RAND Journal of Economics
Vol. 37, No. 2, Summer 2006
pp. 1–17

Estimating a continuous hedonic-choice model with an application to demand for soft drinks

Tat Y. Chan*

Using micro-level scanner data, I study empirically the consumer demand for soft drinks, which is characterized by multiple-product, multiple-unit purchasing behavior. I develop a continuous hedonic-choice model to investigate how consumers choose the best basket of products to satisfy various needs. My model's embedded-characteristics approach both helps to reduce the dimensionality problem in model estimation and generates flexible substitution patterns. Hence, the model is useful in application to data with many product choices that are correlated with each other at the individual level. The estimation results show that interesting substitutability and even a form of complementarity exist among soft drinks.

1. Introduction

■ In many micro-level scanner datasets, consumers buy multiple products and multiple units during each store trip. Conceptually, these product and quantity decisions are correlated at the individual level. For example, if a consumer buys one can of Coke, she may prefer not to buy a can of Pepsi, but the Coke purchase may have little effect on her choice to buy non-cola products such as Sprite or 7-Up. Empirically, though consumers may buy multiple brands, they do not buy all the brands available in a store, and hence we observe corner solutions. These observations require a strategy general enough to explain the correlations among decisions, but also simple enough to remain estimable when many products and corner solutions exist. In this article I develop a continuous hedonic-choice (CHC) model to study such multiple-product, multiple-unit purchasing behavior. I then apply the model empirically, using micro-level scanner data, to estimate consumer demand for soft drinks.

Conventional continuous-demand models treat each product as a separate entity in the utility function.[1] In contrast, the CHC model adopts the characteristics approach, as developed by Stigler (1945), Lancaster (1966), Rosen (1974), Gorman (1980), and others. The approach assumes that

* Washington University in St. Louis; chan@olin.wustl.edu.

This article is based on my 2001 Yale University Ph.D. dissertation. Special thanks go to my dissertation committee, Steve Berry, John Rust, Donald Andrews, and Donald Brown, for guidance and support. I am also grateful to the Editor Ariel Pakes and two anonymous reviewers for comments and suggestions. I have benefited from discussions with Greg Crawford, Gautam Gowrisankaran, Barton Hamilton, Glenn MacDonald, Brian McManus, Robert Pollak, Peter Rossi, P.B. Seetharaman, and participants in several seminars. Finally, I wish to thank David Bell for providing me with the soft drinks scanner data.

[1] The demand function is derived either from maximizing a random direct utility function subject to Kuhn-Tucker conditions (for examples, see Wales and Woodland (1983), and Kim, Allenby, and Rossi (2002)) or from using the dual

Copyright © 2006, RAND.

utility is derived from the characteristics or attributes[2] embodied in products, and that consumption is an activity producing characteristics from products. Hence, demand for products is only a derived demand. A key assumption of this approach is that major product characteristics are additive or combinable, and consumers choose their optimal basket of characteristics under the budget constraint. The characteristic approach has the advantages of helping to predict consumer reaction to new product introductions or quality changes, and of reducing the dimensionality problem by focusing on the characteristics space. Building on earlier characteristics models, this article develops an econometric model that is estimable with a large number of products and allows heterogeneity in consumer preference for product attributes.

A discrete-choice approach may also model multiple-product, multiple-unit purchases. For example, one may observe a consumer buying two units and then rationalize this behavior by assuming that the consumer has made two separate purchasing decisions. An innovative approach was developed by Hendel (1999) and applied by Dubé (2004) to study the behavior of households that purchase carbonated soft drinks. In their models, a consumer consumes only one product during each consumption occasion, but she buys multiple products and units in the store, expecting many such consumption occasions before the next shopping trip. One important simplifying assumption of this framework is that choices are independent across consumption occasions. Therefore, with random coefficients, this approach can parsimoniously generate flexible substitution patterns among products (for example, see Berry, Levinsohn, and Pakes (1995)). However, in this approach, products are substitutes only at the individual and aggregate levels. To allow for complementarity, one has to introduce state dependence in consumption choices.[3] In comparison, the CHC model allows one bundle of products to substitute for another bundle of different products in producing characteristics, so products within a bundle can be complements. Like other static utility-maximization models, CHC does not distinguish utility derived from each consumption occasion. But the utility function may be treated as a reduced-form approximation of the aggregate utility derived from all members of a household within a period. It is flexible enough to include other product attributes such as packages, whose major function may be unrelated to consumption tasks. While this approach is less "structural," it is useful in deriving substitution patterns among products in a parsimonious way.

A recent work by Gentzkow (2004) takes a different approach, also based on discrete choice. He extends the discrete-choice model by defining a utility function directly on product bundles. While this work allows for multiple choices and complementarities, it is not useful for analyzing markets with many products (such as soft drinks), because the number of potential bundles grows exponentially with the number of products and units that consumers can choose. In contrast, the CHC model can generate complementarities while feasibly allowing for multiple products and units. Its computational feasibility arises from the fact that the approach restricts utility to being a function of the total quantity of characteristics, and hence one does not need to compare the utility of all possible bundles to find the optimal bundle. This allows for numerical solutions to be computed quickly, even with a large number of products in the data, when the utility function is strictly concave. Also, the parameter space in the model correlates with the dimension of characteristics and not with products, and hence has a lower dimension.

While the CHC model has the above advantages, it has its own limitations. First, the model relies heavily on the assumption that characteristics are additive, which restricts its application to other categories. For example, two Boeing 707s do not make a Boeing 747, and two sedans do not make a bus.[4] One should use common sense or industry knowledge to scrutinize this

approach that first specifies an indirect utility function and then applies Roy's identity (for examples, see Lee and Pitt (1986, 1987), and Pitt and Millimet (2000)).

[2] "Characteristics" and "attributes" are hereafter used interchangeably.

[3] To the best of my knowledge, no one has empirically modelled state dependence when choices for consumption occasions are unobserved, due to difficulties in computation and identification.

[4] See Trajtenberg (1979) for discussions of modelling when characteristics are "non-concatenable" or "partially concatenable."

© RAND 2006.

assumption before applying the model to the data and, in cases where it is hard to decide, check to see if the model predictions are intuitively appealing and fit the data. Another limitation is that the model assumes perfect divisibility of products, so it is inapplicable to some categories where large discreteness exists. If the consumer choice set is limited, discrete-choice approaches such as Gentzkow's will be more applicable. Alternatively, one can use the traditional discrete-choice approach, or apply Hendel's approach and adopt his simplification strategy, if consumption tasks are independent. Finally, the current model does not allow for unobserved product characteristics; hence, some estimation results may be biased. I will further discuss the above limitations in the model section.

Complementarities arise in the CHC model because one bundle of products substitutes for another in producing characteristics. As an example, let bundle A consist of a 12-pack of Coke and a 2-liter bottle of Sprite, and let bundle B consist of a 12-pack of Sprite and a 2-liter bottle of Coke. Both bundles produce the same characteristics in terms of packaging and soft drink taste, and hence will substitute for each other. However, products within a bundle may be complements in the CHC approach: when a 2-liter bottle of Sprite in bundle A is on promotion, demand for a 12-pack of Coke in that bundle may increase because some consumers will switch from buying B to A. If complementaries arise because, for example, consuming more Sprite increases the marginal utility of consuming Coke as in traditional economics, the model has to allow for the interaction of the quantities consumed of the two attributes. But if characteristics are not additive, so that consumers do not consider the above bundles as similar, the CHC model will not have much explanatory power for complements. In this case, one may need to treat each product as a separate entity, and two products can only be complements if the coefficient of their interaction is positive in the utility function.

The rest of the article is structured as follows. In Section 2 I discuss the data and soft drink attributes. Section 3 discusses the model and some of its limitations. I also provide the details of the estimation algorithm. In Section 4 I discuss the estimation results and report the estimated patterns of substitutability and complementarity among soft drinks. Section 5 concludes.

2. The data

■ I use a scanner dataset collected by Information Resources Inc. (IRI) containing the soft drink purchasing data of 548 households in five stores in a large metropolitan area. The sample period is 104 weeks between June 1991 and June 1993, with 18,212 store trips altogether. There are 1,078 stock-keeping units (SKUs) in the data, with information including (i) marketing activities such as retail prices, features, and displays in each week; (ii) the purchase quantity of each SKU during each store trip; and (iii) household demographics such as household size and income.

