects, which is consistent with the findings in Song

[18] Estimation is done using the Python package PyBLP developed by Conlon and Gortmaker (2020).

(2007). For example, the RCNL with a correction term provides an average welfare estimate of $2.06, which is equivalent to almost 43% of the average price. This estimate is too large to be plausible. One reason is that the U.S. shampoo market already had about 300 brands before the introduction. It is unlikely that adding one new brand can increase consumer welfare by an amount equivalent to almost half the price. The other reason is that the four new products only account for a small market share. Evidence in Section 5.2 suggests that they do not significantly expand the overall market; instead, they mainly steal market shares from existing products. The second finding worth emphasizing is that using models with options can reduce the large welfare estimates relative to models without options. My model, for example, generates welfare estimates that are 36% less than those from the brand-size level pure characteristics model. The RCNL model also reduces the welfare estimate from the brand-level BLP model by 57%. The main reason for such reductions is that models without option choices overlook close substitutes for the new products at the option level. Finally, it is worth noticing that for the pure characteristics model with options, I only report the welfare estimate from my approach but not that from Bajari and Benkard's proposed approach. This is because all the procedures of estimating the price function and the taste parameters are fully developed in this paper but not in their paper.

6 Conclusion

In this paper, I develop an empirical framework of consumer demand that takes into consideration both brand choices and option choices. The framework employs the pure characteristics demand model instead of models with logit errors to deal with issues in estimating consumer welfare. I apply the model to estimate the welfare effects of four new products introduced to the U.S. shampoo market. From the structural estimates and counterfactual analysis, I find that the four new shampoo products increase consumer welfare which is equivalent to 11.64% of the prices on average. Consumers gain the largest welfare increase from the new product offering the relative niche option. Compared with demand models with logit errors, my model can reduce the implausibly large welfare estimates by at least 73%. Ignoring option choices can exaggerate the welfare estimates as well. The pure characteristics model developed in this paper has a better model fit and produces more reasonable elasticities than pure characteristics models that ignore options. Overall, my approach provides an accurate structural estimation of welfare gains from new products, which is crucial in examining whether firms offer new features that boost consumer welfare or just introduce similar products to steal from rivals.

Some works remain for future research. First, the empirical model in this paper omits the supply side, and thus the welfare estimate ignores potential price effects. This is not a big issue if the new products only have relatively small market shares. Still, adding the supply side to the model is worth investigating and can make such a model handle a variety of interesting topics in empirical industrial organization and marketing. Second, in the model, I assume consumers choose a brand and an option simultaneously. It is also worthwhile to consider a situation where consumers make their choices sequentially and incorporate a dynamic setting. Third, while this paper only models a scalar unobserved product characteristic, allowing for multiple unobservables might be an interesting extension to explore.

References

Ackerberg, D. A., & Rysman, M. (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *The RAND Journal of Economics*, *36*(4), 771–789.

Agarwal, N. (2015). An empirical model of the medical match. *The American Economic Review*, *105*(7), 1939–78.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178.

Bajari, P., & Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, *113*(6), 1239–1276.

Bajari, P., & Kahn, M. E. (2005). Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business & Economic Statistics*, *23*(1), 20–33.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, *89*(5), 2015–2048.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 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.

Berry, S., & Pakes, A. (2007). The pure characteristics demand model. *International Economic Review*, *48*(4), 1193–1225.

Borkovsky, R. N., Goldfarb, A., Haviv, A. M., & Moorthy, S. (2017). Measuring and understanding brand value in a dynamic model of brand management. *Marketing Science*, *36*(4), 471–499.

Brenkers, R., & Verboven, F. (2006). Liberalizing a distribution system: The European car market. *Journal of the European Economic Association*, *4*(1), 216–251.

Bronnenberg, B. J., Kruger, M. W., & Mela, C. F. (2008). Database paper—The IRI marketing data set. *Marketing Science*, *27*(4), 745–748.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, *13*(2), 185–213.

