Research Article

Daniel Toro-González, R. Karina Gallardo, Jia Yan, and Jill J. McCluskey*

Quality Differentiation with Flavors: Demand Estimation of Unobserved Attributes

Abstract: This article estimates the demand for mint-flavored gum products using grocery store sales data and accounting for consumers' valuation of quality. Unobserved product attributes, such as flavor quality, are important elements to consider when estimating the demand for gum. The estimation results suggest that gum is an inelastic product. A positive relationship between willingness to pay and unobserved quality was identified, implying that gum industry should be able to command a premium for higher quality mint-flavored products.

Keywords: quality differentiation, unobserved product attributes, gum

***Corresponding author: Jill J. McCluskey,** Washington State University, Pullman, WA 99164, USA, E-mail: mccluskey@wsu.edu
Daniel Toro-González, Universidad Tecnológica de Bolívar, Cartagena, Bolívar, Colombia, E-mail: dtoro@unitecnologica.edu.co
R. Karina Gallardo: E-mail: karina_gallardo@wsu.edu, **Jia Yan:** E-mail: jiay@wsu.edu, School of Economic Sciences, Washington State University, Pullman, WA, USA

1 Introduction

Chewing gum is one of the best performing segments within the confectionery market, and the global market for gum is forecast to reach US$ 20.7 billion by the year 2015 (Global Industry Analysts 2011). The industry is characterized by its product innovation, focused on novel and unique flavors, new ingredients, different product shapes, varied colors, and distinctive packaging techniques (Global Industry Analysts 2011). The production of gum requires ingredients such as a gum base, sweeteners, and a variety of flavors. The gum base is usually a standard mix of synthetic latex and natural rubber extracted from the Sapodilla trees. This ingredient is not a source of product differentiation since only two different gum textures are commercialized, chewing and bubble gum, with 83% and 17% of the market share, respectively. The case of sweeteners is similar, just two categories can be identified, sugar and sugar-free gum, with a share of 42% and 58%, respectively. Flavor is the major source of product differentiation. This is reflected in the high number of products available in the market with differentiated flavors.

Mint is the flavor that displays the largest market share, accounting for approximately 50%, followed by fruit-flavored gum with 19% (Nielsen 2005), of total market share. Mint oil is an important product to the Pacific Northwest (e.g., Washington, Oregon, and Idaho), which is responsible for 83% of the U.S. spearmint production and 50% worldwide (Far West Spearmint Oil Administrative Committee 2013). However, the share of mint oil supplied by U.S. producers has been falling in recent years. Mint oil imports now account for approximately 25% of this market. Dealers buy cheaper and lower quality oil from China and India and then blend these oils with the more expensive high-quality U.S. oil to accommodate each gum manufacturer's standard (Christensen 2010). From 1993 to 2012, the number of spearmint acres harvested dropped 38% and the price per pound dropped 23% (U.S. Department of Agriculture, National Agricultural Statistics Service 2013).

The supply chain of mint oil involves three parties: (1) the mint oil producers who sell mint oil to the dealers, (2) the dealers who mix different mint oils to generate the final flavoring oil mixtures, and (3) gum manufacturers who buy the oil mixtures from dealers to produce gum. For each flavor of mint gum (e.g. "doublemint," mint splash, or cool mint), the gum manufacturer demands a specific mixture of mint oils from the dealer. This specific mint oil mixture is the result of blending mint oils from different qualities, which are measured in terms of the presence of oil components such as limonene, menthone, purine, and esters. The less mixture

This research is funded by USDA-NIFA Specialty Crop Research Initiative Competitive Grant 2009-51181-05790 "Agronomic and Biochemical Impacts of Biotic and Abiotic Stress on Pacific Northwest Flavor Crops." The study has benefited from comments of Ron Mittelhammer and Martin Kolmar. All remaining errors are the responsibility of the authors. This work was funded by USDA's National Institute of Food and Agriculture Specialty Crop Research Initiative project, Agronomic and Biochemical Impacts of Biotic and Abiotic Stress on Pacific Northwest Flavor Crops (2009-51181-05790).

of oils from different growing regions used in the final product leads to the higher flavor quality that is measured in terms of strength and duration.[1]

The U.S. mint oil industry has been decreasing its market share with the increase of mint oil imports (Christensen 2010). In this circumstance, an investigation of the elasticity of substitution between high-quality domestically produced oil and low-quality imported oil is useful to the U.S. mint oil industry to recover their negotiating power with end-user manufacturers. However, it is not possible for the researcher to observe or measure differences in the quality of the mint oil mixed to produce each gum flavor because each gum recipe is private information for the manufacturing firms. One alternative is to analyze the consumers' price elasticity for mint gum accounting for product heterogeneity and unobservable (to the econometrician) product characteristics such as flavor quality.

We assume that consumers have a predetermined ranking of mint flavors based on their previous consumption and their own preferences. Based on the differences in flavor profiles across products, consumers know which flavor delivers the highest utility in consumption. Although the differentiation is subjective, the differences in flavor profiles for each gum product are real rather than merely perceived by the consumers. For a discussion of price competition in a setting when product differentiation is subjective, see Tremblay and Polasky (2002). From the literature on product differentiation in oligopolistic models, we know that unobservable product attributes generate endogeneity of prices, leading to biased estimators. This study considers the unobserved attributes surrounding flavor quality and compares the results using different estimation strategies.