Table 1 lists the top ten brands in total sales. Pepsi is the largest selling brand, closely followed by Coca-Cola, and R.C. is the third. Others include well-known brands such as 7-Up, Dr. Pepper, and Schweppes. Table 2 reports the market share and quantity sales of soft drink "types." The

TABLE 1 **Top Ten Brands with the Largest Market Share**

Brand	Manufacturer	Quantity Sale (in 32 oz.)	Market Share (%)
Pepsi/Crystal Pepsi	PepsiCo Inc.	31,343	11.9
Coca-Cola Classic/Cherry Coke	Coca Cola Co.	21,350	8.1
Diet Rite/Diet RC	Royal Crown Co.	18,154	6.9
Diet Pepsi/Diet Crystal Pepsi	PepsiCo Inc.	15,281	5.8
Diet Coke/Diet Cherry Coke	Coca Cola Co.	14,762	5.6
RC/Cherry RC	Royal Crown Co.	13,745	5.2
7-Up/Cherry 7-Up	Dr. Pepper/Seven-Up Corp.	13,125	5.0
Private Label	Private Label	11,581	4.4
Schweppes	Schweppes Inc.	8,833	3.4
Diet 7-Up/Diet Cherry 7-Up	Dr. Pepper/Seven-Up Corp.	8,436	3.2

© RAND 2006.

TABLE 2 **Sales and Market Share by Product Attributes**

Category	Quantity Sales (in 32 oz.)	Market Share (%)
Regular Cola	81,693	31.1
Diet Cola	61,018	23.2
Regular Flavored Soda	64,313	24.5
Diet Flavored Soda	36,075	13.7
Regular Mixer/Club Soda	17,434	6.6
Diet Mixer/Club Soda	2,298	0.9
Package (more than 1 unit)	152,631	58.1
Large-sized bottle (larger than 32 oz.)	95,429	36.3
Single unit and small-sized container (smaller than 32 oz.)	14,771	5.6

market share of cola and flavored drinks is much larger than that of mixer/club soda, and that of regular drinks is larger than that of diet drinks. The dominance of packed (defined as a packet with more than one unit of containers) and large-bottled (defined as a container size larger than 32 ounces) soft drinks in the market implies that packaging has a major effect on the demand.[5] Table 3 provides some summary statistics of household size and income of consumers in the data.

Table 4 demonstrates the significance of the multiple-product, multiple-unit purchasing behavior. Each entry in the table is the number of purchases corresponding to a specified number of SKUs and purchased units. One unit here can be a single can or bottle, or a pack. About two-thirds of purchases are related to multiple units, and about half are related to multiple SKUs. Indeed, about 80% of the total quantity of sales in the data is from multiple-unit purchasing, and 60% is from multiple-product (SKU) purchasing.

Table 5 demonstrates the behavior of variety-seeking over product attributes. In each row, $\Pr(i \mid j)$ represents the empirical probabilities of purchasing attribute i conditional on purchasing attribute j. There is a high probability of buying both large-bottled and packed drinks. Similar patterns can be found for buying both regular and diet drinks, or drinks with different flavors. Finally, about 10% of either Coca-Cola or Pepsi buyers will pick up the other. These data support the argument that consumers are buying a combination of product attributes for consumption.

□ **Soft drink attributes and products.** Before proceeding to the CHC model, two important issues related to the characteristics approach need to be resolved: (i) What are the attributes that will generate utility? (ii) Are these attributes additive? It is crucial to address these issues to determine whether the CHC model can apply to the data.

In principle, I can include all observed attributes in the model. For model parsimony, I will use only the attributes that have major impact on demand. Obviously, soft drink flavors, including cola, flavored, and mixer/club soda, are important, since they produce different tastes during consumption. Another central attribute is "diet," which produces the function of being healthy but may taste bad. Two brand names, Coca-Cola and Pepsi, are important in affecting consumer preference; otherwise only the cheapest cola will be chosen. These attributes are represented by the indicators *Cola*, *Flavor*, *Mixer*, *Diet*, *Coke*, and *Pepsi* in the model.

TABLE 3 **Some Household Demographics**

Variables	Mean	Maximum	Minimum	Standard Deviation
Household Size	2.682	6	1	1.441
Household Income ($)	34,940	80,000	5,000	24,112

[5] This is not just due to quantity discount. For example, the price per ounce of 12-pack soft drinks is on average higher than that of 2-liter bottles, but the former has a larger market share.

© RAND 2006.

TABLE 4 **Joint Distribution of Store Visits by Number of SKUs and Units Purchased**

SKUs\Units	1	2	3	4	5	6	7	8	9	10+	Total
1	6,561	2,132	520	597	119	266	40	132	12	254	10,633
2		2,215	836	486	249	189	115	92	49	176	4,407
3			638	385	190	149	64	82	37	147	1,692
4				257	116	117	48	56	40	73	707
5					100	68	38	40	22	62	330
6						38	21	19	14	55	147
7							12	15	6	42	75
8								5	4	25	34
9									0	18	18
10+										169	169
Total	6,561	4,347	1,994	1,725	774	827	338	441	184	1,021	18,212

Because a consumer does not go to store every time she wants a drink, she will buy multiple units in advance. Packaging produces the function of convenience for storing or carrying. In the data, large-bottled (i.e., bottles larger than 32 ounces) and packed (i.e., packages with more than one unit) drinks have more than 94% of the market. A consumer may purchase both packs and large bottles for different purposes.[6] But because a dynamic model with storage is beyond the scope of this article, my model assumes packaging simply as an attribute to approximate its impact on the demand. Packaging is represented by indicators *Pack* and *Large*. For model parsimony, other attributes, such as colors or calories, are not included. To avoid omitting important attributes, one has to use intuition and data checking when making choices.

Let us now turn to the issue of additive characteristics. The attributes I discussed above perform different consumption or storage functions. There is no obvious reason why their quantity cannot be added up in the utility function, which is a reduced-form approximation of the aggregate utility derived from all members of a household within a period. However, the additivity assumption also implies that products can be mixed together to produce attributes. It considers two bundles to be perfect substitutes if they produce the same amount of attributes. Using the example I discussed in the Introduction, bundle A that consists of a 12-pack of Coke and a 2-liter bottle of Sprite will be a perfect substitute for bundle B that consists of a 12-pack of Sprite and a 2-liter bottle of Coke. But this may not be true, at least to some consumers. One solution is to consider interactions of attributes: we may assume $Pack \times Coke$, $Large \times Coke$, $Pack \times Flavor$,

TABLE 5 **Empirical Conditional Purchasing Probabilities of Product Attributes**

Conditional Purchasing Probability of	Percentage
Packed | Large-bottle	19
Large-bottle | Packed	20
Regular | Diet	27
Diet | Regular	21
Flavored soda | Cola	33
Cola | Flavored soda	35
Cola | Mixer	29
Flavored soda | Mixer	27
Pepsi | Coca-Cola	8
Coca-Cola | Pepsi	10

[6] Consumption usage may also differ. Flavor erodes quickly once the unit is opened, so packs are more useful for individual consumption. A large bottle, on the other hand, is more useful for parties or big families.

© RAND 2006.

