Marketing Science

Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks

Jean-Pierre Dubé,

Please scroll down for article—it is on subsequent pages

Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks

Jean-Pierre Dubé

Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637, jdube@gsb.uchicago.edu

For several of the largest supermarket product categories, such as carbonated soft drinks, canned soups, ready-to-eat cereals, and cookies, consumers regularly purchase assortments of products. Within the category, consumers often purchase multiple products and multiple units of each alternative selected on a given trip. This multiple discreteness violates the single-unit purchase assumption of multinomial logit and probit models. The misspecification of such demand models in categories exhibiting multiple discreteness would produce incorrect measures of consumer response to marketing mix variables. In studying product strategy, these models would lead to misleading managerial conclusions.

We use an alternative microeconomic model of demand for categories that exhibit the multiple discreteness problem. Recognizing the separation between the time of purchase and the time of consumption, we model consumers purchasing bundles of goods in anticipation of a stream of consumption occasions before the next trip. We apply the model to a panel of household purchases for carbonated soft drinks.

Key words: multiple discreteness; structural modeling; customer behavior; brand choice
History: This paper was received September 25, 2000, and was with the authors 11 months for 3 revisions; processed by Greg Allenby.

1. Introduction

On a given trip, consumers regularly purchase multiple products from the same category. We find evidence of such purchase behavior for carbonated soft drinks (CSDs), ready-to-eat cereals, canned soups, and cookies—four of the top five revenue-generating categories in the dry grocery department in U.S. food stores according to ACNielsen. For each of these categories, more than 20% of the shopping trips resulting in a within-category purchase involve the simultaneous purchase of at least two alternatives. In this paper, we focus on CSDs, which account for more than 10% of total sales in the dry grocery department. In a panel of household shopping trips, only 39% of the trips involving a CSD purchase result in the purchase of a single unit of a single UPC product. In fact, roughly 31% of the trips result in the purchase of two or more products, and the remaining 30% of the trips result in the purchase of multiple units. Table 1 reports on the distribution of total CSD brands and items per shopping trip.

Typically, researchers using household scanner panels estimate expected demand with discrete-choice models (DCMs), such as the conditional logit (Guadagni and Little 1983) and the multinomial probit (Chintagunta 1992, McCullough and Rossi 1994). In categories in which consumers regularly purchase assortments, imposing the single-unit purchase assumption of the DCM will generate incorrect consumer responses to marketing mix variables. For instance, in the context of quantity purchases, Chintagunta (1993) shows that brand choice alone in categories with multiunit quantity purchases only accounts for part of the total price response and, in some instances, generates inelastic price elasticities. In addition to providing incomplete estimates of consumer preferences, the misspecification of the DCM could also lead to incorrect managerial predictions for these categories. Finally, the DCM does not make use of potentially valuable information contained in the quantity data.

Following Hendel (1999), we use a structural approach that allows for a more general form of multiunit and multibrand shopping behavior in a single unified choice model. This structural approach to modeling the assortment choice follows a recent marketing tradition of using microeconomics to guide the specification of an appropriate model of consumer demand. For instance, Erdem and Keane (1996) model consumer learning about product quality, Gonul and Srinivasan (1996) model consumer price expectations, and Erdem et al. (2002) model both price expectations and inventory accumulation. Hendel's approach involves modeling assortment behavior as the outcome of several simultaneous discrete decision problems. Hence, he refers to this type of behavior as multiple discreteness.

Table 1   Carbonated Soft Drinks

Products/Units	1	2	3	4	5	6	7	8	9	10+	Total
1	20,652	11,238	1,447	2,454	245	454	33	282	19	215	37,039
2	0	6,928	2,215	1,817	436	464	146	259	45	166	12,476
3	0	0	1,322	768	302	247	114	109	45	130	3,037
4	0	0	0	335	165	109	63	77	28	69	846
5	0	0	0	0	51	69	27	18	16	41	222
6+	0	0	0	0	0	7	20	12	12	33	84
Total	20,652	18,166	4,984	5,374	1,199	1,350	403	757	165	654	53,704

In the current shopping context, this multiple discreteness arises for several reasons. Typically, marketers interpret shopping behavior in scanner panels as outcomes for a single consumer. In fact, many of the panelists are shoppers making decisions for the entire household. The consumption of the purchased goods typically occurs at some point after the trip. Therefore, the shopper must make multiple decisions in anticipation of the various occasions during which goods will be consumed. A separate choice is made for each anticipated consumption occasion. If preferences vary across these consumption occasions, then one would expect consumers to purchase a bundle of alternatives. Formally, the primitives of our shopping model involve a consumer optimizing a separate subutility function for each anticipated consumption occasion. The derived econometric model predicts the total expected vector of purchases.

Because the consumption occasions are not observed in the purchase data, we are not able to characterize the precise context in which alternatives are consumed. However, the marketing literature provides several explanations for why one would observe shoppers making multiple decisions. For instance, McAlister (1982) documents individuals seeking variety by switching their consumption of CSDs between several alternative flavors over time. At the time of purchase, this form of variety-seeking would induce consumers to select an assortment of alternatives. Similarly, if consumers are uncertain of their tastes at the future time of consumption, they may purchase an assortment to ensure they have the right product on hand (Hauser and Wernerfelt 1991, Simonson 1990, Walsh 1995). Alternatively, the shopper may make several decisions based on the varying tastes of several members of a household, such as children versus adults. Each of these scenarios provides a rationale for why consumers may need to make multiple decisions at the time of purchase. With standard scanner panel data, it is not possible to resolve any particular one of these theories. Instead, we view them as explanations for the multiple decisions in our shopping model.

We find that the proposed model provides a good fit of the aggregate purchases for each alternative. In terms of the characterization of multiple-item shopping, the estimates provide substantial evidence of both observed and unobserved heterogeneity. Demographics play a significant role in determining differences in tastes in addition to identifying differences in the assortment of total purchases on each trip. To help understand some of the underlying structure, we estimate several comparison models that relax certain assumptions. In general, we find that most of our model parameters are fairly robust to these assumptions. We also compare our results to a purely statistical model of "quantity-then-brand choice."

An attractive feature of the current approach is the structural derivation, linking the statistical model to consumer theory. First, we have an intuitive link between our estimated parameters and consumer preferences. When the models are used to conduct marketing simulations, the link to individual utility maximization also allows us to measure consumer welfare. The Hicksian compensating variation can be used to measure consumer willingness to pay in dollars to maintain the status quo versus a hypothetical change in the marketing or category conditions.[1] Finally, the derivation of CSD demand from consumer theory ensures that aggregate demand is "well-behaved," such that a retailer's corresponding profit function satisfies all the conditions necessary for deriving an optimal price vector. For instance, many researchers have used double-log approximations for aggregate demand. Once this specification is used to study category profitability, additional ad hoc restrictions may need to be imposed for the profit function to be maximized (Anderson and Vilcassim 2001). In this sense, a structural derivation has practical appeal.

The paper is organized as follows. The second section describes the model of individual choice and demonstrates its relationship to the standard DCM. In §3, we discuss the econometric specification and the estimation procedure. Section 4 describes the data. In §5 we also report on the results, including parameter estimates and substitution patterns. We also report

---

[1] For instance, Chintagunta et al. (2002) measure the impact of zone pricing on consumers. An earlier version of this paper measured the impact of product deletions on consumer welfare.

in §6 on the results of sensitivity analysis of the proposed model to various underlying assumptions, as well as results from a purely probabilistic comparison model. Finally, we present conclusions in §6.