Only a few studies have been conducted on brand choice and products using mint oil as flavoring ingredient, such as in gum and other mint-flavored products.[2] For toothpaste, previous studies include Kaya et al. (2010), Gutierrez (2005), Yang, Zhou and Chen (2005), and Shin (2008); and for chewing gum, Chung and Szymanski (1997). All of these studies examine factors that affect brand selection. However, no study has analyzed mint-flavored product choices in the presence of unobservable product quality.

The aim of this study is to estimate the demand for mint-flavored gum accounting for the existence of unobservable flavor quality attributes. We postulate that one can depict the demand for gum using a discrete-choice model in an oligopoly context in which prices are endogenously determined by price-setting firms. We also consider the existence of product characteristics unobservable (to the econometrician), but fully considered by gum manufacturers when setting prices and by consumers when purchasing the products. We use quarterly aggregated gum retail sales data for 2005, arranged in a three-dimensional panel of quantities and prices for 49 contiguous states.

The article is organized the following way. In the next section, the data are described. Following that, we present two different modeling strategies: the multinomial logit (MNL) and the nested logit (NL) model. In the subsequent section, we introduce the estimation results and comparisons across the different modeling strategies. Next, counterfactual scenarios are simulated to estimate the changes in market shares derived from changes in the consumer's valuations of the unobservable attributes of the products. Finally, implications are discussed and conclusions are drawn.

2 Data

The database consists of bubble gum and chewing gum household purchases obtained from the AC Nielsen Homescan survey aggregated by brand for each state. Summary statistics are presented in Table 1. We define each state as an independent market with information for 2005 four quarters (each quarter is composed of three months).[3] The definition of the level of aggregation that constitutes a market was made based on the information availability for each brand within the geographical aggregation. Although information for the state and county levels are both available; in most of the cases, there are not enough individuals to statistically represent each county. On the other hand, for each state, there is sufficient individual-level information.

The number of aggregate observations is 4,403, which is now reported in Table 1. The number of observations responds to an unbalanced quarterly data panel for 49 states and 67 brands of gum.

The original disaggregated database consists of 106,496 observations of daily household grocery store purchases of gum (bubble and chewing) from the AC Nielsen homescan survey for 2005. The information was collected in 2,106 counties distributed across the 49 contiguous states. The overall sample consists of 19,702

1 Typically, mint oil contains menthofuran. This substance reduces palatability. Oils that contain high levels of menthofuran are considered lower quality, whereas oils with lower level of this substance are considered high quality (Mint Dealer Company Representative 2010).

2 An estimate of 45% of the total U.S. mint oil production is devoted to the production of gum and the same percentage to the production of toothpaste. The remaining 10% in used in other confectioneries, pharmaceutical applications, flavor for liqueurs, and aroma therapy (Heritage Productions 2000).

3 We choose the year 2005 because it contains more information about product attributes and consumers characteristics than any other year available in the survey.

Table 1 Definition of variables and summary statistics ($n = 4{,}403$)

Variable name	Type	Description	Mean	Std. Dev.	Min	Max
Price	Numerical	Per ounce	0.825	0.347	0.005	2.890
Units	Numerical	No. of units of the item	1.528	1.212	1	29
Product characteristics						
Form	Categorical	Pieces (63%), sticks (26%) or other (11%)	1.587	0.604	1	3
Flavor	Categorical	Mint (34%), fruit (19%), spice (6%), variety (15%), sour (4%), other flavors 20%	2.231	1.702	1	6
Mint flavor	Categorical	Peppermint (13%), Spearmint (11%), and other (8%), No mint (65%)	1.050	1.174	0	3
Texture	Categorical	1 if chewing gum and 0 if bubble gum	0.829	0.377	0	1
Volume	Numerical	Size of the package in ounces	0. 398	2. 135	0.01	28
Sugar content	Dummy	1 if Sugar free and 0 otherwise	0.585	0.493	0	1
Coupon	Dummy	1 if consumers use a coupon and 0 otherwise	0.100	0.301	0	1
Household characteristics						
Household size	Numerical	No. individuals in household	2.867	1.448	1	9
Income level*	Categorical	Income intervals begin with annual incomes under $5,000 and the highest interval is $100,000 and over	19.413	5.643	3	27
Children	Dummy	1 if children under 18 in the household	0.401	0.490	0	1
Marital status	Dummy	1 if married	0.678	0.467	0	1
Age	Numerical	Ages starting at 25, top is 65 + with				
Other variables						
Projection 61k**	Numerical	Expansion factor	3,184.664	3,427.413	139	31,230
Region***	Categorical	East (17%), West (19%), South (37%), Central (25%).	2.746	0.916	1	4
State	Categorical	49 contiguous states	27.871	16.324	1	56

Notes: *Income categories are used as build by the AC Nielsen with gaps of $3k until an income level of $15k that represents 6% of the sample. From $15k up to $50k, the gaps are equivalent to $5k and represent 45% of the sample. Then there are two categories with a gap of $10k ($50–60k and $60–70k), and finally from $70 to 100k and more than $100k these two last categories represent 25%; **the variable projection 61k is built in the AC Nielsen data set. This numeric variable is a projection factor that can be used to multiply quantities and expenditures to provide representative estimates of the U.S. aggregates. This projection is estimated as the ratio of the number of households and the households in the sample. Then it produces a demographic weighting and a household population projection. It also considers the correlation between household demographics and item purchases. It also includes a correction for the lower income households that are under-sampled due to difficulties in recruiting them (see Harris and Blisard 1995); ***the four U.S. Census regions in the United States.

participant households with 15 purchases, on average, during the sample period. The minimum purchases for a household is one, and the maximum is 176. Around 24% of the households have between two and five purchases registered in the sample, and approximately 70% of the sample corresponds to households with more than six purchases during 2005.