TABLE 6 **Descriptive Statistics: Market Share, Prices, Features, and Displays**

			Price (per 32 oz.)					
Product Name	Leading Example	Market Share (%)	Mean	Minimum	Maximum	Standard Deviation	Feature Mean	Display Mean
1. Diet Colas	Diet RC 12-oz. Can	.35	.58	.32	.82	.08	.06	.01
2. Large-Bottle Diet Colas	2-lit. Bottle Diet RC	4.20	.58	.35	.85	.17	.18	.10
3. Pack Diet Colas	12-Pack Diet RC	3.35	.81	.48	1.18	.15	.16	.13
4. Large-Bottle Diet Coke	21-lit. Bottle Diet Coke	1.72	.67	.32	.85	.15	.32	.12
5. Pack Diet Coke	12-Pack Diet Coke	8.16	.79	.43	1.20	.17	.26	.14
6. Large-Bottle Diet Pepsi	2-lit. Bottle Diet Pepsi	1.89	.68	.32	.85	.14	.26	.08
7. Pack Diet Pepsi	12-Pack Diet Pepsi	5.23	.84	.46	1.24	.22	.27	.14
8. Regular Colas	RC 12-oz. Can	.95	.56	.16	.64	.08	.07	.02
9. Large-Bottle Regular Colas	2-lit. Bottle RC	6.42	.47	.33	.85	.12	.19	.12
10. Pack Regular Colas	12-Pack RC	4.43	.77	.48	1.24	.17	.25	.17
11. Large-Bottle Regular Coke	2-lit. Bottle Coke	1.51	.64	.32	.85	.15	.34	.13
12. Pack Regular Coke	12-Pack Coke	7.48	.78	.43	1.19	.17	.30	.17
13. Large-Bottle Regular Pepsi	2-lit. Bottle Pepsi	2.82	.67	.32	.85	.15	.27	.09
14. Pack Regular Pepsi	12-Pack Pepsi	8.27	.83	.46	1.24	.20	.28	.15
15. Diet Flavors	Diet 7-Up 12-oz. Can	1.03	.87	.46	2.45	.23	.05	.04
16. Large-Bottle Diet Flavors	2-lit. Bottle Diet 7-Up	6.52	.61	.34	.85	.14	.18	.08
17. Pack Diet Flavors	12-Pack Diet 7-Up	6.87	.84	.58	1.20	.12	.17	.09
18. Regular Flavors	7-Up 12-oz. Can	2.02	.86	.31	1.79	.36	.07	.01
19. Large-Bottle Regular Flavors	2-lit. Bottle 7-Up	9.11	.58	.34	.83	.14	.21	.10
20. Pack Regular Flavors	12-Pack 7-Up	11.25	.83	.56	1.19	.15	.17	.11
21. Mixer	16-oz. Bottle Canada Dry	.18	1.94	1.21	3.33	.28	.09	.10
22. Large-Bottle Mixer	2-lit. Bottle Canfield	4.54	.65	.46	.87	.13	.10	.11
23. Pack Mixer	6-Pack Canfield	1.65	1.01	.79	1.35	.10	.02	.01

and *Large* × *Flavor* as four "attributes" instead. However, as I introduce more interactions, the model will converge to the traditional approach that treats each product as a separate entity, and hence lose its simplicity and intuition. In this article I adopt an alternative strategy by assuming an individual-specific taste[7] for each product. The fact that a consumer may prefer bundle *A* to bundle *B* is due to her unobserved tastes for the four products. Details about the individual-specific tastes for products will be provided later on.[8]

As discussed, there are eight observed attributes in the model: *Cola, Flavor, Mixer, Diet, Coke, Pepsi, Pack*, and *Large*. Since it is infeasible to estimate a model with 1,078 SKUs in the data, I group them into 23 "products," each with a unique combination of the above attributes. Hereafter, a soft drink product will refer to a group of SKUs sharing the same product attributes.[9]

Table 6 reports the 23 product names, their leading brands, and market share. I use the average prices, features, and displays, weighted by the total market share of all SKUs in each product, to construct product prices, features, and displays. Large-Bottle Regular Flavors and Pack Regular Flavors, including brands such as Sprite and 7-Up, have the largest market share. Pack Regular Pepsi, Pack Diet Coke, and Pack Regular Coke are other products with a large market share.

[7] This should not be confused with unobserved characteristics, which are not modelled in this article. I shall provide a detailed explanation later in the article.

[8] Other soft drink attributes may not satisfy the additive assumption. For example, caffeine-free colas have a very small market share in the data, while caffeine-free flavored drinks have a very large share. It is difficult to justify using the assumption of additive characteristics. I estimated a model that included *Caffeine* and found that the model without *Caffeine* provided a better fit to the aggregate market share.

[9] Estimation results may be biased due to the aggregation, but adding more brand names in the model can be inconsistent with the characteristics approach. One alternative is to assume latent brand attributes and then use a factor analytic model (for examples, see Elrod and Keane (1995) and Elrod (1998)).

© RAND 2006.

3. The model

■ The characteristic approach assumes that utility is derived from product attributes, and consumption is an activity producing attributes (outputs) from products (inputs). Let J be the number of products and C the number of attributes in the market. Also let $\mathbf{y}$ be a vector of the quantities purchased of products, and let $\mathbf{z}$ be a vector of the amount of attributes produced from $\mathbf{y}$. I assume a continuous quantity represented by standardized units (one standardized unit is 32 ounces, or about one liter in the model). The "consumption technology" is constant and similar to the settings in Lancaster (1966) and Gorman (1980), implying that the attributes are linearly additive. It is written as

$$\mathbf{z} = \mathbf{A}'\mathbf{y}, \tag{1}$$

where $\mathbf{A}$ is a $J \times C$ matrix. Each element indicates how much of an attribute can be produced by one unit of a specific product. They are either zero or one in the application. For example, a 2-liter bottle of Coke (two units of the product) produces approximately two units of the attributes *Cola, Coke*, and *Large*, and zero units of others.

The utility function of consumer i in period t is denoted by $V(y_{0,it}, \mathbf{z}_{it}; \boldsymbol{\xi}_i, \boldsymbol{\tau}_{it})$, where $y_{0,it}$ is the quantity of the outside good purchased, $\boldsymbol{\xi}_i$ a vector of household demographics, and $\boldsymbol{\tau}_{it}$ a vector of preference coefficients that are individual and maybe time specific. $V(\cdot)$ can be treated as a reduced-form expression of the aggregate household utility derived from all members during all future consumption occasions. It approximates how attributes produce consumption and other functions that in turn will produce utility. The consumer problem is to maximize this utility function under the budget and nonnegativity constraints. Let w_i be i's budget, $\mathbf{p}_t$ a $J \times 1$ vector of observed prices, and p_0 the price of the outside good. The objective function is the following:

$$\begin{aligned} &\max_{\{y_{0,it}, \mathbf{z}_{it}\}} V(y_{0,it}, \mathbf{z}_{it}; \boldsymbol{\xi}_i, \boldsymbol{\tau}_{it}) \\ &\text{s.t.} \qquad \mathbf{z}_{it} = \mathbf{A}'\mathbf{y}_{it};\ \mathbf{p}_t'\mathbf{y}_{it} + p_0 y_{0,it} = w_i;\ \mathbf{y}_{it},\ y_{0,it} \geq 0. \end{aligned} \tag{2}$$

I assume that $V(\cdot)$ is a quasi-concave, twice continuously differentiable function over $\mathbf{z}$. This assumption ensures the existence of a unique demand function, $y(\mathbf{A}, \mathbf{p}_t; \boldsymbol{\xi}_i, \boldsymbol{\tau}_{it})$. It also helps to compute a numerical solution for (2) quickly in the model estimation. Following standard assumptions, $y_{0,it}$ enters the utility function linearly, and p_0 is normalized to one. Substituting the budget constraint and the consumption technology into $V(\cdot)$, (2) can be transformed into the following:

$$\max_{\{\mathbf{y}_{it}\}} V(\mathbf{A}'\mathbf{y}_{it}; \boldsymbol{\xi}_i, \boldsymbol{\tau}_{it}) - \lambda_i \mathbf{p}_t'\mathbf{y}_{it}; \qquad \text{s.t. } \mathbf{y}_{it} \geq 0, \tag{3}$$

where λ_i is an individual specific, time-invariant price coefficient representing the marginal utility of income.

If I make the restriction that $\mathbf{y}_{it}[j] = 0$ and $\sum_{j=1}^{J} \mathbf{y}_{it}[j] = 1$, where $\mathbf{y}_{it}[j]$ is the jth row of $\mathbf{y}_{it}$, the CHC model becomes a discrete-choice model. Since sizes are different across products, I can introduce an additional variable $\mathbf{K}$ to measure container sizes. Then the above utility function will be $V(\mathbf{A}'(\mathbf{K} \cdot \mathbf{y}_{it}); \boldsymbol{\xi}_{it}, \boldsymbol{\tau}_{it})$, where "$\cdot$" is an operator of element-by-element multiplication between two vectors.

□ **Details of the model.** To incorporate the effect of household size on quantity purchased, a per capita quantity of soft drinks purchased is defined as

$$\overline{\mathbf{y}}_{it} = \mathbf{y}_{it}/(h_i)^{\alpha}, \tag{4}$$

where h_i is the household size of consumer i, and α represents the effect of household size. Correspondingly, I define $\overline{\mathbf{z}}_{it} = \mathbf{A}'\overline{\mathbf{y}}_{it}$ as the per capita quantity of attributes produced from $\overline{\mathbf{y}}_{it}$. Consumer i evaluates how much each household member on average will consume when making her purchasing decision, hence $\overline{\mathbf{y}}_{it}$ substitutes $\mathbf{y}_{it}$ in (3).