2. Household CSD Demand

Our model is similar to that of Hendel (1999), who looks at the multiple discreteness in firms' profit-maximizing computer holdings decisions. We consider a random utility framework in which households make optimal shopping decisions subject to a budget constraint. For a given category on a given trip, the consumer anticipates several future consumption occasions and chooses some quantity of the optimal brand for each occasion. Allowing preferences to vary across these consumption occasions leads to consumers purchasing bundles of products and varying quantities of each. An interesting feature of Hendel's framework is that it can be shown to be a generalization of the typical discrete-choice model (e.g., logit or probit).

In practice, we do not observe the specific consumption occasions. Intuitively, we expect the source of these occasions to reflect such factors as many family members with varying tastes, the replenishment of overall household CSD inventory, and uncertain future tastes. However, without time-of-consumption data, we are unable to model the specific context in which products are consumed. Instead, we model the distribution of consumption occasions, consisting of the expected number of occasions and the tastes for each. The model first generates an integer number representing the number of consumption occasions for which a shopper must make purchases during the trip. Each of these consumption occasions has its own set of corresponding preferences. For each consumption occasion, the shopper chooses an optimal quantity of one of the alternatives. The model predicts the total purchase vector for a given trip, integrating across the decisions made for each of the anticipated consumption occasions.

Two recent studies have also modeled the incidence of multiple-item purchases.[2] Harlam and Lodish (1995) use a variant of the DCM to estimate multiple brand purchases, but not the quantity of each alternative chosen. Because the model formulation is not structural, this specification will not be able to provide profit and consumer welfare measurements in assessing marketing strategy. Kim et al. (1999) propose an alternative imperfect substitutes specification capable of addressing both the multiple brand and the quantity decisions. At the time of a trip, a household

optimizes a separate subutility function for each product alternative, rather than for each expected consumption occasion. The purchase of assortments of goods reflects heterogeneity of rates of diminishing marginal returns for the various alternatives. In contrast, we explain assortments by heterogeneity in tastes across expected consumption occasions. The imperfect substitutes model would be difficult to apply to CSDs, because the large number of product alternatives would require identifying a large number of subutility functions.

2.1. The Model

Because the expected number of consumption occasions on a given trip follows a count process, a logical starting point is to assume a Poisson distribution.[3] The total expected number of decisions a consumer must make during a trip is an integer drawn from a Poisson distribution. We allow the mean of this Poisson distribution to vary across consumers and across trips. Formally, during a shopping trip on date t, a household h purchases a basket of various alternatives in the category in anticipation of J_{ht} different future consumption occasions. J_{ht} is an integer value that is derived from a Poisson distribution with a mean λ_{ht}:

$$J_{ht} \sim P(\lambda_{ht}), \quad (1)$$

where $\lambda_{ht} = D'_{ht}\delta$, D_{ht} is a vector of household characteristics and shopping behavior that may vary over time, and δ is a vector of parameters.

For each of these $j = 1, \ldots, J_h$ (we drop the time subscript for expositional convenience) expected consumption occasions, the household generates utility from consuming from among the $i = 1, \ldots, I$ products in the category:

$$u_j^h(Q_{ij}, D_h, \Theta) = \left(\sum_{i=1}^{I} \Psi_{ij}^h Q_{ij}\right)^{\alpha} S_h,$$

$$j = 1, \ldots, J, \ h = 1, \ldots, H, \quad (2)$$

where $S_h = D'_h\gamma$. In the above specification, Ψ_{ij}^h represents the household's perceived quality for alternative i on consumption occasion j. The value Q_{ij} is the quantity chosen for alternative i. The parameter α captures the curvature in the utility function. So long as the estimated value of α lies between 0 and 1, the model maintains the concavity property needed for an interior solution.[4] S_h captures the effect of household characteristics on the scale of purchases, and γ is a vector of parameters. Because the utility

---

[2] Manchanda et al. (1999) investigate purchase incidence decisions across categories.

[3] In our results section, we discuss the sensitivity of our findings to the Poisson assumption.

[4] Kim et al. (1999) use a product-specific α to allow diminishing returns to differ across products.

function has a perfect substitute specification, a single alternative will be chosen for the consumption occasion. The model permits any positive quantity of this alternative because of the curvature, α. Moreover, because the consumer has such J_h decisions to make on each trip, the aggregate purchase vector may contain several alternative products and varying quantities of each.

We define perceived quality in the following manner:

$$\Psi_{ij}^h = \max(0, X_i' \beta_j^h + \xi_i)^{m_h} \qquad (3)$$

$$\beta_j^h = \tilde{\beta} + D_h' \mu + \Omega \nu_j^h. \qquad (4)$$

Quality has the typical hedonic structure where X_i is a vector of product i's attributes, β_j^h is the vector of household h's tastes for attributes on consumption occasion j, and ξ_i is a vector of product-specific fixed effects (brand intercepts). We use a random coefficients specification for the tastes in Equation (4). Vector $\tilde{\beta}$ captures the portion of tastes common to all households and consumption needs, and μ is a vector of interactions between demographics and tastes for attributes. Finally, Ω is a diagonal matrix whose elements are standard deviations, and ν_j^h is a vector of independent standard normal deviates. Thus, for each household, the taste vector will be distributed normally with, conditional on demographics, mean $\tilde{\beta} + \gamma D_h$ and covariance matrix $\Omega\Omega'$ (i.e., Ω contains the Cholesky factors of the covariance matrix of β). In addition, we include the term $m_h = D_h'\omega$ to capture potential differences in households' tastes for quality, where ω is a vector of parameters. Note that because m_h does not vary across brands, it allows households to purchase more expensive alternatives, regardless of the underlying product attributes. As such, this term introduces a vertical component to preferences. Finally, the specification allows for zero demand (no-purchase), depending on the values of the product valuations Ψ.

In formulating the household's decision problem at the time of a trip, we assume preferences are quasilinear to separate the utility from purchases of the I soft drink products and a composite commodity of other goods, z. Conditional on J_h, the total utility of household h at the time of a shopping trip is given by:

$$U^h = \sum_{j=1}^{J_h} u_j^h(Q_{ij}, D_h, \Theta) + z. \qquad (5)$$

The household's expenditure constraint is given by:

$$\sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij} + z \leq y_h,$$

where p_i is the price of product i and y_h is the household's total shopping budget. Given the specification

of the utility function, the expenditure equation binds and may be substituted into (5) to give:

$$U^h = \sum_{j=1}^{J_h} u_j^h(Q_{ij}, D_h, \Theta) - \sum_{j=1}^{J_h} \sum_{i=1}^{I} p_i Q_{ij} + y_h. \qquad (6)$$

Conditional on the number of anticipated consumption occasions, J_h, the household's problem will be to pick a matrix with columns Q_j ($j = 1, \ldots, J_h$) to maximize (6).