The chewing gum industry is highly concentrated with three firms (Wrigley, Cadbury,[4] and Amurol) dominating the market (see Table 2). The three-firm concentration ratio (CR3) is almost 90%. The Herfindahl–Hirschman Index (HHI), calculated by averaging across quarters and considering "other producers" as not having a significant share, is 4,338. According to the U.S. Department of Justice, Federal Trade Commission (2013) an HHI above 2,500 indicates high concentration. We do not have information about most of the small producers, but we know that all of the "other firms" have less than 1% of the market and do not affect the HHI substantially.[5]

Table 2 Market shares of gum sales for main suppliers by quarter (2005)

Producer	Quarter				Average
	1	2	3	4	
WRIGLEY'S	60.0	60.0	58.0	55.0	58.3
CADBURY*	25.0	28.0	29.0	29.0	27.8
AMUROL	1.8	1.5	1.8	2.5	1.9
OTHER PROD	14.0	11.0	12.0	14.0	12.8
C3**	**86.8**	**89.5**	**88.8**	**86.5**	**87.9**

Notes: *Subsidiary of Hershey's; **C3 is a concentration measure that adds up the market shares of the three firms with higher market share in the market. Source: AC Nielsen Homescan data set expanded by projection 61k.

4 Cadbury is a subsidiary of Hershey.

5 In fact, the calculated HHI considering "other producers" as just one firm with 11% of the market share is 4,481.

There are 67 gum varieties and brands in the data set that are manufactured mainly by three firms. Ninety-seven percent of gum sales were grouped into 66 different brands, the remaining 3% of the observations were gum products sold by brands with a frequency of less than 100 observations and they were aggregated into one label as "other brands." Thus, the total number of brands in the sample is 67. There are 4,403 observations, which from an unbalanced quarterly data panel for 49 states and 67 brands of gum.

Table 3 presents a description of the gum sales data by flavor. The average price (in dollars per ounce) paid is 82 cents with a standard deviation of 34 cents. Mint-flavored gum accounts for 51% of the total sales share. The highest average price is paid for sour-flavored gums at $1.20, followed by variety flavors.[6] The prices for mint, fruit, and spice are substantially lower than variety, sour, and other flavors.[7] In terms of the form or shape of the product, 94% of the gum is sold as pieces or sticks, the average price paid for the consumers is 87 cents for gum in pieces form and 75 cents for gum in stick form.

Table 3 Gum sales frequency, price, and standard deviation of price, by flavor

Flavor	Frequency (%)	Price ($/ounce)	Standard deviation unit price
Mint	51	0.790	0.280
Fruit	19	0.750	0.300
Spice	12	0.780	0.350
Variety	4	1.100	0.570
Sour	2	1.200	0.420
Other	13	1.000	0.420
Total	100	0.820	0.340

Source: AC Nielsen Data, calculations by the authors.

The mint-flavored category (51% of the sample) is divided into three sub categories, peppermint (19.6%), spearmint (13.8%), and other, not recognizable or artificial types of mint (17.9%).[8] The U.S Department of Agriculture, Economic Research Service (1995) reports that U.S. consumers prefer peppermint flavor to spearmint. Peppermint has more end uses than spearmint and is the number-one mint used in chewing gum production, which is the most important use of mint. Peppermint-flavored gum has a higher average price per ounce than spearmint, 84 cents versus 78 cents, respectively (see Table 4). By gum texture, the data consist of 83% chewing gum and 17% bubble gum. Almost 60% of the purchases registered in the sample are sugar free.

Table 4 Market share and average prices for each gum mint flavor

Mint types	Share	Avg. price ($/ounce)	Standard deviation unit price
Peppermint	19.54	0.84	0.31
Spearmint	13.85	0.78	0.27
Other mint	17.92	0.74	0.26
No mint	48.58	0.86	0.40

Source: AC Nielsen Data, calculations by the authors.

Based on the product diversity in our data, consumers' preferences are varied with respect to observable characteristics, such as flavor and form. Unfortunately, econometricians cannot observe all flavor quality attributes especially those in terms of strength and duration. Consequently, to model consumers' behavior, we rely on the subset of observable attributes in addition to our knowledge of the existence of other unobserved attributes. The AC Nielsen survey data enable us to observe consumers' socioeconomic characteristics, including household size, age, income level, presence of children in the household, and marital status (see Table 1 for descriptive statistics). These socioeconomic characteristics are used as covariates in the estimations.

3 Empirical framework

For purposes of estimation, we define the combination of brand and flavor as a product. This definition allows us to divide the set of differentiated products into subsets of homogeneous products. Examples of brands with their flavors are peppermint, used in brands such as Wrigley's Doublemint, Wrigley's Extra, and Wrigley's Freedent. Spearmint is used in brands such as Wrigley's spearmint or Care-Free Koolerz. Fruit flavors are used in brands such as Wrigley's Juicy Fruit, Adams Bubblicious, or Adams Dentyne Tango. Spicy flavors are used in brands such as Wrigley's Big Red or Adams Dentyne Fire. Also, we control for other elements such as the form of the gum and size of the package.