© RAND 2006.

To incorporate the effect of household income on the price sensitivity, the price coefficient is parameterized as a function of the per capita household income level w_i:

$$\lambda_i = \lambda / \left(\frac{\ln(w_i)}{\ln(w_0)} \right)^{\delta}, \tag{5}$$

where w_0 is the lowest per capita household income level in the data, λ is the price coefficient of consumers with income at w_0, and δ is the impact as household income grows.

A major identifying assumption is made in the model: I assume that there are no unobserved product characteristics. Therefore, there is no simultaneity issue for marketing variables such as prices, features, and displays. Although this assumption can be challenged, it solves a data problem: good instruments for retail prices that change weekly are not available.[10] Instead, I assume that there exist both time-invariant and time-varying individual-specific tastes for products, which are represented by two $J \times 1$ vectors ς_{1i} and ς_{2it} respectively.

I also assume that promotional activities such as features and displays will affect product tastes. I construct a $J \times 1$ vector of variables, $feature_t$, as a measure of the feature activity: The data indicate whether an SKU is featured ("1") or not ("0") in period t. For each product j, $feature_{jt}$ is equal to the sum of all of j's SKUs' feature indicators multiplied by their relative market share. A $J \times 1$ vector of variables, $display_t$, is constructed in a similar way. I further assume that product tastes enter the utility function linearly. As a result, the consumer problem in (3) can be rewritten as

$$\begin{aligned} &\max_{\{\mathbf{y}_{it}\}} U(\mathbf{A}'\overline{\mathbf{y}}_{it}) + (\varsigma_{1i} + \varsigma_{2it} + \gamma^f feature_t + \gamma^d display_t)'\overline{\mathbf{y}}_{it} - \lambda_i \mathbf{p}_t' \overline{\mathbf{y}}_{it} \\ &\text{s.t.} \qquad \mathbf{y}_{it} \geq 0, \end{aligned} \tag{6}$$

where parameters γ^f and γ^d represent the effects of feature and display. For simplicity, the parameters are assumed to be equal across consumers.

$U(\cdot)$ is a function corresponding to how consumer utility is derived from attributes. One specification used in the estimation model is

$$U(\mathbf{A}'\overline{\mathbf{y}}_{it}) = \sum_{c=1}^{C} u_c(\mathbf{A}'\overline{\mathbf{y}}_{it}) = \sum_{c=1}^{C} [\beta_{c,i}(\mathbf{A}'\overline{\mathbf{y}}_{it} + 1)^{\rho_c}], \tag{7}$$

where $\mathbf{A}_c$ is the cth column of the matrix $\mathbf{A}$. Parameters β and ρ imply the intercept and curvature in $U(\cdot)$: the marginal utility with respect to $\overline{z}_c = \mathbf{A}_c'\overline{\mathbf{y}}$ is $\beta_{c,i}\rho_c(\overline{z}_c + 1)^{\rho_c - 1}$, which equals $\beta_{c,i}\rho_c$ at $\overline{z}_c = 0$. To guarantee concavity, I assume that $\rho_c \in (0, 1)$ if $\boldsymbol{\beta}_{c,i} \geq 0$ (i.e., attribute c is preferred) and $\rho_c \geq 1$ if $\beta_{c,i} < 0$ (i.e., attribute c is not preferred). This specification allows attributes such as *Diet* to be preferred by some consumers but disliked by others for various reasons. For simplicity ρ_c is restricted to be equal across consumers. The "+1" in (7) ensures that marginal utility is finite at the zero-consumption level; hence consumers may not buy all attributes.

The way that each $u_c(\cdot)$ enters linearly in (7) underlies a weak separability condition on the characteristics space. One of the implications is that if a consumer prefers Coke to Pepsi, she also prefers Diet Coke to Diet Pepsi as the quantity of attribute *Diet* has no impact on her preferences for attributes *Coke* or *Pepsi*. One way to release this separability restriction is to include pairwise interactions between the quantities of product attributes in the utility function. Let K be the total number of interactions in the function, and let $(k, 1)$ and $(k, 2)$ be the indices of the two attributes in the kth interaction. Another specification in the estimation model is the following:

$$U(\mathbf{A}'\overline{\mathbf{y}}_{it}) = \sum_{c=1}^{C} u_c(\mathbf{A}'\overline{\mathbf{y}}_{it}) + \sum_{k=1}^{K} u_k(\mathbf{A}_{k,1}'\overline{\mathbf{y}}_{it}, \mathbf{A}_{k,2}'\overline{\mathbf{y}}_{it})$$

[10] Though quarterly or even longer period cost data are sometimes collected as instruments for price in the marketing literature, they can be poor instruments for prices in individual stores that vary weekly.

© RAND 2006.

$$= \sum_{c=1}^{C}[\beta_{c,i}(\mathbf{A}_c'\overline{\mathbf{y}}_{it} + 1)^{\rho_c}] + \sum_{k=1}^{K}[\beta_{k,i}(\mathbf{A}_{k,1}'\overline{\mathbf{y}}_{it} \cdot \mathbf{A}_{k,2}'\overline{\mathbf{y}}_{it} + 1)^{\rho_k}]. \tag{8}$$

Again, for simplicity, ρ_k's are restricted to be equal across consumers.

While the utility function should be treated as a reduced-form approximation of the aggregate household utility, it is important that the above specifications have properties that are consistent with the economics of multiple-product, multiple-unit purchasing. The underlying reason a consumer may buy multiple products is preference heterogeneity in different consumption occasions or among household members. The above specifications approximate this heterogeneity in the way that a consumer may have a high preference for different flavors or brands. A consumer may buy multiple units because she does not visit the store every time she wants to drink. When there is a promotion, she may buy more for stockpiling reasons or simply to substitute soft drinks for other drink categories. But she will not buy infinitely because there is either an inventory holding cost or a saturation level for consumption. In addition, the consumer may seek variety, so she may look for Sprite if she already has many units of Coke in her shopping cart. All of these behaviors are approximated in the way that the utility function is increasing but also concave in quantity purchased y. As the price of a product falls, its quantity purchased will rise at a decreasing rate.

Let $D = C$ under the specification in (7), and $D = C + K$ under the specification in (8). Individual specific $\boldsymbol{\beta}_i$ is defined by a random coefficients approach,

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0 + \mathbf{e}_i; \qquad \mathbf{e}_i = \boldsymbol{\sigma}_\beta \cdot \boldsymbol{\eta}_i, \tag{9}$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\sigma}_\beta$ are two $D \times 1$ vectors of parameters to be estimated, and $\mathbf{e}_i$ is a vector of stochastic deviates assumed to be i.i.d. over individual consumers and attributes. $\boldsymbol{\eta}_i$ is assumed to have the distribution of *normal*$(\mathbf{0}, \mathbf{I}_D)$, where $\mathbf{I}_D$ is an identity matrix with dimension D.

The two unobserved tastes for products, ς_{1i} and ς_{2it} in (6), are assumed to be i.i.d. over individual consumers and products. The latter is also assumed to be i.i.d. over time. To avoid the problem that a draw of either one is so high that demand becomes infinite, they are restricted to be negative:

$$\begin{aligned} \log(-\varsigma_{1i}) &= \sigma_{\varsigma 1} \cdot \boldsymbol{\upsilon}_{1i}, \\ \log(-\varsigma_{2it}) &= \sigma_{\varsigma 2} \cdot \boldsymbol{\upsilon}_{2it}, \end{aligned} \tag{10}$$

where $\sigma_{\varsigma 1}$ and $\sigma_{\varsigma 2}$ are two parameters to be estimated. Both $\boldsymbol{\upsilon}_{1i}$ and $\boldsymbol{\upsilon}_{2it}$ are assumed to be distributed as *normal*$(\mathbf{0}, \mathbf{I}_J)$, where $\mathbf{I}_J$ is an identity matrix with dimension J.