Given the additive separability of (6) across each of the J_h expected consumption occasions, we can treat each decision independently. The optimal quantity of brand i for occasion j solves the first-order condition:

$$\alpha (\Psi_{ij}^h)^\alpha (Q_{ij}^*)^{\alpha-1} S_h - p_i = 0.$$

Rewriting the first-order condition in terms of Q_{ij}^* gives:

$$Q_{ij}^* = \left(\frac{\alpha (\Psi_{ij}^h)^\alpha S_h}{p_i} \right)^{1/1-\alpha}, \qquad (7)$$

which is the optimal quantity of product i for consumption occasion j. The optimal quantity of a given brand is increasing in S_h, which is why we describe this term as impacting the scale of purchases. The fact that consumers must purchase integer quantities does not pose a problem, because the subutility functions (2) are concave and monotonically increasing in Q_{ij}. These properties ensure that we only need to consider the two contiguous integers to Q_{ij}^*. We then compare the $2 \cdot I$ potential quantities, picking the one yielding the highest utility. Each of these optimal quantities has a corresponding latent utility $u_j^* = (u_{j1}^*, \ldots, u_{jI}^*)$, where $u_{ji}^* = \max_Q u_j^h(Q_{ij}, D_h, \Theta)$. The perfect substitutability ensures that a household selects brand i if $u_{ji}^* = \max(u_{j1}^*, \ldots, u_{jI}^*)$. Households carry out a comparable decision for each expected consumption occasion.

For each trip, we observe the sum of all of these optimal quantities in the form of an aggregate purchase vector. The model's predicted value for this aggregate purchase vector has the following form:

$$EQ_h(D_h, X, \Theta) = \sum_{J_h=1}^{\infty} \sum_{j=1}^{J_h} \int_{-\infty}^{\infty} \cdots \int Q_j^{h*}(D_h, X, \beta_j^h, \Theta)$$
$$\cdot f(\beta \mid D_h, \Theta) p(J \mid D_h, \Theta) \, \partial\beta, \qquad (8)$$

where $f(\beta \mid D_h, \Theta)$ is the normal probability density function of the taste vector conditional on household characteristics and model parameters, and $p(J \mid D_h, \Theta)$ is the Poisson probability density function of the number of expected consumption occasions conditional on household characteristics and model parameters. Note that although we make distributional and functional form assumptions on the primitives of the model, econometrically, we estimate the expression (8), which is derived from the model.

2.2. Comparison with the Standard DCM

An interesting feature of the random utility framework in the previous section is that it is a direct extension of the standard DCMs. Disregarding the expected consumption occasions and assuming that consumers are restricted to single-unit purchases, then α no longer plays any role and (6) reduces to:

$$u_{hi} = X_i \beta S_h - p_i, \qquad i = 1, \dots, I.$$

In this formulation, we are no longer able to identify $m(D_h)$, so we set it to one for all households. We can also divide through by S_h to obtain:

$$\widetilde{u_{hi}} = X_i \beta - \frac{1}{S_h} p_i, \tag{9}$$

where the inverse of S_h is usually interpreted as the price-response parameter. Adding a random disturbance term directly in (9) gives the standard random utility DCM (McFadden 1981). The proposed model differs from the DCM in several ways. The curvature in the utility function enables consumers to purchase multiple units of a given product alternative instead of choosing either zero or one unit. The number of consumption occasions, generated by the Poisson distribution, implies that consumers solve multiple-choice problems on a trip. Allowing preferences to vary across these consumption occasions, a different alternative may be selected for each occasion. As a result, the total category shopping basket may include a variety of product alternatives and multiple units of each alternative chosen.

In general, existing marketing models are not suitable for categories exhibiting multiple discreteness. For instance, treating each unit purchased as an independent outcome from a DCM still conditions on the number of items in the shopping basket (e.g., the number of decisions made per trip). This model will not be able to predict expected demand out of the sample because the number of decisions would be unknown. Similarly, the "brand-then-quantity choice" specification of Krishnamurthi and Raj (1988) would require conditioning on the total number of brands chosen. Ignoring the structural derivation, we construct an alternative statistical model that captures the multiple discreteness and, at the same time, generalizes the standard DCM. In the Appendixes, we outline the derivation of a "quantity-then-brand choice" model. In the section on "Model Sensitivity," we compare the estimates of this statistical model to those of the proposed model described previously.

3. Estimation Procedure

In this section, we discuss the intuition for the econometric model we estimate. The two main features of

the methodology are the use of method of simulated moments and a nonparametric correction for serial dependence in the prediction errors. We defer the formal details regarding the estimation procedure to the Appendixes.

Referring back to §2.1, our model consists of the following system (where we have omitted the time subscripts):

$$EQ(D_h, X, \Theta) = \sum_{J_h=1}^{\infty} \sum_{j=1}^{J_h} \int \cdots \int_{-\infty}^{\infty} Q_j^{h*}(D_h, X, \beta_j^h, \Theta)$$
$$\cdot f(\beta \mid D_h, \Theta) p(J \mid D_h, \Theta) \, \partial \beta \tag{10}$$

$$\lambda_{ht} = D_{ht}' \delta \tag{11}$$

$$S_h = D_h' \gamma \tag{12}$$

$$m_h = D_h' \omega \tag{13}$$

$$\beta_j^h = \tilde{\beta} + D_h' \mu + \Omega \nu_j^h. \tag{14}$$

Our objective is to estimate (10), the I-dimensional vector of total purchases for each alternative on a given shopping trip. In addition to the parameters, Θ, the purchase vector is a function of data matrices D_h, household characteristics, and X, product characteristics.

To estimate the model parameters, we use the generalized method of moments (GMM) approach, constructing conditional moments based on (10). To handle the complex multivariate integrals, we simulate the moments using Monte Carlo methods (McFadden 1989, Pakes and Pollard 1989), which allows us to obtain consistent parameter estimates. We use 30 draws and assume this number is sufficient to eliminate any noticeable simulation error. For each store trip, we take 30 independent draws from the Poisson distribution to simulate the number of expected consumption occasions. For each of these draws, we then take $(N + I - 1) \times K$ draws from the standard normal distribution to simulate the taste coefficients for these occasions, in which N is the dimension of X and K is a sufficiently large number to place an upper bound on the number of occasions simulated for each household. These draws are then used to construct 30 simulations of the expected purchase vector for each trip. Simulated moments are obtained by averaging across the 30 simulated trip vectors. Technical details of the GMM procedure are provided in the Appendixes.

The panel structure of the data potentially introduces additional problems with the efficiency of our estimates. We attempt to control for both cross-sectional and intertemporal persistence in the data. To deal with the cross-sectional aspect of the data, we include several state variables, such as temperature and seasonal dummies, to capture contemporaneous

aggregate demand shocks that could affect households in a similar fashion. However, most households also have fairly long purchase histories, which could exhibit persistent unobserved shocks (McCulloch and Rossi 1994, Seetharaman 1999 provide parametric time-series methods for multiperiod probit models). In the GMM procedure, we use a nonparametric correction for serial dependence in the prediction error based on Conley (1999). In this way, we correct the standard errors of our parameter estimates for serial dependence. Details for this standard error correction are presented in the Appendixes.

3.1. Identification

We now discuss several identification-related issues that arise for the proposed econometric model. In the data, we observe total demand, but we do not observe the specific expected consumption occasions. Nonetheless, we are able to identify the process that generates these occasions. The main identification problem involves the distinction between a household purchasing, for example, five units of CSDs to satisfy five expected occasions versus five CSDs to satisfy a single occasion. Because the random tastes are independent across consumption occasions, a household with several decisions will tend to purchase several different alternatives. However, a household with a single consumption occasion will only purchase one alternative. The number of consumption occasions determines the joint distribution of the total number of units of CSDs purchased and the number of different brands.