6 Variety stands for packages of assorted flavors.

7 The category "other" corresponds to those products with no available information about their flavor and for which was not possible to assign one by product name, such as: Gourmet, Holiday Stripe, Island Squeeze, Mystery Magic, and Radical Red.

8 This category includes types of mint agglomerate product flavors, such as "Arctic Chill" and "Cool Frost" in which it was not possible to determine clearly which type of mint was used.

The use of a discrete-choice model for estimation is advantageous for several reasons. First, it enables using aggregate-level information. In the context of the demand estimation in presence of unobserved product attributes, considering that unobserved product attributes are correlated with prices, endogeneity of prices can be addressed in an aggregate discrete-choice model. However, our purpose is to examine whether and how the observed price variation can be attributed to unobserved product attributes rather than to understand consumers' preference heterogeneity. In this sense, this study does not account for any individual consumer preferences. Second, using a discrete-choice model enables solving the dimensionality problem by projecting the products onto characteristics space. Third, it provides a tractable link between consumer theory and econometrics, enabling us to study markets with differentiated products in a structural model framework. Finally, since the estimation must account for non-observable product-specific characteristics or demand factors, the model allows for the possibility of prices being correlated with unobserved demand factors.

Assume the indirect utility of consumer i, for product j depends on the characteristics of the product and the consumer $U(x_j, \xi_j, p_j, v_i; \theta_d)$, where x_j and ξ_j are the observed and unobserved product characteristics, respectively. The price of each product is represented by p_j. Consumer-specific terms affecting utility are v_i and θ_d. The vector x_j represents the observable characteristics including flavor, if the product is sugar free, the form of the product, texture, and size. Vector ξ_j represents the unobserved product characteristics or the product attributes that the econometrician cannot measure or observe, but producers consider when setting their prices, and the consumers take into the account to make their choices, e.g. the average quality of each product derived from the individual consumer's valuation.

Consider a specification for the log indirect utility function where the unobserved consumer-specific taste parameters are captured by the error terms:

$$u_{ij} = x_j\beta - \alpha p_j + \xi_j + \epsilon_{ij} \quad [1]$$

The random term ξ_j can be interpreted as the mean of consumers' valuations of all the unobserved product characteristics, and the error term (ϵ_{ij}) represents the distribution of consumer preferences around ξ_j. However, we must consider that gum manufacturers know their own product characteristics, including those that are unobserved to the econometrician. Factors, such as higher quality ingredients, are considered to estimate their production costs, and they likely use this information to set the prices of their products. In this way, product prices (p_j) are likely to be correlated with those unobservable product quality attributes (ξ_j).

Another source of market price variation is the competition among products. Greater variety of products in a specific market implies intense price competition, affecting the prices of the products negatively. Recent literature has incorporated the product assortment as an endogenous variable to the model (see Draganska et al. 2008; Draganska, Mazzeo and Seim 2009; Mazzeo 2002; Seim 2006; Allender and Richards 2010). Taking advantage of the modeling strategy developed by this literature, we assume product assortment is exogenous and it affects the price.

Other strategies that affect product pricing may include the introduction of "private labels" which are brands developed by individual retailers as a strategic tool to compete with national brands. This strategy seems to increase retailers' profits and national brand prices (Bontemps, Orozco and Réquillart 2008; Cotterill and Putsis 2000). However, there are no records of private labels for gum products.

3.1 Multinomial and nested logit models setup

Assuming that the error term is independently and identically distributed (*i.i.d.*) across products and consumers as an "extreme value" distribution, we can represent the traditional market shares multinomial logit (MNL) model in the usual way:

$$s_j = \frac{e^{\delta_j}}{\sum_{k=0}^{J} e^{\delta_k}} \quad [2]$$

The functional form characterized in eq. [2] is the closed form of the MNL model, representing the probability of choosing good j among all other goods, including an outside good. Where $\delta_j = x_j\beta - \alpha p_j + \xi_j$ stands for the mean utility for product j, and $k = 0$ is the outside good that represents the consumer's expenditure in any other goods but gum. In this case, we use the state population as reference for potential market for the product to define the outside good.

Following Berry (1994), demand can be estimated by "inverting" the market-share equation to find the implied mean levels of utility for each good. A feature of the method is that allows for estimation by traditional instrumental variable (IV) techniques. By normalizing the mean utility of the outside good to zero and assuming the relation between observed and predicted market shares

is invertible, we can represent this relation in a linear form as:

$$\ln(s_{jm}) - \ln(s_{0m}) = \delta_{jm} = x_{jm}\beta - \alpha p_{jm} + \xi_j \quad [3]$$

On the left-hand side of the equation is the observed market share for each market m of each product j relative to the outside good, and on the right-hand side is the mean utility for product j.

As mentioned previously, gum manufacturers know their own product characteristics, including those unobserved for the econometrician, and they use this information to set the prices of their products in each market. In this way, product prices (p_{jm}) are likely to be correlated with those unobservable product characteristics (ξ_j), so the explanatory "observable" variables are not completely exogenous to the model, specifically the price, generating an identification problem due to endogeneity.

An additional potential problem with the MNL approach is that assumes that the probability of each alternative is related equally with the probability of other alternatives, this is the independence of irrelevant alternatives (IIA) assumption does not hold. An approach to this problem is allowing consumer tastes to be correlated across products j in a restricted fashion by using the nested logit (NL) modeling approach.