Finally, the set of parameters to be estimated using the specification in (7) is $\boldsymbol{\theta} = \{\alpha, \delta, \gamma^f, \gamma^d, \boldsymbol{\gamma}', \boldsymbol{\beta}_0', \boldsymbol{\sigma}_e', \sigma_{\varsigma 1}, \sigma_{\varsigma 2}\}$. Using the specification in (8), I will estimate additional parameters $\boldsymbol{\rho}$ and $\boldsymbol{\beta}_0$ for the interactions.[11]

□ **Limitations of the model.** As discussed, one of the major restrictions of the CHC model is the assumption of linear additive characteristics and perfectly divisible products. Even with individual-specific tastes for products, the model still predicts that two bundles are close substitutes if they produce the same amount of observed attributes. The applicability of the CHC model depends on the tradeoff between parsimony and feasibility in model estimation, and how close the model can approximate the data. It is important to use common sense or industry knowledge to scrutinize the assumptions and also to check whether model predictions are intuitively appealing and fit the data. For example, if the individual-specific tastes are estimated to have a large variation, it may imply that the additivity assumption is too restrictive. In this case the CHC model does not have a close approximation for the demand surface, and hence may not be applied.

[11] To reduce the number of parameters, I restrict that $\rho_{Large} = \rho_{Pack}$, and that ρ's for interactions in (8) are equal. Since it is unlikely to have diminishing brand name effect, I assume that $\rho_{Coke} = \rho_{Pepsi} = 1$. I also restrict that $\sigma_{Large} = \sigma_{Pack}$, $\sigma_{Coke} = \sigma_{Pepsi}$, and σ's for all interactions are equal.

© RAND 2006.

Another major limitation is that this article does not model unobserved product characteristics and hence treats other marketing variables as exogenous. As a result, the estimated price elasticities are potentially biased if demand shocks exist in the data that are correlated with pricing decisions. The setup in the model is mainly due to data restrictions. I try to control for seasonality shocks by estimating the model using subsamples from different periods in the data. This will be discussed in the next section.[12]

Finally, the CHC model uses a static utility maximization framework, hence ignoring the interesting issue of dynamic planning over time, which relates to the multiple-unit decisions. If consumers buy more during promotions only to stockpile for future consumptions, long-term price elasticities are not as large as in this model. To address this issue, a dynamic model that includes inventory-holding behavior has to be developed. This is outside the scope of the current article. Furthermore, without modelling the consumer decisions of where and when to buy, one cannot study the real impact of some promotional activities such as product features, of which the main purpose is to increase store traffic.

□ **Model estimation.** The utility function in the model is only defined up to a monotone transform. In the estimation model, I normalize the price coefficient parameter λ in (5) to one. Consumers with a higher income are expected to be less price sensitive.

I use the simulated method of moments (SMM), developed by Pakes (1984), Pakes and Pollard (1989), and McFadden (1989), in model estimation. Its main advantage is that estimators are consistent with a finite number of simulated draws. The SMM only uses moment conditions. The computation burden is vastly reduced because there is no need to evaluate the high-dimensional probability integrals. The estimator efficiency is not a major concern here, since I have more than 18,000 observations in data.

The estimation procedure involves a nested algorithm for estimating the parameters $\boldsymbol{\theta}$: An "inner" algorithm computes a simulated quantity purchased, which solves the problem in (6) for each trial value of $\boldsymbol{\theta}$, and an "outer" algorithm searches for the value of $\boldsymbol{\theta}$ that minimizes a criterion function. The inner algorithm is repeated every time $\boldsymbol{\theta}$ is updated by the outer algorithm.

The procedure for the inner algorithm is the following: Let $\boldsymbol{\varepsilon}_{it} = \{\boldsymbol{\eta}_i', \boldsymbol{\upsilon}_{1i}', \boldsymbol{\upsilon}_{2it}'\}'$ be the stochastic components for consumer i in period t. I draw $\tilde{\boldsymbol{\varepsilon}}_{it,1}, \ldots, \tilde{\boldsymbol{\varepsilon}}_{it,ns}$, where ns is the number of simulation draws,[13] from the distribution functions F_ε that are known from specifications (9) and (10). Conditional on parameters $\boldsymbol{\theta}$, covariates $\mathbf{X}_{it}$, and in each simulation draw $\tilde{\boldsymbol{\varepsilon}}_{it,s}$, I use a derivative search procedure called sequential quadratic programming (SQP) to solve the nonlinear constrained optimization problem in equation (6) for the optimal quantity purchased $\tilde{y}(\mathbf{X}_{it}; \boldsymbol{\theta}; \tilde{\boldsymbol{\varepsilon}}_{it,s})$. In this method, $\tilde{\mathbf{y}}$ is updated in a series of iterations beginning with a starting value (e.g., a vector of zeros).[14] Let $\tilde{\mathbf{y}}_s$ be the current value for iteration s; let the succeeding value be $\tilde{\mathbf{y}}_{s+1} = \tilde{\mathbf{y}}_s + \rho\boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is a direction vector, and let ρ be a scalar step length.[15] The procedure is repeated for every simulation draw to generate simulated data $\tilde{y}(\mathbf{X}_{it}; \boldsymbol{\theta}) = (1/ns) \cdot \sum_{s=1}^{ns} \tilde{y}(\mathbf{X}_{it}; \boldsymbol{\theta}; \tilde{\boldsymbol{\varepsilon}}_{it,s})$. When the utility function is concave and its Jacobian and Hessian matrices can be written down in analytical forms, each $\tilde{\mathbf{y}}$ converges quickly.

The outer algorithm searches for the estimator $\boldsymbol{\theta}_n$. Let $\mathbf{X}_{it}$ include a constant, relative income ratio $\ln(w_i)/\ln(w_0)$, household size h_i, and a vector of product prices $\mathbf{p}_t$, product features $\mathbf{f}_t$, and product displays $\mathbf{d}_t$. Under the identifying assumption that there is no simultaneity issue for marketing variables, $\mathbf{X}_{it}$ are valid instruments in model estimation. Let $\mathbf{y}_{it}$ be the quantity

---

12 There is a potential endogeneity issue caused by product aggregation. Since brand names are not included in the model as attributes, they may correlate with prices and hence bias estimation results. However, "products" in this model with the same flavors usually have the same bundle of brands. Therefore, I argue that the preference for other brand names has been included in the preference for flavors.

13 The number of draws for each observation is five in the estimation.

14 To make sure that the global optimum is achieved, I tried several starting values. In this application they all converged to the same solution.

15 A built-in algorithm is available in *Gauss*. Detailed discussions of the various procedures in computing ρ and $\boldsymbol{\delta}$ can be found in *Constrained Optimization*, 1995, Aptech Systems Inc.

© RAND 2006.

purchased, as observed in the data, and we have the following moment condition:

$$E(\mathbf{y}_{it} - \tilde{y}(\mathbf{X}_{it}; \boldsymbol{\theta}_0) \mid \mathbf{X}_{it}) = \mathbf{0}, \tag{11}$$

where $\boldsymbol{\theta}_0$ are the true parameters. The estimator $\boldsymbol{\theta}_n$ is obtained by the following:

$$\boldsymbol{\theta}_n = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min}\, Q_n(\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\Theta}{\arg\min} \left\| \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T_i} \sum_{t=1}^{T-i} [(\mathbf{y}_{it} - \tilde{y}(\mathbf{X}_{it}; \boldsymbol{\theta}))'\mathbf{X}_{it}] \right\|_{W_n(\boldsymbol{\theta})}, \tag{12}$$

where $\|\cdot\|$ is a norm function and $W_n(\boldsymbol{\theta})$ is a positive definite matrix. Details of how to compute the optimal $W_n(\boldsymbol{\theta})$ and the standard errors for the SMM estimator can be found in Gourieroux and Monfort (1996). I use the Nelder-Meade (1965) nonderivative simplex method to search for $\boldsymbol{\theta}_n$. The number of iterations required in this procedure rises rapidly as the parameter space grows; hence model parsimony is important.

4. The results

■ I estimate the model under three specifications for the utility function. The first, specification A, ignores product feature and display effects, and uses the weak separability assumption as in (7). Specification B includes product features and displays, and also uses specification (7). The last, specification C, includes both product features and displays, and includes the interactions of the quantity of attributes as in (8). For model parsimony, I will only choose important interactions: I first estimate the model under specification B, then check the implied substitution patterns among soft drinks to identify some pairs of product attributes that may cause seemingly strange relationships such as complements. By doing so, I choose three interactions in the specification: *Cola* × *Flavor*; *Large* × *Pack*; and *Coke* × *Pepsi*. Comparison of the results provides a useful robustness check.