Choi, H.-J., Wohlgenant, M. K., & Zheng, X. (2013). Household-level welfare effects of organic milk introduction. *American Journal of Agricultural Economics*, *95*(4), 1009–1028.

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.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American Economic Review*, *70*(3), 312–326.

Dubé, J.-P. (2005). Product differentiation and mergers in the carbonated soft drink industry. *Journal of Economics & Management Strategy*, *14*(4), 879–904.

Erdem, T., Keane, M. P., & Sun, B. (2008). A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science*, *27*(6), 1111–1125.

Fosgerau, M., Monardo, J., & De Palma, A. (2022). The inverse product differentiation logit model. *Available at SSRN 3141041*.

Gandhi, A., & Houde, J.-F. (2020). Measuring substitution patterns in differentiated-products industries. *NBER Working paper*, (w26375).

Giacomo, M. D. (2008). GMM estimation of a structural demand model for yogurt and the effects of the introduction of new brands. *Empirical Economics*, *34*(3), 537–565.

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

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.

Hausman, J. A., & Leonard, G. K. (2002). The competitive effects of a new product introduction: A case study. *The Journal of Industrial Economics*, *50*(3), 237–263.

Hendel, I., & Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, *74*(6), 1637–1673.

Jiang, J., & Chen, X. (2021). Pure characteristics demand models and distributionally robust mathematical programs with stochastic complementarity constraints. *arXiv preprint arXiv:2102.04933*.

Joo, J. (2023). Rational inattention as an empirical framework for discrete choice and consumer-welfare evaluation. *Journal of Marketing Research*, *60*(2), 278–298.

Miller, N. H., & Weinberg, M. C. (2017). Understanding the price effects of the MillerCoors joint venture. *Econometrica*, *85*(6), 1763–1791.

Morozov, I. (2023). Measuring benefits from new products in markets with information frictions. *Management Science*.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, *69*(2), 307–342.

Nevo, A. (2003). New products, quality changes, and welfare measures computed from estimated demand systems. *Review of Economics and statistics*, *85*(2), 266–275.

Pang, J.-S., Su, C.-L., & Lee, Y.-C. (2015). A constructive approach to estimating pure characteristics demand models with pricing. *Operations Research*, *63*(3), 639–659.

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

Song, M. (2007). Measuring consumer welfare in the CPU market: An application of the pure-characteristics demand model. *The RAND Journal of Economics*, *38*(2), 429–446.

Song, M. (2010). The quality adjusted price index in the pure characteristics demand model. *Journal of Business & Economic Statistics*, *28*(1), 190–199.

Sun, H., Su, C.-L., & Chen, X. (2017). SAA-regularized methods for multiproduct price optimization under the pure characteristics demand model. *Mathematical Programming*, *165*(1), 361–389.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. CRC press.

Appendices

A List of Inside Products

Table A1: Information on Inside Products

Brand	Size	Option
Firm: Alberto Culver		
Alberto Vo 5	22.5	*Volumizing*
Alberto Vo 5	15	*Clarifying*
Alberto Vo 5	15	*Moisturizing*
Alberto Vo 5	15	*Other*
Alberto Vo 5	15	*Regular*
Alberto Vo 5	15	*Volumizing*
Alberto Vo 5	14.5	*Other*
TRESemmé	32	*Clarifying*
TRESemmé	32	*Moisturizing*
TRESemmé	32	*Regular*
TRESemmé	32	*Volumizing*
TRESemmé	20	*Color*
Firm: Diamond Products Inc.		
White Rain	20	*Regular*
White Rain	20	*Volumizing*
White Rain	15	*Regular*
White Rain Naturals	20	*Clarifying*
White Rain Naturals	20	*Other*
White Rain Naturals	20	*Volumizing*
Firm: L'Oréal USA		
Garnier Fructis	25.4	*Regular*
Garnier Fructis	13	*Color*
Garnier Fructis	13	*Other*
Garnier Fructis	13	*Regular*
Garnier Fructis	13	*Volumizing*
L'Oréal Vive	13	*Other*
L'Oréal Vive Pro	13	*Color*
L'Oréal Vive Pro	13	*Other*
L'Oréal Vive Pro	13	*Regular*
L'Oréal Vive Pro	13	*Volumizing*
Firm: Procter & Gamble		
Aussie	16	*Moisturizing*
Aussie	16	*Regular*
Aussie	16	*Volumizing*
Herbal Essences	12	*Color*
Herbal Essences	12	*Moisturizing*
Herbal Essences	12	*Other*
Herbal Essences	12	*Volumizing*
Pantene	25.4	*Clarifying*
Pantene	25.4	*Color*
Pantene	25.4	*Moisturizing*
Pantene	25.4	*Other*