Preserving the assumption that tastes are distributed via extreme value, but allowing consumer tastes to be correlated across products j in a restricted way, we setup a nested logit (NL) model. We group gum products into exhaustive and mutually exclusive sets according to their flavor $g = 0, 1, \ldots, 6,$ where the outside good $g = 0$, is assumed to be the only member of group 0.[9] If we denote the set of products in group g as $\mathcal{J}_g$, for product $j \in \mathcal{J}_g$, the indirect utility of consumer i can be represented by:

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma_g)\epsilon_{ij}, \quad [4]$$

where $\delta_j = x_j\beta - \alpha p_j + \xi_j$ and ϵ_{ij} is i.i.d. extreme value. For consumer i, the variable ζ is common to all products in group g and has a distribution function that depends on σ_g, with $0 \leq \sigma_g < 1$. Parameter σ_g measures homogeneity of products within each group. As the parameter σ_g approaches one, the within-group correlation of utility levels goes to one, and as σ_g approaches to zero, the within-group correlation goes to zero.

We can interpret eq. [4] as a model involving random coefficients ζ_{ig} only on group-specific dummy variables. That is, if d_{ig} is a dummy variable equal to one if $j \in \mathcal{J}_g$ and equal to zero otherwise, we can rewrite eq. [4] as:

$$u_{ij} = \delta_j + \sum_g d_{ig}\zeta_{ig} + (1 - \sigma_g)\epsilon_{ij} \quad [5]$$

Thus, we can derive an analytic expression for mean utility levels in each market similar to the MNL model represented by eq. [3] with just one additional term for each group as[10]:

$$\ln(s_{jm}) - \ln(s_{0m}) = \delta_{jm} \equiv x_{jm}\beta - \alpha p_{jm} + \sum_{g=0}^{6} \sigma_g \ln(\bar{s}_{j/g}) + \xi_j \quad [6]$$

The new element compared with eq. [3] is the natural log of the within-group market share ($\bar{s}_{j/g}$). Using the NL model represented by eq. [6], the estimates of β, α, and σ_g can be obtained from a linear IV regression of differences in log market shares on product characteristics, prices, and the log of the within-group share.

However, as in the case of prices in MNL model specification, since the within-group share is also related with the unobserved characteristics via consumer preferences, $\bar{s}_{j/g}$ is endogenous, suggesting the need for additional exogenous variables that are correlated with the within-group share but not with the unobserved product valuation. In both specifications, the MNL given by eq. [3] and the NL given by eq. [6], the error term (ξ_j) is a structural component of the model and represents the average consumer valuation of the unobserved product attributes such as quality.

Summarizing, the differences between models is that the NL model relaxes the assumptions of the MNL model. The basic idea of the NL model is to extend the MNL model in order to allow groups of alternatives to be similar to each other in an unobserved way; that is, to have correlated error terms (Heiss 2002). In addition, as highlighted by Berry (1994), the NL model may be preferred when the researcher wants to model substitution effects depending only on predetermined classes of products, as it is the case in this project.

Basically, both models describe how market shares are generated from modelers' mind. Evaluated at the "true" value of parameters, the difference between the predicted (from the models) and observed market share depends only on the unobserved product attributes. These models can then be identified if we have a set of instruments, conditional on which the mean of the unobserved product attributes is zero. In both cases, the MNL and NL, eqs 3 and 6, the specification is linear in parameters, allowing the use of traditional IVs to eliminate endogeneity.

9 The rest of the categories was assigned as: fruit ($g = 1$), mint ($g = 2$), other ($g = 3$), sour ($g = 4$), spice ($g = 5$), and variety ($g = 6$).

10 For details refer to Berry (1994).

3.2 Instruments

Recall that the endogeneity originates in the relation between prices and the product-specific characteristics. We use a set of IVs that are related with the prices but not with the unobserved characteristics captured by the error term (ξ_j). As instruments, we use two groups of variables: the prices of the same products in other markets (Hausman 1996) and the distance from the production plants to account for the geographical location of each market. In this sense, prices of brand j in different markets will be correlated due to the common marginal costs, but they will be uncorrelated with the market-specific valuations of the product.[11] The distances from the production plants as IV for prices are proxy variables for the transportation costs, which are determinant of the supply function, so it is related with the prices but not with the unobservable market-specific valuation of the product.[12] Assuming the use of the correct IV the conditional mean of the error term equals zero given the repressors, so we can interpret the coefficients as the structural parameters of the model, see Cameron and Trivedi (2005). In the case of the NL model, where the within-group share is also endogenous, we use the within number of products as a variable of the market structure or the degree of competition. The market structure is correlated within each group, but not with the unobservable attributes.[13]

By construction, the number of products is inversely correlated with the within market shares, but uncorrelated with the market-specific valuation of the product. From economic theory, we can support the use of number of products within the market as instruments by framing the situation as a sequential-decision game. The decision made by each firm on the number of products is made before the realization of consumer preferences. At this stage, firms do not know consumers' preferences. Hence, the number of products is not related with the consumers' valuations of the unobservable product attributes. Even though this approach allows consumer tastes to be correlated across products, giving more reasonable substitution patterns compared with the MNL model, the grouping of products or the choice of sets is made a priori without any basis in theory or empirical support, but instead is justified by the market segments we are interested in analyzing, according to flavor. Given the set of instruments in both models, we have more moment conditions than parameters to be estimated, so it is a case of over-identified parameters that we estimate using a two-stage least squares (2SLS) estimator.

In terms of the identification strategy for the demographic variables, we rely on the variation across states in order to identify the parameters of the model.