Several features of the model help us identify this distribution. Note that the mean of the Poisson generating the number of expected consumption occasions, the scale of purchases and the perceived product qualities, λ_h, S_h, and Ψ_{ij}^h, respectively, are all functions of household characteristics, D_h. Suppose household size increases both λ_h and S_h. Allowing household size to influence both terms enables the model to distinguish between having a lot of consumption occasions and purchasing a large quantity for a given consumption occasion. Similarly, suppose income increases S_h and interacts positively with premium brands in Ψ_{ij}^h. Allowing income to enter both these terms enables the model to distinguish between the number of units purchased and the product selected for a given consumption occasion. Although several different sets of parameter values could give the same likelihood for expected total purchases, they would not have the same likelihood for the joint distribution of total products and total units purchased. Because the sample households tend to purchase baskets containing several different soft drink products, the data identify this joint distribution.

An additional identifying assumption we make involves the independence of tastes across consumption occasions. We implicitly rule out cross-occasion externalities. It is not immediately clear how our results would change if we allowed for correlation in tastes across consumption occasions. However, if we were able to characterize specific consumption occasions, one might observe a phenomenon such as selecting cola for one situation increasing the likelihood of purchasing a non-cola for another situation.

Finally, the large number of product alternatives in the CSD category makes it infeasible to estimate a full matrix of correlated product-specific random effects. These correlated random effects typically play an important role in allowing for flexible substitution patterns between product alternatives. For instance, one might expect Diet Coke to be a closer substitute to Diet Pepsi than regular Coke. If the price of Diet Pepsi rises, one would want the model to permit Diet Pepsi consumers to switch primarily to Diet Coke rather than to other nondiet alternatives. Rather than capturing this "closeness" via correlated brand intercepts, we use the Lancasterian *characteristics approach* (e.g., Berry 1994, Fader and Hardie 1996). In the model, we characterize product alternatives in terms of their underlying product attributes (as described in the section on "Data"). The random coefficients for the product attributes enable flexible substitution patterns. Suppose a household's utility-maximizing choice for a given expected consumption occasion is a 6-pack of Diet Pepsi. If a household exhibits strong preferences for low-calorie goods on a consumption occasion, the household will tend to substitute other low-calorie alternatives for Diet Pepsi. If the consumption occasion also exhibits strong preferences for a specific package size, such as the 6-pack of cans, then the household will tend to substitute other low-calorie 6-packs for 6-packs of Diet Pepsi. The advantage of this characteristics approach is the parsimonious representation of substitution patterns. Effectively, consumer preferences are projected onto product attributes, which vastly reduces the number of parameters needed.

4. Data

The scanner data, collected by ACNielsen, cover the Denver area between January 1993 and March 1995. These data include consumer information for a random sample of 1,920 households, as well as weekly store-level information for 58 supermarkets with more than $2 million all-commodity volume. The store-level information consists of weekly prices, sales, and feature and display activity for 26 diet and regular products with a combined share of 51% of the household-level category sales. The list of 26 products

consists of all universal product codes (UPCs) with at least a 1% share of total-sample CSD volume each. Of these 26 products, 12 are owned by Pepsi, 8 are owned by Coke, and 6 are owned by Cadbury-Schweppes. The household-level data cover all shopping trips for these items. For each trip, we know the date, the store chosen, and the quantities purchased. For each alternative available within the store, we know the prices and whether the product was featured in a newspaper or as an in-store display. Combining the store and purchase data sets, we observe the full set of prices and marketing mix variables for all the alternatives on a given trip.

We treat different package sizes of a given product as separate goods. Thus, our analysis is at the UPC level. In the beer category, Allenby and Shively (2001) view different product sizes as a nonlinear pricing scheme based on consumers' volume choice. In the CSD context, the inherent difference in storability of small cans versus large bottles leads us to view sizes as different products altogether.

For each shopping trip, we construct a quality measure for each product based on both fixed and time-varying attributes (reported as per 12-oz serving). The fixed attributes consist of the ingredients of the product, which we collect from the nutritional information printed on the product packages: total calories, total carbohydrates, sodium content (in milligrams), and a set of dummy variables that indicate the presence of caffeine, phosphoric acid, citric acid, caramel color, and clear. We also include package-size indicators: 6-pack of 12-oz cans (6×cans), 12-pack of 12-oz cans (12×cans), and 6-pack of 16-oz bottles (6 × bottles), omitting 67.6-oz bottles, because we include an intercept. The time-varying attributes are shelf prices and marketing mix variables: feature ads and displays. Because the A&W root beer and cream soda are indistinguishable in terms of these observable attributes and marketing variables, we combine them into one generic A&W brand.

We also include controls for household-specific loyalty to brands and specific products. We include a dummy variable, brand loyalty, indicating whether the brand was purchased on the previous trip. We also include a dummy variable, product loyalty, indicating whether a specific product was purchased on the previous trip. This second variable distinguishes loyalty to a brand (e.g., Coke vs. Pepsi), opposed to loyalty to a specific package size of a brand, 6-pack of Coke versus a 12-pack of Coke. Linking these parameters back to the economics of consumer choices implies that consumers are myopic. Implicitly, consumers condition on the past but do not account for the impact of current decisions on future choices. It is difficult to provide a rational justification for these parameters when consumers are forward looking (see

Table 2   Descriptive Statistics (Averaged Across Trips)

Variables	Mean	Standard Deviation	Minimum	Maximum
Kids	0.3865	0.4870	0	1
Family size	2.6976	1.4034	1	9
Income bracket	4.2470	1.9616	1	9
Female under 35	0.1964	0.3973	0	1
Time between trips (days)	6.8498	13.7602	0	763
Max. temperature (F)	64.6149	19.8264	8	99
Holiday	0.1747	0.3797	0	1
Price ($)	2.1515	0.3782	0.2	6
Feature ad	0.3203	0.0579	0	1
Display	0.4174	0.0503	0	1

Chintagunta et al. 2001 for a discussion). Despite the theoretical limitations, they have been found to improve the ability to predict purchases (Guadagni and Little 1983, Erdem 1996, Keane 1997).

Summary statistics of the demographic variables and time-varying product attributes used in the estimation appear in Table 2. Table 3 breaks down the fixed characteristics by flavor group, providing a rough sense of the relative positions of the different products in attribute space. These characteristics pick up such differences as the fact that Diet Coke, which contains citric acid, does not have the same flavor as regular Coke, which does not contain citric acid. The nondiet colas, lemon/lime, and pepper drinks are quite similar, with around 150 calories on average. The root beers and the citrus beverages are substantially higher, with about 170, and the new age are substantially lower, with 120. Phosphoric acid is used in all regular colas, all the diet colas, and in all the peppers. Although citric acid is found in all the fruit drinks, it is also used in many of the other products. The caramel and clear attributes span all of the products except for the citrus (Mountain Dew), which are yellow.

5.   Results

5.1.   Baseline Model
We now present the results from our base model. In the following section, we discuss the sensitivity of our model to some of the specification and distributional assumptions. Model parameter estimates appear in Tables 4 and 5. We begin by discussing results for the means and standard deviations of random taste parameters in (14). These are the elements of the vector β in Equation (3).[5] As expected, both feature ads and displays have a strong positive influence on perceived product quality. The loyalty

---

[5] Mean tastes for fixed product attributes are obtained by regressing product fixed effects on the measured attributes. The regression uses the estimated variances of the fixed effects as weights.