As discussed earlier, I assume that there are no unobserved characteristics and hence no endogeneity issue. Still, seasonal and holiday factors may affect consumers' tastes for products and also promotion activities in stores.[16] Hence, there is a potential correlation between the time-varying tastes for products, $\varsigma_{2,it}$, and other marketing variables such as prices. To minimize this impact, I break down the data into two parts and estimate them separately. Model I uses the store visits during summers and during and before Thanksgiving, Christmas, and New Year, and model II uses the store visits during other weeks. Both models are estimated using the above three specifications.

Table 7 reports the estimates from the two models under the three specifications. Estimates are similar under different specifications; hence, I will only discuss the results in models I-C and II-C.

The coefficient INCOME is the parameter δ in equation (5), representing the effect of per capita income on price sensitivity. Positive estimates in both models show that richer consumers are less price sensitive. The coefficient FMY SIZE is the parameter α in (4), representing the effect of household size on total quantity purchased. Estimates in both models are positive but smaller than one, implying that larger households buy more, but at a decreasing rate. One possibility is that large households may visit stores more frequently.

The coefficients for COLA, FLAVOR, and MIXER are all significantly positive. FLAVOR is in general the most preferred attribute, followed by COLA. Positive estimates for MIXER show that mixer/club soda drinks are appealing. The fact that these drinks have a small market share is due to their high prices. Interestingly, the higher estimate in model I-C perhaps implies a higher consumption rate for mixer drinks during summers and holidays. The coefficient DIET is not significantly different from zero, implying that about half of consumers like diet drinks

[16] Chevalier, Kashyap, and Rossi (2003) find empirical evidence that stores tend to use soft drinks as a loss-leader pricing category during summers and holidays.

© RAND 2006.

TABLE 7 Parameter Estimates

	Model I						Model II					
	I-A		I-B		I-C		II-A		II-B		II-C	
	Estimates	Standard Error	Estimates	Standard Error	Estimates	Standard Error	Estimates	Standard Error	Estimates	Standard Error	Estimates	Standard Error
1. INCOME	.08312	.02487	.09897	.00919	.11474	.00954	.13821	.01229	.20607	.01208	.24214	.00887
2. FMY SIZE	.57192	.02642	.54759	.02134	.58508	.01612	.46630	.02301	.52820	.02188	.53837	.02211
3. COLA	1.29350	.01102	1.39962	.01680	1.46192	.01822	1.53213	.00636	1.73050	.04416	1.80770	.00948
4. FLAVOR	1.98559	.06658	2.08934	.07644	2.11577	.04055	2.05215	.10784	2.42628	.08585	2.46521	.06468
5. MIXER	1.01747	.12767	.70619	.13324	.55545	.10863	.36058	.09388	.13300	.03296	.11565	.03141
6. DIET	−.00068	.00094	−.00065	.00097	−.00058	.00112	−.00090	.00076	−.00076	.00109	−.00034	.00083
7. LARGE	−.19914	.00400	−.13625	.01676	−.12960	.01099	−.26667	.02418	−.31254	.01716	−.33667	.02048
8. PACK	1.36109	.01581	1.57569	.03982	1.62624	.01870	1.26748	.03436	1.18179	.02336	1.22060	.02694
9. COKE	.03568	.00026	.03952	.00025	.04842	.00043	.02835	.00027	.04163	.00018	.04209	.00028
10. PEPSI	.03098	.00025	.04651	.00033	.05614	.00046	.04222	.00027	.05565	.00021	.06141	.00032
11. COLA ∗ FLAVOR					.00918	.00015					.00228	.00009
12. LARGE ∗ PACK					.00612	.00017					.00415	.02035
13. COKE ∗ PEPSI					.00666	.00008					.00200	.00002
14. FEATURE			−.06153	.00063	−.06066	.00101			−.06714	.00049	−.05961	.00058
15. DISPLAY			.01009	.00086	.00905	.00081			.00789	.00025	.01360	.00072
16. σ(COLA)	1.17539	.00716	1.41892	.01193	1.48855	.00791	1.14122	.00527	1.22006	.00617	1.05418	.03144
17. σ(FLAVOR)	.09393	.04399	.11725	.04929	.17598	.02386	.10685	.05480	.17234	.04958	.18818	.02823
18. σ(MIXER)	.78912	.03570	1.27688	.07095	1.41023	.06852	.73852	.03189	.54693	.01263	.51885	.01453
19. σ(DIET)	.07252	.00217	.12868	.00316	.18430	.00479	.18303	.00625	.19919	.00572	.20299	.00543
20. σ(LARGE, PACK)	.85183	.00728	1.11059	.03760	1.12172	.01425	1.21537	.04712	1.24778	.02897	1.26348	.03099
21. σ(COKE, PEPSI)	.00377	.00020	.00187	.00038	.00128	.00027	.00784	.00042	.01442	.00019	.02931	.00032
22. σ(INTERACTIONS)					.00768	.00006					.00131	.00001
23. ρ(COLA)	.25000	.00390	.24632	.00338	.24807	.00236	.18817	.00265	.19038	.00409	.17967	.00236
24. ρ(FLAVOR)	.26099	.00935	.23835	.00862	.23817	.00419	.22063	.01083	.20075	.00697	.20280	.00545
25. ρ(MIXER)	.32102	.02610	.27817	.02352	.27351	.01967	.41009	.02794	.55658	.01383	.55574	.01434
26. ρ(DIET)	.74988	.00839	.72505	.00712	.69825	.00728	.72678	.00894	.68780	.00850	.68330	.00729
27. ρ(LARGE, PACK)	.41687	.00535	.41648	.00755	.41247	.00289	.45909	.00775	.45813	.00544	.45260	.00519
28. ρ(INTERACTIONS)					.39638	.00537					.22545	.01046
29. $\sigma(\rho 1)$	.02185	.00041	.01207	.00025	.01200	.00046	.01594	.00038	.02302	.00023	.02074	.00020
30. $\sigma(\rho 2)$	.13178	.00096	.10255	.00058	.11628	.00094	.09321	.00055	.08253	.00044	.08352	.00048

Model I uses observations during summer, Thanksgiving, Christmas, and New Year holidays; model II uses observations during other periods. INCOME is the coefficient of income effect on price sensitivity; FMY SIZE is the coefficient of household size on quantity purchased.

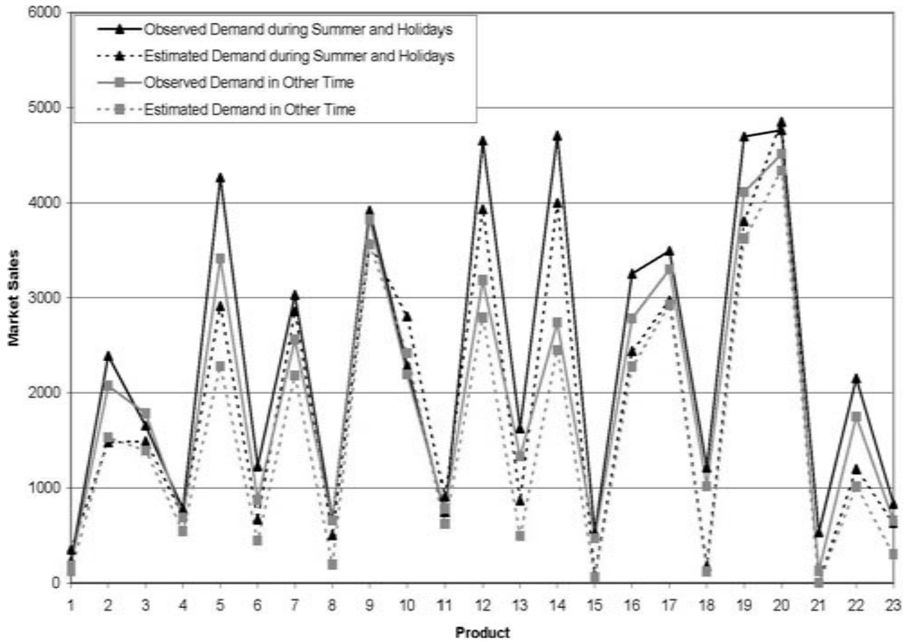
and half dislike them. The coefficient PACK is significantly positive, which is consistent with the fact that packs perform an important function: carrying and storing. The coefficient LARGE is negative, but the estimate σ_{LARGE} is also large. Despite the fact that taste is lost quickly once the bottle is opened, large bottles are still preferred by many consumers because they provide storage convenience. Also, packs and large bottles are much preferred during summer and holidays, perhaps due to higher consumption needs. Finally, if two attributes are substitutes in the utility function, the coefficient of their interaction will be negative. Results show that COLA ∗ FLAVOR, LARGE ∗ PACK, and COKE ∗ PEPSI are all positive, though they are also close to zero.