Continued on the next page.

Table A1: Information on Inside Products(continued)

Brand	Size	Option
Pantene	25.4	*Volumizing*
Pantene	13.5	*Clarifying*
Pantene	13.5	*Color*
Pantene	13.5	*Moisturizing*
Pantene	13.5	*Other*
Pantene	13.5	*Volumizing*
Pantene Pro V	13.5	*Other*
Pantene Pro V	13.5	*Color*
Pert	25.4	*Regular*
Pert	13.5	*Clarifying*
Pert	13.5	*Volumizing*
Firm: Unilever		
Dove	25.4	*Moisturizing*
Dove	12	*Clarifying*
Dove	12	*Moisturizing*
Dove	12	*Volumizing*
Dove Advanced	12	*Color*
Dove Advanced	12	*Moisturizing*
Suave	22.5	*Clarifying*
Suave	15	*Clarifying*
Suave	15	*Other*
Suave Naturals	27	*Regular*
Suave Naturals	22.5	*Moisturizing*
Suave Naturals	22.5	*Regular*
Suave Naturals	15	*Moisturizing*
Suave Naturals	15	*Other*
Suave Naturals	15	*Regular*
Suave Professionals	14.5	*Clarifying*
Suave Professionals	14.5	*Color*
Suave Professionals	14.5	*Moisturizing*
Suave Professionals	14.5	*Other*
Suave Professionals	14.5	*Volumizing*

B Information on Other CPG Categories

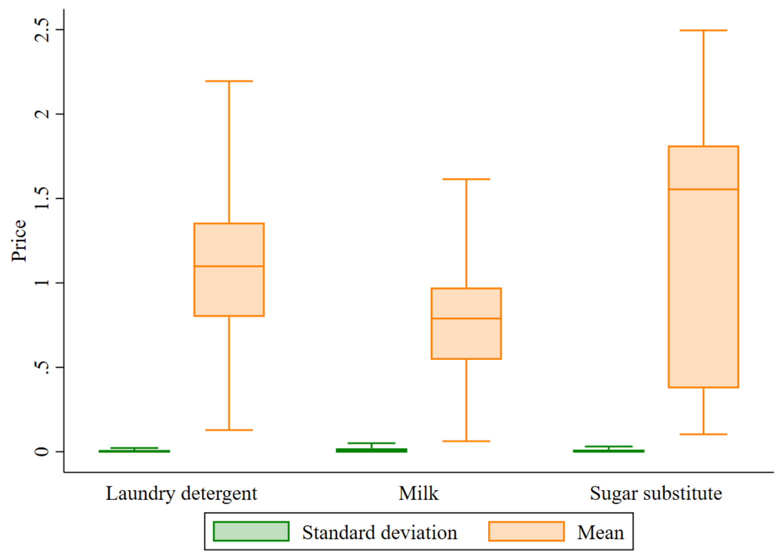
Table B1: Numbers of Inside Products for Other Categories

CPG Category	Firms	Brands	Options	Products
Laundry detergent	8	18	6	40
Milk	8	17	4	40
Sugar substitute	4	5	7	24