Table 3   Continuous Attributes by Flavor and Diet Versus Regular (Averages)

Flavor	Continuous Variables			Indicator Variables					
	Calories	Mg Sodium	Carbohydrates	Caffeine	Phosphorus	Citric Acid	Caramel	Clear	No.
Cola									
Regular	150 (7.5)	40.5 (7.4)	41 (1.51)	7	7	4	7	0	7
Diet	0 (0)	34.7 (8.7)	0 (0)	6	9	9	9	0	9
Lemon/lime									
Regular	143.3 (5)	61.7 (16.4)	38.333 (0.5)	0	0	2	0	2	2
Diet	0 (0)	35 (0)	0 (0)	0	0	1	0	1	1
Root beer									
Regular	168.3 (4.1)	44.2 (14.6)	44.8 (1.5)	0	0	0	1	0	1
Citrus									
Regular	170 (0)	70 (0)	46 (0)	3	0	3	0	3	3
Pepper									
Regular	148.6 (3.8)	45.7 (8.9)	35.1 (15.5)	3	3	0	3	0	3

parameters, brand loyalty, and product loyalty show that conditioning on past brand choices and specific UPC choices provide information in predicting current preferences. These parameters suggest that loyalty to a specific brand might be stronger than loyalty to a given UPC. For instance, consumers are slightly more loyal to Coca-Cola in general than to a specific package size of Coca-Cola. Interestingly, their effect sizes are quite small, suggesting that consumer tendencies toward specific products may be driven more by preferences than habit. We also observe significant unobserved heterogeneity in consumer perceptions of product-specific quality (the intercept term). Although we find that the mean household places

little weight on whether a CSD is diet, we also find a substantial amount of heterogeneity. In particular, the distribution suggests an almost even split between positive and negative valuations of diet. At the same time, households place weight on higher energy (carbohydrate) beverages.

Previous marketing studies have found seemingly contradictory results with regard to the effects of demographics. With a few exceptions, the research generally finds weak and inconsistent effects (discussed in Kalyanam and Putler 1997). In the case of CSDs, however, we expect different-flavored products to cater to different demographic segments. For instance, households with a female head under 35 years old tend to have higher preferences for diet products.[6] In fact, we might find additional explanatory power from dummies such as female head with a college degree.[7] Similarly, larger households place slightly more weight on products with more 12-oz servings, such as the 12-pack (as captured by hh size ∗ serv). Households with kids place a higher weight on products with caffeine than those without. Despite controlling for demographic interactions, we still find evidence of unobserved heterogeneity in tastes for package size (6-packs, 12-packs, ...) and diet, suggesting that demographics alone are insufficient to explain taste differences.

We now move on to Table 5, which contains the parameters characterizing the mean of the Poisson distribution, vector λ_h in Equation (11), the exponent m_h in Equation (13), the scale component S_h in Equation (12), and curvature parameter α_h. Similar to

Table 4   Estimated Taste Parameters of the Quality Function Ψ

Variables in Equation (14)	β Mean (S.E.)	Ω Standard Deviations (S.E.)
β: Feature ad	1.52 (0.03)	0.06 (0.01)
β: Display	2.00 (0.04)	0.79 (0.05)
β: Brand loyalty	0.27 (0.03)	—
β: Product loyalty	0.02 (0.16)	—
β: Constant	1.02 (0.16)	1.88 (0.03)
β: Diet	−0.06 (0.01)	0.26 (0.02)
β: Sodium	−0.03 (0.00)	—
β: Carbs	0.23 (0.02)	—
β: Caffeine	0.47 (0.04)	—
β: Phosphorus	−0.04 (0.07)	—
β: Citric	0.04 (0.04)	1.23 (0.04)
β: Caramel	0.14 (0.11)	0.22 (0.02)
β: No color	0.77 (0.07)	—
β: Cans × 6	1.60 (0.02)	1.19 (0.11)
β: Cans × 12	0.57 (0.03)	0.30 (0.02)
β: Bottles × 6	0.72 (0.11)	0.19 (0.02)
μ: Kids ∗ Caffeine	0.40 (0.03)	—
μ: (Household size ∗ Servings)	0.01 (0.00)	—
μ: Female head < 35	1.13 (0.03)	—
Store trips	169,788	

Note. S.E. = standard error.

[6] Diet Pepsi was reintroduced in Europe with double caffeine as Pepsi Max to overcome its "feminine" image.

[7] "Just who's buying all these soft drinks, anyway?" *Beverage Industry*, 84(3), 1993.

Table 5   Estimated Nonlinear Coefficients in Utility Function

Variables in Equations (11), (12), and (13)	Coefficient (S.E.)
λ: Kids	0.336 (0.001)
λ: Family size	0.041 (0.000)
λ: Time since last CSD	0.001 (0.000)
λ: Time since last trip	−0.003 (0.000)
λ: Temperature	0.004 (0.000)
λ: Holdiay	0.009 (0.000)
s_h: Constant	1.549 (0.104)
s_h: Family size	0.958 (0.049)
s_h: Time since last trip	0.012 (0.003)
s_h: Time since last CSD	0.013 (0.003)
m_h: Income	2.305 (0.172)
α	0.009 (0.001)

Note. S.E. = standard error.

Kalyanam and Putler (1997), we find that demographics are key for identifying product holdings. We are able to capture this additional role for demographic variables by using the quantity information, which is omitted in standard DCM models of brand choice. The expected number of consumption occasions, λ_h, depends positively on the presence of children and on family size. Thus, larger households with children should have more expected consumption occasions and, in turn, should purchase more brands on a given trip. Holiday weeks (such as Christmas, Labor Day, and Memorial Day) exhibit the expected large positive effect on needs. Similarly, temperature increases the number of expected consumption occasions. Finally, the time since the last CSD purchase increases λ_h, but the time since the last shopping trip seems to offset this effect.

The scale of purchases, S_h, also increases with family size. Therefore, one would expect larger households to purchase larger quantities of the brands selected. Unlike the positive relationship with λ_h, which increases the scope of products purchased, S_h impacts the actual quantity, as seen in (7). Similarly, the scale of purchases increases with both the time since the last trip and time since the last CSD purchase. The strong positive effects are consistent with the notion that household inventories have depleted and need replenishing. The vertical component, m_h, increases with income, suggesting that households with higher income have a higher taste for quality (e.g., a higher willingness to pay for high-quality goods). Finally, the estimated values of α are positive and below one, which is consistent with the notion that the utility function is concave.

The reported standard errors have been corrected to account for potential serial dependence. We attempt to control for as many of the observed potential dynamic factors, such as timing of trips, loyalty, and inventories, as possible. Despite these controls, we still find unexplained persistence in the residuals.

Accounting for time-series increases some of the standard errors by as much as a factor of 1.8 in comparison with the simple robust standard errors.[8] Nonetheless, almost all the parameters remain significant after this correction, probably because of the extremely large sample. For now, we have not derived an explicit source for this persistence, focusing solely on the precision of our standard errors. Because most marketing studies do not correct for unobserved time-series, these findings suggest the need for further research into how well existing static models capture both heterogeneity as well as choice dynamics.

5.2. Aggregate Demand and Substitution Patterns

We use elasticities to measure consumer sensitivity to marketing mix variables. Because households do not face the same mix of marketing variables across shopping trips, we report elasticities in response to uniform percentage changes in marketing mix variables. Following Ben-Akiva and Lerman (1985), we report the sum of the elasticity on each trip, $\varepsilon_{p_k}^i$, weighted by the share of total unit sales: $\varepsilon_{p_k}^i = \sum_{t=1}^{T} \varepsilon_{p_k}^{ti}(Q_{ti}/\sum_{t=1}^{T} Q_{ti})$, $i = 1, \ldots, I$.