Surprisingly, the results show a significantly negative effect of feature, and only a small positive effect of display. In my model specifications, feature and display affect not only the decision to buy or not to buy, but also how much to buy. Suppose, for example, store features attract a lot of "cherry pickers" who only buy small quantities. Conditional on purchase, the average quantity purchased when the product is featured will be lower than when it is not, though the total quantity purchased is higher. The estimated feature effect, therefore, will be negative in order to fit the data. A similar explanation also applies to display. As a result, the estimates may

© RAND 2006.

FIGURE 1

OBSERVED AND ESTIMATED DEMAND IN DIFFERENT PERIODS



be biased. As I discussed previously, the absence of modelling the consumer decisions of where and when to buy is one of the limitations in the article.[17]

Figure 1 compares the observed and estimated aggregate demand of soft drinks derived from the models I-C and II-C respectively. Each product number corresponds to the product described in Table 6. The models fit the data quite well, though they underestimate the demand for some drinks with small market share. More important, by using the characteristics approach, the model is able to explain the differences in aggregate demand among soft drinks. For example, as Pack Regular Coke (product 12) is more popular than Large-Bottle Regular Coke (product 11), the model predicts that Pack Diet Coke (product 5) should also be more popular than Large-Bottle Diet Coke (product 4), which is exactly the case in the observed data. The model also predicts higher demand for soft drinks during summers and holidays. Since the number of weeks during these periods is about half of the number of weeks during other periods, the average weekly demand is much higher. These results provide some face validity for applying the CHC model to the data.

□ **Demand elasticities, substitutes, and complements.** There are no analytical expressions for demand elasticities. To investigate the substitution patterns implied by the estimation results, I numerically compute the elasticities based on the sample average prices. I use a bootstrap procedure with 50 iterations to compute the standard errors for elasticities. Table 8 reports the own- and cross-elasticites of eight regular cola and flavored drinks, based on the estimation results from model I-C. Results from model II-C are similar. Entries in the parentheses are the 10th and 90th percentiles of elasticities computed from bootstrapping, respectively.[18]

Own-price demand elasticities are very elastic, probably because the product aggregation level is low. For example, drinks with different packaging sizes are aggregated into different

[17] This can also be due to product aggregation. I estimated a discrete-choice model, which treats purchasing each brand as a separate choice and ignores the aspect of multiple units, using the same product aggregation, and find the coefficient FEATURE positive. Therefore, product aggregation does not seem to be the reason.

[18] Complete results are available from the author upon request.

© RAND 2006.

TABLE 8 **Some Own- and Cross-Price Elasticities of Soft Drinks**

	LBR Colas	PR Colas	LBR Coke	PR Coke	LBR Pepsi	PR Pepsi	LBR Flavors	PR Flavors
LBR	−4.96	.30	.47	.38	.42	.30	.64	−.32
Colas	(−4.99, −4.92)	(.30, .31)	(.46, .49)	(.37, .40)	(.41, .44)	(.28, .31)	(.61, .67)	(−.34, −.31)
PR	.24	−9.37	.07	2.67	.06	1.93	−.11	1.14
Colas	(.23, .25)	(−9.47, −9.28)	(.07, .08)	(2.62, 2.72)	(.05, .06)	(1.89, 1.97)	(−.12, −.10)	(1.10, 1.20)
LBR	2.25	.33	−9.90	.47	.91	.45	1.04	−.59
Coke	(2.13, 2.37)	(.31, .35)	(−10.01, −9.77)	(.45, .50)	(.89, .93)	(.41, .48)	(1.01, 1.08)	(−.64, −.52)
PR	.25	1.85	.10	−8.08	.05	1.89	−.11	.96
Coke	(.24, .26)	(1.82, 1.89)	(.09, .10)	(−8.13, −8.04)	(.05, .06)	(1.86, 1.91)	(−.12, −.10)	(.92, 1.02)
LBR	3.04	.41	1.70	.65	−11.09	.52	.96	−.46
Pepsi	(2.69, 3.15)	(.38, .43)	(1.65, 1.74)	(.60, .70)	(−11.35, −10.96)	(.49, .55)	(.91, 1.00)	(−.53, −.41)
PR	.28	2.29	.09	2.73	.07	−10.65	−.17	1.15
Pepsi	(.24, .31)	(2.24, 2.34)	(.08, .09)	(2.69, 2.77)	(.07, .09)	(−10.78, −10.54)	(−.19, −.15)	(1.11, 1.21)
LBR	.79	−.16	.29	−.22	.19	−.18	−5.49	.96
Flavors	(.76, .82)	(−.17, −.16)	(.27, .30)	(−.23, −.22)	(.18, .20)	(−.19, −.17)	(−5.57, −5.37)	(.94, 1.00)
PR	−.16	.59	−.06	.76	−.04	.54	.36	−6.44
Flavors	(−.17, −.15)	(.57, .61)	(−.07, −.06)	(.72, .77)	(−.04, −.04)	(.52, .56)	(.34, .37)	(−6.54, −6.32)

LBR = Large-Bottle Regular. PR = Pack Regular.

products, so there is packaging size switching in addition to the conventional brand switching. Demand elasticities in other flavor categories are in general lower than those in the cola category. Price elasticities of other colas are lower than those of Pepsi drinks, but the results are less clear when compared with Coke drinks. This is because, on the one hand, the average prices of Coke and Pepsi drinks are higher, so their elasticities should be higher; on the other hand, these brands have loyal customers, and hence their elasticities will be lower given the same prices.

Turning to the cross-elasticities, I define two soft drinks as "substitutes" if the demand for one soft drink rises in tandem with a price increase in the other, and as "complements" if it is the opposite. Therefore, a positive entry in column j and row k, $j \neq k$, in Table 8 indicates that products j and k are substitutes, and a negative entry indicates that they are complements.

Among substitutes, the results seem quite reasonable. In general, two drinks with the same flavor and packaging but different brands are the closest substitutes. For example, Large-Bottle Regular Pepsi and Coke, and Pack Regular Pepsi and Coke, are very close substitutes. Surprisingly, substitutability is low if the same two drinks differ in packaging, which perhaps indicates that the consumption and usage needs for cans and large bottles are different.

One of the most interesting results is that complements exist. In general, two drinks are complements if their flavors and packaging are different. For example, when a 2-liter bottle of Sprite is on sale, the demand for a 12-pack of regular Coke will increase (as Large-Bottle Regular Flavor and Pack Regular Coke are complements). The main behavioral explanation is that consumers seek variety in flavors and packaging. In the data, about 17% of sales are generated from buying both cola and flavored soda, and 12% come from buying both packs and large bottles. If a consumer who likes both Sprite and Coke has chosen a 2-liter bottle of Sprite, she will tend to buy packs, instead of large bottles, of Coke. In this case she chooses the bundle of a 12-pack of Coke and a 2-liter bottle of Sprite to substitute for other bundles, such as a 2-liter bottle of Coke and a 12-pack of Sprite, to produce her desired flavor and packaging attributes. Hence, the demand for a 12-pack of Coke will increase when a 2-liter bottle of Sprite is on sale, though the total demand for Coke of all packages may decrease. Note that complementarity exists despite the fact that interactions *Cola* × *Flavor* and *Large* × *Pack* have been included in the utility function; hence the result is not due to the separability assumption as in (7). Also note that the result does not require a 2-liter bottle of Sprite to be "mixed" with a 12-pack of Coke during consumption, since the attributes *Pack* and *Large* produce functions such as storage that are unrelated to consumption tasks.

© RAND 2006.