Table B2: List of Options for Other Categories

CPG Category	Option list
Laundry detergent	Liquid-regular, Liquid-concentrated, Liquid-ultra concentrated, Powder-regular, Powder-concentrated, Powder-ultra concentrated
Milk	0%, 1%, 2%, Whole milk
Sugar substitute	Aspartame, Brown sugar, Nutra sweet, Saccharin, Saccharin & dextrose, Sugar substitute blend, Sucralose

Figure B1: Standard Deviations and Means of Price for Brand-sizes in Other Categories



Note: For each brand-size combination, the standard deviation and mean of price are calculated across all options in each store/week. Outside values are excluded from the box plots.

C Proof of Theorem 2

For simplicity of notation, I omit the t subscript. Let J_b denote the number of products from brand b. In equilibrium, the observed price p_j is equal to the price function $\boldsymbol{p}(\boldsymbol{X}_j, \xi_j)$ where $\xi_j = \xi_b$ for $j = 1, ..., J_b$. Then the average price of brand b is given by

$$\bar{p}_b = \frac{1}{J_b}\sum_{j=1}^{J_b} p_j = \frac{1}{J_b}\sum_{j=1}^{J_b} \boldsymbol{p}(\boldsymbol{X}_j, \xi_b) = \boldsymbol{\phi}(\boldsymbol{X}_1, \boldsymbol{X}_2, ..., \boldsymbol{X}_{J_b}, \xi_b) = \boldsymbol{\phi}(\boldsymbol{X}_b, \xi_b) \tag{C1}$$

where $\boldsymbol{X}_b$ includes all characteristics of all products from brand b and $\boldsymbol{\phi}(\cdot)$ rewrites the average into a function of $\boldsymbol{X}_b$ and ξ_b. Then, the conditional distribution of the average price is

$$
\begin{aligned}
F_{p|\boldsymbol{X}=\boldsymbol{X}_b}(\bar{p}_b) &= \Pr[\boldsymbol{\phi}(\boldsymbol{X},\xi) \leq \bar{p}_b | \boldsymbol{X} = \boldsymbol{X}_b] \\
&= \Pr[\xi \leq \boldsymbol{\phi}^{-1}(\boldsymbol{X},\bar{p}_b) | \boldsymbol{X} = \boldsymbol{X}_b] \\
&= \Pr[\xi \leq \boldsymbol{\phi}^{-1}(\boldsymbol{X}_b,\bar{p}_b)] \\
&= \boldsymbol{\phi}^{-1}(\boldsymbol{X}_b,\bar{p}_b) = \xi_b.
\end{aligned}
\tag{C2}
$$

The second line of equation (C2) follows from theorem 1, the third line holds because ξ_b is independent of the option-level characteristics, and the last line holds because ξ_b follows $U[0,1]$.

D Power of the Over-identification Test

I construct a simulated data set to see how the test power changes with the sample size. In the simulation, price is assumed to follow a simple data generating process (DGP) according to

$$
p_j = \exp(\xi_b) + u_j,\ u \sim \mathcal{N}(0,1). \tag{D1}
$$

There are four options available for each brand. And they are divided into two even groups, G_1 and G_2. The unobserved characteristic in G_1 is denoted as ξ_{1b} and is randomly drawn from $U[0,1]$. The unobserved characteristic in G_2 is denoted as ξ_{2b}, and $\xi_{2b} = \xi_{1b} + 1$. That is, the unobserved characteristics are different across the two groups for each brand. Then I use the GRF method to estimate the two sets of ξ_b and perform the Friedman test. The sample size is chosen to be $N = 100, 200, 300, 400$ and for each sample size, 10,000 Monte Carlo replications are simulated. Table D1 presents the rejection rates at the 5% level. For all sample sizes, rejection rates are greater than 80%, the rule-of-thumb power at the 5% significance level. And they increase as the sample size becomes larger. Therefore, the over-identification test has enough power.