Table 6 presents estimated own-price, feature ad, and display elasticities for the baseline model. All of the own-price elasticities are greater than one, which is consistent with typical static category-pricing models. Because feature ads and displays are binary variables, we report the change in demand from switching from a zero value to a one. We find that advertising has the largest impact on caffeine-free diet colas, the 12-pack of peppers, and on the lemon-lime products. In contrast, regular colas appear to have the lowest advertising and display responses. We also find that advertising has a relatively small effect on the 6-packs of cans and bottles, and a relatively large effect on 67.6-oz bottles. We find similar effects from display. From a retail manager's perspective, these results suggest that the ability to stimulate consumer response from marketing tools, such as newspaper advertising and in-store displays, will vary for different flavors and for different package types.

In a separate Appendix (available from the authors on request), we present the cross-price elasticities from the baseline model. The predicted substitution patterns show that, most importantly, consumers seem to respond to price changes by switching to another product of the same size. Most of the observed substitution patterns reflect realistic interactions. Almost all products substitute primarily to a cola. Also, 6-packs of caffeine-free Diet Pepsi are very substitutable with 6-packs of Diet Pepsi. Mountain Dew and Dr. Pepper are generally predicted as likely substitutes. Surprisingly, we find little interaction

---

[8] We do not find any evidence of cross-sectional dependence.

Table 6   Own Price, Ad, and Display Elasticities for the Baseline Model and Multinomial Comparison Model

Products	Baseline Model			Multinomial Model	
	Price	Feature Ad	Display	No Price in λ	Prices in λ
PEPSI 6P	−2.38	1.41	1.69	−0.59	−1.29
COKE CLS 6P	−2.11	2.25	2.49	−0.59	−1.32
PEPSI DT 6P	−3.61	2.01	1.98	−0.62	−1.35
COKE DT 6P	−2.47	2.32	2.47	−0.67	−1.46
DR PR 6P	−3.14	2.78	2.81	−0.73	−1.51
MT DW 6P	−3.04	2.85	3.06	−0.72	−1.37
PEPSI DT CF 6P	−3.56	2.69	2.87	−0.68	−1.44
A&W CF 6P	−3.59	3.57	3.84	−0.73	−1.54
PEPSI 16 oz	−2.25	1.42	1.50	−0.85	−1.68
PEPSI 12P	−2.16	2.44	2.69	−0.64	−1.40
COKE CLS 12P	−2.13	2.40	2.67	−0.64	−1.39
COKE DT 12P	−2.50	2.33	2.31	−0.72	−1.53
PEPSI DT 12P	−2.66	3.16	3.63	−0.74	−1.61
DR PR 12P	−2.47	3.95	4.20	−0.81	−1.69
MT DW 12P	−3.02	3.09	3.04	−0.82	−1.54
COKE DT CF 12P	−2.76	4.05	3.45	−0.78	−1.61
SP CF 12P	−2.57	4.71	4.97	−0.78	−1.53
PEPSI DT CF 12P	−2.92	6.20	6.14	−0.81	−1.70
PEPSI 67.6 oz	−2.62	2.10	2.23	−0.42	−1.02
COKE CLS 67.6 oz	−2.80	3.41	3.42	−0.43	−1.04
PEPSI DT CL 67.6 oz	−2.66	3.16	3.40	−0.47	−1.10
COKE DT 67.6 oz	−2.81	4.65	4.77	−0.48	−1.19
DR PR 67.6 oz	−2.94	3.18	3.32	−0.49	−1.16
7UP R CF 67.6 oz	−2.57	3.13	3.57	−0.45	−1.02
7UP DT CF 67.6 oz	−2.61	3.41	3.47	−0.47	−1.01
MT DW 67.6 oz	−3.23	4.11	4.16	−0.49	−1.07

Notes. 6P = 6-pack; CLS = classic; DT = diet; DR PR = Dr. Pepper; MT DW = Mountain Dew; CF = caffeine free; 12P = 12-pack; SP = Sprite; CL = Cola; R = regular.

between Sprite and 7UP, mainly because they do not include comparable sizes in the choice set.

5.3.   Model Sensitivity

In this section, we check the sensitivity of our baseline model to various underlying assumptions. In Tables 7 and 8, we present results from comparison models to investigate the sensitivity of our taste parameters to the specification of consumption occasions. We also report the mean-square error from out-of-sample predictions in Table 8, where we re-estimate each model using the first eight quarters of data, using the ninth quarter as a holdout sample.

The baseline model's parameter estimates appear in the first column. In the second column (heterogeneous Poisson), we check the sensitivity of our Poisson assumption by mixing it with a normally distributed random intercept. This modification offsets some of the restrictive properties of the Poisson, such as the equal mean and variance. The use of this more flexible distribution leads to smaller parameter magnitudes in the mean of the Poisson, as well as the vertical component, m_h. However, we do not see a notable impact on the taste parameters. We also see a

relatively lower predictive ability out of sample than the baseline model.

In the third column (reference prices), we control for price expectations using proxies for reference prices (see Kalyanaram and Winer 1995 for a survey). Using a price index for the previous trip, as well as for the current trip, we implicitly assume that consumers use past observations of shelf prices to form an internal reference point with which to assess the current price level (Winer 1985).[9] The omission of price expectations could limit the model if, for instance, households defer some of their current purchases in anticipation of better prices on a subsequent trip. Interestingly, these additional controls reduce the size of the random intercept in λ. However, as before, the taste parameters appear to be fairly robust to this change. Out of sample, this model performs very similarly to the simple heterogeneous Poisson. Although beyond the scope of the current analysis, others have modeled forward-looking behavior by consumers more formally in the context of discrete-choice models when consumers have expectations about prices (Gonul 1999) or quality (Erdem and Keane 1996).

In the final column (deterministic J_h), we investigate the role of modeling the consumption occasions. We check the sensitivity of the Poisson specification by restricting the number of decisions made on a given trip to be equal to the actual number of alternatives purchased, making J_h deterministic ($J_{ht} = \sum_{i=1}^{I} I_{\{Q_{it}^h > 0\}}$). This model would have limited appeal for managers running policy simulations, because it cannot predict the changes in consumers' demand in response to a counterfactual change in marketing conditions. Although many of the parameters appear qualitatively unchanged, the scale of purchases, S_h, now depends very heavily on family size. Offsetting this effect, the interaction between family size and package size (in the tastes) becomes insignificant. Similarly, the responses to feature ads and displays increase dramatically, suggesting these parameters may be relatively more sensitive to the ability of the model to forecast the number of consumption occasions. These changes are likely related to the fact that conditioning on the total number of brands in the shopping basket also provides information about no-purchase. To provide a reasonable out-of-sample predictive fit, we do not condition on the observed number of brands. Instead, we set J_h equal to household h's average number of brands per trip in the estimation sample (average brands). We also set J_h equal to the predicted number of brands using a simple regression of total brands on the covariates used in λ_h in the baseline model (regression). For both cases, we

[9] We thank Russ Winer for suggesting this approach.