4 Estimation results

In this section, we present the results of the estimations of the MNL and NL models. The results for the initial benchmark estimation are presented in column 1 in Table 5, these coefficients correspond to an estimation ignoring the non-orthogonal relation between covariates and error term, specifically the relation between prices and unobservable product attributes. The coefficient for

Table 5 Parameter estimates for gum demand using a multinomial logit model (MNL), multinomial logit model – instrumental variables (MNL-IV), and a nested logit model – instrumental variables (NL-IV)

Covariates	(1)MNL	(2)MNL-IV	(3)NL-IV
Price	−0.030	−0.266***	−0.130*
	(0.022)	(0.079)	(0.068)
Form (Piece)	0.008	0.269***	0.055
	(0.024)	(0.067)	(0.064)
Form (Stick)	0.011	−0.163**	−0.112*
	(0.038)	(0.074)	(0.065)
Size	−0.029***	−0.120***	−0.088***
	(0.010)	(0.026)	(0.023)
Sugar free	0.161***	0.001	−0.024
	(0.041)	(0.050)	(0.044)
Texture	0.184***	0.444***	0.006
	(0.026)	(0.058)	(0.062)
Mint	0.316***	0.752***	0.723***
	(0.057)	(0.053)	(0.071)
Incomes	−0.007	−0.093***	−0.072***
	(0.007)	(0.022)	(0.019)
Household size	0.016	0.022	0.022
	(0.011)	(0.030)	(0.025)
Children	−0.008	−0.032	−0.034
	(0.011)	(0.032)	(0.027)
Age	−0.036***	−0.076***	−0.053***
	(0.009)	(0.024)	(0.020)
Observations	4,403.000	4,403.000	4,403.000
R^2	0.118	0.305	0.529

Notes: Standard errors in parenthesis. *,**,***denote statistically significance at the 0.1, 0.05, and 0.01 levels.

11 We try different sets of instruments like: market structure (number of firms), rival products characteristics, and wages.

12 We estimate a first-stage regression using other prices and distances as determinants of prices for each product. In both cases, the coefficients were statistically significant.

13 Other instruments evaluated were rival product characteristics, some demographics, and advertisement expenditure. Even though the results were not very different from what we present, the tests show that these other instruments where weaker than those we use.

price is not statistically different from zero, which suggests that the demand for gum is completely inelastic to prices, which is inconsistent with economic theory. This anomaly is likely caused by an endogeneity problem and could be explained if unobserved product quality is considered. As in Trajtenberg (1989), prices appear to have a positive, or in this case nil, effect on consumers.

From the estimation results, it is also possible to see the positive marginal utilities generated by some of the product characteristics such as flavor, sugar, and texture. In the case of the size, the marginal utility is negative, suggesting that consumers prefer to carry smaller packages of gum. The second column in Table 5 presents the coefficients for the MNL model represented by eq. [3], now controlling for the presence of endogenous variables by using the prices of other products and the distances from the production plants as IVs, specifically to address the endogeneity of prices (column 2). In this estimation, the results are more compatible with economic theory. The coefficient for price is negative, as expected, and statistically significant.

Column 3 presents the estimation results for the NL model represented by eq. [6]. As in the previous case, we use prices of other products and the distances from the production plants as IVs to address the endogeneity of prices, but we also use the within-market number of products as instrument for the within-market shares to allow for product heterogeneity. In this case, the coefficient on price increases in absolute value when the model incorporates the IVs. An interpretation of this finding is that products with higher unmeasured quality characteristics sell at higher prices. These results suggest that ignoring the correlation between price and the demand error can lead to biased results. We notice that the coefficient for price in the first estimation is not statistically different from zero. In contrast, when endogeneity is accounted for with the use of IVs, the coefficients are negative and statistically significant, which is consistent with economic theory.

When accounting for endogeneity, the form of the product, the size, and the flavor are statistically significant in explaining the market shares. If the gum is in the form of a stick, the market share decreases with respect to the other forms of products. If the size of the product decreases, the market share of the product increases. If the product is mint flavored, the probability of a higher market share also increases. Consumers' income level and the age both decrease the probability of purchasing the good (see Table 5). The decrease in gum consumption with the income level might be related to substitution effects with respect to other type of non-gum breath freshener products, such as Altoids or Mentos, which are not included in the sample.

One can also understand the importance of unobservable characteristics by examining the fit of the logit demand estimation. The explanatory power of the model increases after product heterogeneity is allowed by the nesting strategy. In the case of the MNL-IV model, the R^2 statistic is 30%. For the NL-IV, the R^2 statistic increases to 52%, reducing significantly the percentage of the variance in mean utility levels associated with the unobserved product characteristics (see Table 5).

In Table 6, we present results indicating homogeneity of products within each group. The estimated values of the NL coefficients (σ_g) are all between 0 and 1 ($0 < \sigma_g < 1$), suggesting that the NL specification is consistent with the random utility maximization theory. Because a larger value of σ_g implies a stronger within-group correlation of utility, when σ_g is closer to one, consumers view products within the group as closer substitutes compared to products within another group with a smaller σ given the observed product attributes (such as price and size). As such, results in Table 6 indicate that products within the three groups – fruit, mint, and spice flavor, are more differentiated products for the consumers compared to products within other groups. This result seems to suggest that for consumers, two different mint-flavored gums are more different than two assorted flavored gums. We calculate the elasticities for different flavor profiles. We find that mint-flavored gum is slightly more inelastic (–0.100) than non-mint (–0.120), but the difference in the coefficients is not statistically significant.