TABLE 9 Own- and Cross-Price Elasticities Among Soft Drink Categories

Product Category	Other Colas	Coca-Cola	Pepsi	Flavors	Mixer
Other Colas	−6.46	2.43	1.83	.66	.07
	(−6.51, −6.39)	(2.39, 2.48)	(1.79, 1.85)	(.64, .69)	(.06, .09)
Coca-Cola	2.84	−7.33	2.63	1.02	.11
	(2.79, 2.88)	(−7.37, −7.29)	(2.60, 2.66)	(.99, 1.07)	(.09, .15)
Pepsi	3.29	3.94	−9.93	1.12	.12
	(3.26, 3.33)	(3.89, 4.00)	(−10.05, −9.84)	(1.08, 1.15)	(.10, .15)
Flavors	.53	.60	.42	−4.20	.18
	(.50, .56)	(.57, .63)	(.39, .44)	(−4.30, −4.06)	(.15, .23)
Mixer	.37	.36	.25	1.03	−5.39
	(.34, .43)	(.33, .40)	(.23, .29)	(.91, 1.22)	(−5.98, −4.97)

The computed results are based on estimated model I that includes only observations during holidays and summer.

There is an alternative explanation for the finding of complements: demand for products can be correlated due to a time shock. For example, a consumer may have larger needs for soft drinks during holidays. She will search for the lowest-price store and purchase a larger quantity with several brands. This creates the appearance of complements.[19] The concern is valid due to the potential endogeneity issues in my model. I acknowledge that this is a major limitation in the result interpretation. As a partial solution, I estimate the model from two separate periods. If the time shock is purely due to seasonality, I will not find complements during off-seasons. Yet results show that in both periods the same pattern exists. Note that the complementarity results require that consumers desire not only multiple flavors but also different packages. An anecdotal support I found from my own experience was the presence of both cans and large bottles in the refrigerators of many families. One explanation I was given was that some parts of the refrigerator are designed for cans and other parts for large bottles. So the efficient way to use refrigerator space is to buy both.

□ **Defining markets of soft drinks.** The CHC model allows soft drinks to be imperfect substitutes so that consumers can choose multiple brands. This flexibility allows one to include in the model different types of drinks such as mineral water, sport drinks, and even alcoholic drinks such as beer, without defining a "market" *a priori*. One can apply the model to empirically test the boundary of a market, and use the result to address policy issues such as the market power of a manufacturer and the impact of mergers and acquisitions in an industry. As an illustration, I divide the 23 soft drink products into five main categories: Other Colas, Coca-Cola, Pepsi, Flavors, and Mixer, and compute their total own- and cross-price elasticities in Table 9 by assuming that the prices of products in each category are changed simultaneously by the same percentage.[20] Again, table entries in parentheses are the 10th and 90th percentiles of elasticities computed from a bootstrap procedure using 50 iterations.

Own-price elasticities are between −4.2 and −9.9, consistent with the oligopoly theory, which predicts prices falling in the range of elastic demand. The elasticities of Pepsi and Coke are higher than Other Colas, and that in turn is higher than Flavors and Mixer. It is also interesting to find that complements at a more aggregate product level disappear, because the substitution effect is more dominant than complementarity.

Coke, Pepsi and Other Colas are close substitutes with each other. Their cross-elasticities with Flavors are much lower, and that with Mixer is the lowest. Other soft drinks are low substitutes to Mixer (see the last column). Based on these results, one may define "colas" as one single market, and "mixer" as another. Flavored drinks fall somewhere in between. Since the substitutability

[19] I thank an anonymous reviewer for raising the issue and providing detailed examples.

[20] To save space, I only report the results using observations during holidays and summers. Results of using data in other periods are similar.

© RAND 2006.

between colas and flavored drinks is low, it seems safe to conclude that from the anti-trust perspective, Coca-Cola's ownership of Sprite is of less concern than, say, an acquisition of Royal Crown Co. Though beyond the scope of the current article, one may generalize these results to study the optimal pricing strategy and hence the impacts of acquisitions and mergers in the soft drink industry.

5. Conclusion

■ This article empirically studies the consumer demand for soft drinks, characterized as a multiple-product, multiple-unit purchasing behavior, using micro-level scanner data. Although this purchasing behavior is important, particularly in grocery shopping data, problems in modelling and estimation arise when a large number of products exist. The CHC approach models how consumers choose the best basket of products and hence how the quantity purchased decisions of each product correlate at the individual level. The model is built upon the assumptions that (1) utility is derived from the characteristics or attributes in products, (2) consumption is an activity producing characteristics from products, and (3) characteristics are additive in the utility function. These assumptions, though restrictive, help to reduce the dimensionality problem in estimation and generate flexible patterns of substitutability and even complementarity among products. Therefore, the model is particularly useful for the data where many possible "bundles" exist, and consumers demonstrate significant variety-seeking behavior over product attributes.

Own- and cross-price elasticities derived from the estimation results are mostly reasonable. A surprising result is that a form of complementarity exists among soft drinks. For example, a 2-liter bottle of Coke is a close substitute for a 2-liter bottle of Sprite but a complement to a 12-pack of Sprite. While an alternative explanation exists for the finding, the proposed model provides an intuitive link between these relationships and how products substitute for each other in producing characteristics bundles. A counterfactual experiment shows various degrees of substitution and hence the existence of local markets among soft drinks. By allowing products to be "imperfect" substitutes, the model helps to define market boundaries and hence derives useful policy implications.

References

Berry, S., Levinsohn, J., and Pakes, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63 (1995), pp. 841–890.

Chevalier, J.A., Kashyap, A.K., and Rossi, P.E. "Why Don't Prices Rise During Periods of Peak Demand? Evidence from Scanner Data." *American Economic Review*, Vol. 93 (2003), pp. 15–37.

Dubé, J.-P. "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks." *Marketing Science*, Vol. 23 (2004), pp. XXX–XXX.

Elrod, T. "Choice Map: Inferring a Product-Market Map from Panel Data." *Marketing Science*, Vol. 7 (1988), pp. 21–40.

——— and Keane, M.P. "A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data." *Journal of Marketing Science*, Vol. 32 (1995), pp. 1–16.

Gentzkow, M. "Valuing New Goods in a Model with Complementarity: Online Newspapers." Mimeo, Harvard University, 2004.

Gorman, W.M. "A Possible Procedure for Analysing Quality Differentials in the Egg Market." *Review of Economic Studies*, Vol. 47 (1980), pp. 843–856.

Gourieroux, C. and Monfort, A. *Simulation-Based Econometric Methods*. Oxford University Press: New York, 1996.

Hendel, I. "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns." *Review of Economic Studies*, Vol. 66 (1999), pp. 423–446.

Kim, J., Allenby, G.M., and Rossi, P.E. "Modeling Consumer Demand for Variety." *Marketing Science*, Vol. 21 (2002), pp. 229–250.

Lancaster, K.J. "A New Approach to Consumer Theory." *Journal of Political Economy*, Vol. 74 (1966), pp. 132–156.

Lee, L.F. and Pitt, M.M. "Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach." *Econometrica*, Vol. 54 (1986), pp. 1237–1242.

——— and ———. "Microeconometric Models of Rationing, Imperfect Markets, and Non-Negativity Constraints." *Journal of Econometrics*, Vol. 36 (1987), pp. XXX–XXX.

McFadden, D. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica*, Vol. 57 (1989), pp. 995–1026.

© RAND 2006.

Pakes, A. “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” *Econometrica*, Vol. 54 (1986), pp. 755–785.

——— and Pollard, D. “Simulation and the Asymptotics of Optimization Estimators.” *Econometrica*, Vol. 57 (1989), pp. 1027–1057.

Pitt, M.M. and Millimet, D.L. “Estimation of Coherent Demand Systems with Many Binding Non-Negativity Constraints.” Working Paper no. 99-4, Brown University, 2000.

Rosen, S. “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition.” *Journal of Political Economy*, Vol. 82 (1974), pp. 34–55.

Stigler, G.J. “The Cost of Subsistence.” *Journal of Farm Economics*, Vol. 27 (1945), pp. 303–314.

Trajtenberg, M. “Quantity Is All Very Well But Two Fiddles Don’t Make a Stradivarius.” Mimeo, the Maurice Falk Institute for Economic Research in Israel, 1979.

Wales, T. and Woodland, A. “Estimation of Consumer Demand Systems with Binding Non-Negativity Constraints.” *Journal of Econometrics*, Vol. 21 (1983), pp. 263–285.

© RAND 2006.