Table 7   Nonlinear Coefficients

Variables	Baseline	Heterogeneous Poisson	Reference Prices	Deterministic J_h	
λ_h: Intercept	—	0.330 (0.001)	0.076 (0.000)	—	
λ_h: Kids	0.336 (0.001)	0.040 (0.000)	0.328 (0.001)	—	
λ_h: Family size	0.041 (0.001)	0.001 (0.000)	0.040 (0.000)	—	
λ_h: Time since last CSD	0.001 (0.000)	0.003 (0.000)	0.001 (0.000)	—	
λ_h: Time since last trip	−0.003 (0.002)	−0.003 (0.000)	−0.003 (0.001)	—	
λ_h: Temperature	0.004 (0.001)	0.010 (0.000)	0.003 (0.001)	—	
λ_h: Holdiay	0.009 (0.004)	0.056 (0.02)	0.010 (0.004)	—	
λ_h: Favorite products	—	—	1.007 (0.004)	—	
λ_h: Overall prices	—	—	0.466 (0.002)	—	
λ_h: Lag favorite products	—	—	−0.427 (0.002)	—	
λ_h: S.D. of intercept	—	1.535 (0.084)	0.056 (0.000)	—	
s_h: Constant	1.549 (0.104)	0.978 (0.048)	1.545 (0.079)	0.132 (0.024)	
s_h: Family size	0.958 (0.049)	0.013 (0.001)	0.986 (0.049)	8.662 (0.037)	
s_h: Time since last trip	0.012 (0.003)	0.013 (0.004)	0.013 (0.001)	0.010 (0.002)	
s_h: Time since last CSD	0.013 (0.003)	2.344 (0.143)	0.013 (0.003)	0.009 (0.005)	
m_h: Income	2.305 (0.172)	0.009 (0.520)	2.358 (0.165)	2.699 (0.520)	
α	0.009 (0.001)	0.009 (0.001)	0.010 (0.001)	0.008 (0.001)	
Hansen's J (degrees-of-freedom)	629.49 (206)	649.94 (204)	641.24 (201)	534.97 (212)	
				Average Brands	Regression
Out-of-sample MSE	0.0407	0.0417	0.0418	0.0475	0.0491
Store trips	169,788	169,788	169,788	169,788	

Notes. S.D. = standard deviation; MSE = ?; lag = ?.

find that the baseline model provides a better out-of-sample fit according to the mean-square error.

To pursue the role of consumption occasions further, we now consider a purely statistical model that treats the total number of items in the shopping basket ($Q_t^h = \sum_{i=1}^{I} Q_{it}^h$) as a Poisson random variable. The allocation of these items across the 26 brands is modeled as a multinomial distribution. The corresponding success probabilities for this multinomial distribution are modeled as standard random coefficients logits. The derivation of this model is provided in the Appendixes. We report empirical results in Table 9. Because the treatment of prices in the Poisson quantity regression is arbitrary, we report a specification with no prices included, as well as a specification with the mean of all 26 prices for the shopping trip.[10]

Interestingly, demographic variables seem to play a comparable role in determining the expected total number of product alternatives purchased. Note that the taste parameters do not have precisely the same interpretation as in the baseline model. In the "quantity-then-brand choice" model, we need to normalize one of the alternatives (we use A&W), so that

our results are relative to this base brand.[11] Marketing mix parameters have the expected signs, although there seems to be much less heterogeneity in the response to feature ads and displays. Unexpectedly, the response to last product purchased is negative, in contrast with the baseline model, where it is positive. This result would suggest that households are loyal to brands, but vary the package sizes they purchase. We also note the strong role of including prices in the mean of the quantity regression, λ. First, the mean response to prices (β in the brand probabilities) increases substantially. Second, many of the variance terms (Ω) increase, implying a larger degree of heterogeneity.

In Figure 1, we see that the baseline model and the multinomial model provide comparable aggregate sales predictions, although the multinomial model underpredicts demand for the larger-share items. We attribute this problem to the difficulty in relating quantity decisions for each brand to their respective shelf prices. In the multinomial model, the quantity price elasticities are all captured through a single parameter (the coefficient on mean trip price in λ). Similarly, the single-price parameter may not account well for different package sizes (e.g., cents off a 6-pack of cans may not be equivalent to cents off a 2-liter bottle). In contrast, the baseline model allows consumers to assess each brand's quality to price

---

[10] We also attempted to include each of the 26 product prices in the Poisson model, adding an additional 26 parameters. Although this specification did not alter the other model parameters dramatically, several of the price coefficients had positive signs. For a category monopolist, this would imply that setting the corresponding prices to infinite would cause consumers to have infinitely large shopping baskets.

[11] This normalization also prevents us from recovering the mean tastes for product attributes using the weighted regression approach for the baseline model.

Table 8 Taste Coefficients for Fixed Attributes in Quality Function—First Stage Only

Variables	Baseline Model	Heterogeneous Poisson	Reference Prices	Deterministic J_h
β: feature ad	1.52	1.41	1.41	7.32
	(0.03)	(0.03)	(0.04)	(0.05)
β: S.D. feature ad	0.06	0.06	0.06	0.44
	(0.01)	(0.01)	(0.01)	(0.02)
β: Display	2.00	1.79	1.79	7.69
	(0.04)	(0.04)	(0.05)	(0.06)
β: S.D. display	0.79	0.78	0.79	1.63
	(0.05)	(0.03)	(0.07)	(0.03)
β: Brand loyalty	0.27	0.28	0.28	0.37
	(0.03)	(0.04)	(0.05)	(0.03)
β: Product loyalty	0.02	0.02	0.02	−0.00
	(0.16)	(0.09)	(0.01)	(0.02)
Ω: S.D. constant	1.88	1.87	1.89	1.22
	(0.03)	(0.02)	(0.04)	(0.01)
Ω: S.D. diet	0.26	0.26	0.26	0.38
	(0.02)	(0.02)	(0.03)	(0.01)
Ω: S.D. citric	1.23	1.24	1.24	0.10
	(0.04)	(0.05)	(0.04)	(0.02)
Ω: S.D. caramel	0.22	0.21	0.21	0.00
	(0.02)	(0.02)	(0.02)	(0.01)
Ω: S.D. cans $\times$ 6	1.19	1.21	1.21	0.00
	(0.11)	(0.10)	(0.10)	(0.02)
Ω: S.D. cans $\times$ 12	0.30	0.30	0.30	0.00
	(0.02)	(0.01)	(0.02)	(0.01)
Ω: S.D. bott $\times$ 6	0.19	0.19	0.19	0.01
	(0.02)	(0.03)	(0.02)	(0.06)
μ: (Kids $*$ Caffeine)	0.40	0.39	0.39	2.15
	(0.03)	(0.02)	(0.03)	(0.08)
μ: (Household size $*$ Servings)	0.01	0.01	0.01	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
μ: (Female head $<$ 35 $*$ Diet)	1.13	0.94	0.94	2.01
	(0.03)	(0.04)	(0.04)	(0.03)
Store trips	169,788	169,788	169,788	169,788

Note. S.D. = standard deviation.

Table 9 Taste Coefficients for the "Quantity-Then-Brand Choice" Model

Variables	No Price in λ		Prices in λ	
	$\bar{\beta}$ Mean (S.E.)	Ω S.D. (S.E.)	$\bar{\beta}$ Mean (S.E.)	Ω S.D. (S.E.)
β: Price	−2.839	0.059	−6.241	3.026
	(−0.039)	(0.019)	(0.041)	(0.027)
β: Feature ad	0.036	0.031	0.007	0.027
	(0.006)	(0.005)	(0.006)	(0.005)
β: Display	0.571	0.0025	0.504	0.0016
	(0.005)	(0.004)	(0.006)	(0.005)
β: Product loyalty	2.761		2.209	
	(0.003)		(0.0034)	
β: Brand loyalty	−1.114		−0.613	
	(0.029)		(0.032)	
β: Intercept		0.056		0.524
		(0.002)		(0.015)
β: Diet		0.019		0.969
		(0.003)		(0.005)
β: Citric		0.026		0.525
		(0.004)		(0.005)
β: Caramel		0.160		0.369
		(0.008)		(0.004)
β: Cans $*$ 6		0.018		0.336
		(0.003)		(0.009)
β: Cans $*$ 12		0.053		1.178
		(0.003)		(0.006)
β: Bottles $*$ 16		0.049		2.636
		(0.013)		(0.01)
μ: Kids $*$ Caffeine	0.1552		−0.014	
	(0.0077)		(0.008)	
μ: (Family size) $*$ Servings	0.000		−0.001	
	(0.0004)		(0.001)	
μ: (Female $<$ 35) $*$ Diet	−0.543		−0.701	
	(0.0057)		(0.009)	
λ: Kids	0.309		0.127	
	(0.002)		(0.002)	
λ: Family size	−0.136		0.057	
	(0.001)		(0.001)	
λ: Time since last CSD	0.018		0.027	
	(0.000)		(0.001)	
λ: Time since last trip	−0.013		−0.011	
	(0.000)		(0.000)	
λ: Temperature	−0.008		0.0017	
	(0.000)		(0.000)	
λ: Holiday	0.299		0.400	
	(0.005)		(0.005)	
λ: Price	—		−0.382	
	—		(0.003)	
Store trips		169,788		