Table 6 Estimated product homogeneity within each gum flavor group

Parameter	NL-IV
σ (g = Fruit)	0.148**
	(0.072)
σ (g = Mint)	0.156**
	(0.064)
σ(g = Other)	0.290***
	(0.076)
σ (g = Sour)	0.303***
	(0.101)
σ (g = Spice)	0.152*
	(0.083)
σ (g = Variety)	0.268***
	(0.078)

Notes: Standard errors in parenthesis. *,**,***denote statistically significance at the 0.1, 0.05, and 0.01 levels., controlling for producers, time period, and region.

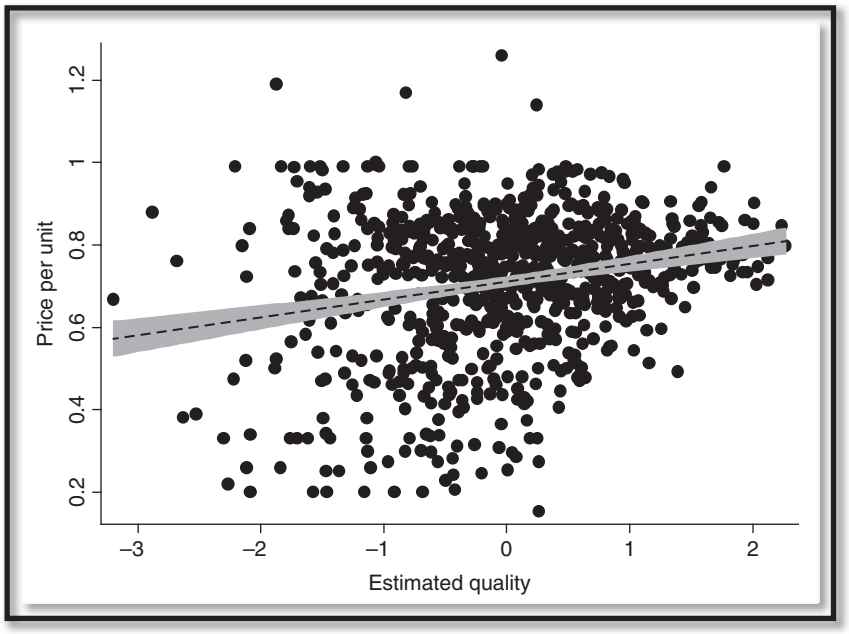

Figure 1 Relationship between mint-flavored gum prices and inferred quality
Source: AC Nielsen Data, calculations by the authors.

Income elasticities are small for all flavor categories, but statistically different from zero, on the order of 0.00023 and 0.0004 for mint and non-mint flavor categories, respectively. The small, almost nil income elasticity, supports the idea that budget constraints are not binding for gum products. We used household income for the estimation, which is high compared with the price of gum products.

After controlling for unobservable product characteristics in mint-flavored products, we observe some correlation between product prices and the product unobserved quality in the case of mint, 21%. This is the consumers' willingness to pay increase with the quality of mint flavor (see Figure 1). The study of gum demand using this approach contrasts with the larger computational burden of the Berry, Levinsohn, and Pakes (BLP) model (Berry, Levinsohn, and Pakes 1995). The NL model might be preferred when the researcher wants to model substitution effects depending only on predetermined classes of products (Berry 1994), such as flavors, in this case.

4.1 Counterfactual exercises

In this section, using the estimation results, we answer hypothetical questions about the mint-flavored gum market. Counterfactual scenarios help us understand the role of the unobservable product attributes such as quality on the market shares. We partition the sample according to the consumer's quality valuation in two sub groups: high-quality goods (HQ) and lower-quality goods (LQ), according to the distribution around the mean of the consumers' valuations of the unobserved product characteristics.

In these counterfactual exercises, we used the model represented by eq. [6] with parameter estimates from the NL-IVs shown in column 3 of Table 5. The hypothetical changes in quality do not respond to any particular research question or interest but to evaluate the effects of different quality distributions patterns on the market shares.

We find that the actual market shares for HQ is in average 3.5%, and for LQ is 1.2%. Considering changes in the factors that affect the consumer's quality valuation, e.g. the quality of the inputs used in the production process, we note that if the mint products are homogeneous ($\sigma_{\text{Mint}} = 1$), this is, if the consumer's quality valuation is the average for all products. The market shares of, each gum product is around 1.8%.

Consider a case not as homogeneous as the previous one. For this exercise, we estimate the parameter σ for each subsample, HQ and LQ, such as $\sigma_{HQ} = 0.285$ and $\sigma_{LQ} = 0.027$. If consumers' perceived differentiation reduces by half for HQ ($\sigma_{HQ} = 0.142$) products and doubles for LQ products ($\sigma_{LQ} = 0.054$), the market shares are also distributed more equally 2.3% for HQ and 1.3% for LQ gum products, than in the actual case.

5 Conclusions

In this study, we estimate demand for mint-flavored gum products with retail sales data while accounting for consumers' valuation of unobserved quality and other product attributes. We find that the unobserved product attributes are important to consider when estimating the demand for quality-differentiated products. With a nested logit model, almost 47%[14] of the variance in mean utility levels is associated with the unobserved product characteristics such as flavor quality. In terms of the parameter estimates, we find that the presentation of the product in terms of its form is important to consumers. Smaller packaged products have an advantage in terms of their market shares. Consumers' income level and age decrease the probability of gum purchase.