Table D1: Rejection Rate of the Friedman Test

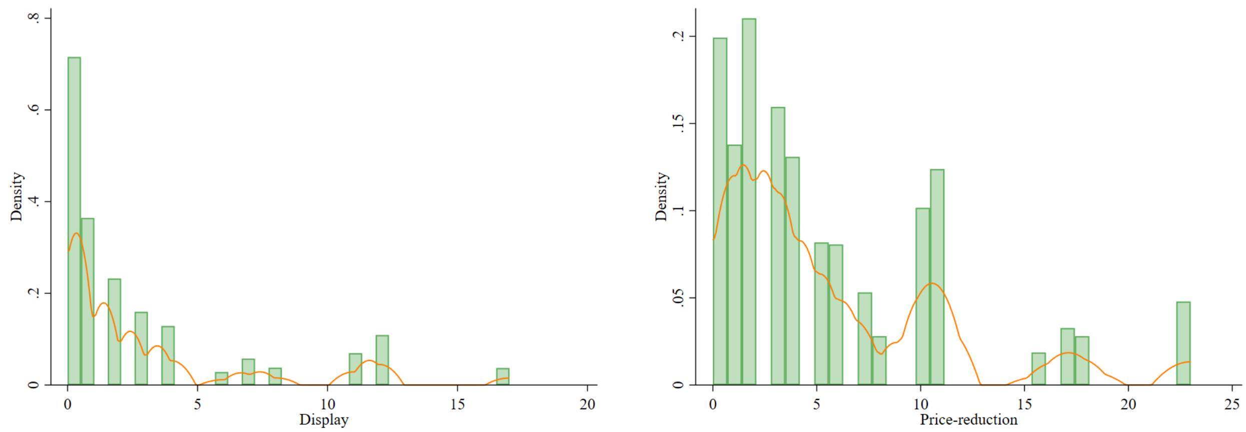
Sample size	100	200	300	400
Rejection rate	80.52%	86.44%	88.47%	90.11%

Note: For each sample size, 10,000 replications are simulated. The significance level is 5%.

E Variation in Display and Price-Reduction

Figure E1: Histograms of Display and Price-Reduction

(a) (b)



Note: The figure plots the density of display and price-reduction in the micro-level data. The green lines represent the kernel density estimates.

F Households' Purchasing Probabilities

I evaluate households' purchasing probabilities given prices and promotional variables. The micro data set provides households' weekly shampoo purchases by UPCs and stores. I combine them with the market-level data to get all available weekly prices, displays, and price-reductions for these UPCs in the stores visited by the households in 2006. Then I use probit regressions to show how promotional variables affect whether or not households purchase the products. In Table F1, Column (1) shows that price-reduction has a significantly positive coefficient when I control for prices and other covariates. This suggests that households are more likely to make the purchase when the product has a price-reduction promotion. A similar effect is shown for display in Column (2). Combining the two variables in one regression (Column (3)), I find that both exhibit positive coefficients as well, and that price-reduction has a larger coefficient than display. Therefore, both promotional variables can increase households' probabilities of purchasing shampoo products; and the households are affected more by price-reduction than by display.

Table F1: Households' Purchasing Probabilities

	Purchasing Probabilities		
	(1)	(2)	(3)
Price-reduction	0.454***		0.407***
	(0.048)		(0.045)
Display		0.365***	0.309***
		(0.041)	(0.040)
Price	-0.097***	-0.314***	-0.103***
	(0.027)	(0.025)	(0.026)
Constant	-0.542***	-0.295***	-0.498***
	(0.078)	(0.086)	(0.079)
Obs.	40162	40162	40162

Note: The probit regressions also include dummies on week, store, brand, option, and size. Standard errors are clustered at the UPC level and shown in parentheses. There are 207 clusters. *** denotes the 1% significance level.