Our estimation results suggest that mint-flavored gum is more inelastic to changes in price than other flavors, and there exists an important variability in the valuation for quality among gum products. Given that mint-flavored gum is more inelastic to changes in prices than other flavors and the positive relationship identified between willingness to pay and unobserved quality, mint gum industry should be able to command a premium for higher quality product. The finding that U.S. consumers are willing to pay a premium for higher quality products is useful information for the U.S. mint oil industry, as they compete with cheaper foreign imports and lose negotiating power with gum manufacturers. Even though the tradeoff between lower costs and higher quality might not disappear, consumer preferences for high-quality products seem to guarantee the existence of a significant market share for high-quality mint oils in the U.S. gum market.

References

Allender, W. J., and T. J. Richards. 2010. "Market Diversion and Market Power: California Eggs." *Review of Industrial Organization* 36(1):37–58.

Berry, S. T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25(2):242–62.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4):841–90.

Bontemps, C., V. Orozco, and V. Réquillart. 2008. "Private Labels, National Brands and Food Prices." *Review of Industrial Organization* 33(1):1–22.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics Methods and Applications*. New York: C. U. Press, Ed.

Christensen, R. 2010. (Manager of the Far West Spearmint Oil Administrative Committee) interview by R. Karina Gallardo, August 10, 2010.

Chung, S. W., and K. Szymanski. 1997. "Effects of Brand Name Exposure on Brand Choices: An Implicit Memory Perspective." *Advances in Consumer Research* 24:288–94.

Cotterill, R. W., and W. P. Putsis. 2000. "Market Share and Price Setting Behavior for Private Labels and National Brands." *Review of Industrial Organization* 17(1):17–39.

Draganska, M., M. Mazzeo, and K. Seim. 2009. "Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing Decisions." *Quantitative Marketing and Economics* 7(2):105–46.

Draganska, M., S. Misra, V. Aguirregabiria, P. Bajari, L. Einav, and P. Ellickson. 2008. "Discrete Choice Models of Firms' Strategic Decisions." *Marketing Letters* 19(3–4):399–416.

Far West Spearmint Oil Administrative Committee. 2013. "Brief Summary of Spearmint Oil Marketing Order #985." Accessed April 5, 2013. www.farwestspearmint.org/history.htm.

Global Industry Analysts. 2011. "Gums: A Global Strategic Business Report." Accessed November 30, 2011. http://www.prweb.com/pdfdownload/8067444.pdf.

Gutierrez, B. P. 2005. "Determinants of Toothpaste Brand Choice in Urban Philippines." *Philippine Management Review* 12:45–71.

Harris, M.J. and N. Blisard. 1995. "Characteristics of the Nielsen Homescan Data." Working paper. Washington, DC: U.S. Department of Agriculture, Economic Research Service.

Hausman, J. 1996. "Valuation of New Goods under Perfect and Imperfect Competition. Edited by NBER." *Studies in Income and Wealth* 58:209–247.

Heiss, F. 2002. "Structural Choice Analysis with Nested Logit Models." *Stata Journal* 2(3):227–52.

Heritage Productions. 2000. "American Mint." Accessed March 15, 2012. http://www.mintstills.com/peppermintspearmint.html.

Kaya, T., E. Aktaş, İ. Topçu, and B. Ülengin. 2010. "Modeling Toothpaste Brand Choice: An Empirical Comparison of Artificial Neural Networks and Multinomial Probit Model." *International Journal of Computational Intelligence Systems* 3(5):674–87.

Mazzeo, M. J. 2002. "Competitive Outcomes in Product-Differentiated Oligopoly." *The Review of Economics and Statistics* 84(4):716–28.

Mint Dealer Company Representative, interview by R. Karina Gallardo, August 8, 2010.

Nielsen, A. 2005. "Homescan Data." U.S. Department of Agriculture Third Party Research Agreement, November 8, 2010.

Seim, K. 2006. "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." *The RAND Journal of Economics* 37 (3):619–40.

Shin, S. 2008. *Disentangling Preferences, Inertia and Learning in Brand Choice Models*. PhD dissertation. Rochester, NY: University of Rochester.

Trajtenberg, M. 1989. "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners." *Journal of Political Economy* 97:444–79.

Tremblay, V. J., and S. Polasky. 2002. "Advertising with Subjective Horizontal and Vertical Product Differentiation." *Review of Industrial Organization* 20(3):253–65.

14 This percentage is determined by the proportion of the variance that is not explained by the model ($1 - R^2$).

U.S. Department of Agriculture. Economic Research Service. 1995. "Peppermint and Spearmint: An Economic Assessment of the Feasibility of Providing Multiple-Peril Crop Insurance." Accessed August 1, 2012. http://www.rma.usda.gov/pilots/feasible/PDF/mint.pdf.

U.S. Department of Agriculture. National Agricultural Statistics Service. 2013. "Crop Production 2012 Summary." Accessed July 11, 2012. http://www.nass.usda.gov/Statistics_by_Subject/index.php?sector=CROPS.

U.S. Department of Justice. Federal Trade Commission. 2013. "Herfindahl-Hirschman Index." Accessed July 18, 2013. http://www.justice.gov/atr/public/guidelines/hhi.html.

Yang, Z., N. Zhou, and J. Chen. 2005. "Brand Choice of Older Chinese Consumers." *Journal of International Consumer Marketing* 17(4):65–81.