Notes. S.E. = standard error; S.D. = standard deviation.

ratio for both quantity and brand choices. Referring back to Table 6, we can also see that the multinomial model predicts own-price elasticities that are much lower than those of the baseline model. In the specification that omits prices entirely from the mean of the quantity regression, the elasticities are all less than one in magnitude, which would be inconsistent with standard category management pricing models. Note that this finding of inelastic brand choice probabilities is consistent with previous work studying the role of quantities (e.g., Chintagunta 1993). Including prices in the quantity regression results in more elastic estimates of demand, albeit substantially lower than those of the baseline model. Although not reported, using these estimates to study category pricing implies negative marginal costs (the implied margins are larger than the observed shelf prices). These results may suggest the importance of allowing prices to affect individual brand choices, as well as brand-specific quantity decisions.

6. Conclusions

Although the typical logit and probit DCMs have provided useful predictions for consumer purchases in many product categories, their restrictive single-unit purchase assumption seems inappropriate for several categories, such as CSDs. Instead, we estimate a model that allows consumers to purchase a bundle of products within a category. In addition, demographic variables, which have typically been found

Figure 1   Model Sensitivity: Predicted Aggregate Sales (In-Sample) for Baseline Model and Multinomial Model
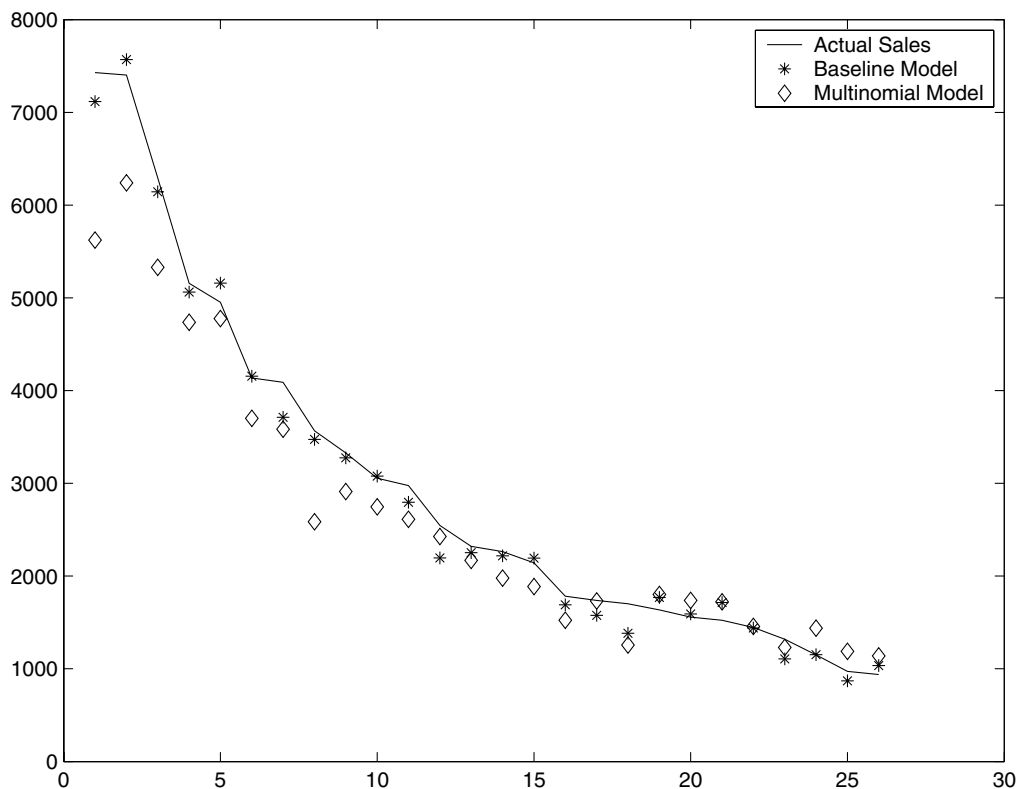
to provide little information in marketing applications, are instrumental in determining the joint distribution of total product alternatives and total units purchased on a given trip. Demographics also partially explain observed differences in tastes for product attributes. We also find that the correction for a 15-day lag in the persistence of the residuals has a substantial impact on estimated standard errors of choice model parameters. The estimates of consumer preferences seem to be fairly robust to some of the crucial underlying model assumptions.

The proposed specification derives from a microeconomic framework of consumer choice. In addition to providing intuition for the underlying model parameters, the structural model could also be useful for policy simulations that require economic metrics such as consumer willingness to pay. To demonstrate the advantages of the derivation, we compare our estimates to an alternative statistical "quantity-then-brand choice" model that also captures multiple discreteness. Whereas the alternative model provides comparable in-sample predictions of aggregate demand, it does not provide sensible price elasticities.

The current results suggest that accounting for multiple discreteness may be relevant for measuring the sensitivity of consumer demand to marketing mix. Future research might consider the additional role of attribute satiation on a given shopping trip.

Although simulating the consumption occasions, one could introduce flavor interaction terms that reflect which flavor combinations have been selected across expected consumption occasions. In addition to linking decisions across consumption needs, these interaction terms would also provide a statistical test for complementarities between flavors.

Finally, the current analysis accounts for variety in consumer purchases during a trip. Future research may benefit from examining how shopping dynamics influence consumer shopping baskets. For instance, the number of consumption occasions during a trip could be a function of both expectations about future prices and the rate of consumption or inventory depletion. Whereas we attempt to proxy for these effects, current research may help provide guidance in how to model these aspects formally (Erdem et al. 2002).

Appendix A: GMM Procedure

Using (10), we define the prediction error $\varepsilon_{ht}(D_{ht}, X_{ht}, \Theta_0) = EQ(D_{ht}, X_t, \Theta) - Q_{ht}$, where the vector Q_{ht} contains the observed purchases. At the true parameter values, Θ_0:

$$E\{\varepsilon_{ht}(D_{ht}, X_{ht}, \Theta_0)\} = \vec{0}_I, \qquad h = 1, \ldots, H \text{ and } t = 1, \ldots, T_h.$$
(15)

We also assume that:

$$E\{\varepsilon_{ht}(D_h, \Theta_0)\varepsilon_{hk}(D_{ht}, \Theta_0)'\} = \Omega_{|t-k|},$$
(16)

where $\Omega_{|t-k|}$ is a finite ($I \times I$) matrix. Using (15), we can construct conditional moments based on $Z_{ht} = D_{ht} \otimes I_J$, which is independent of the unobservables (Hansen 1982):

$$E\{Z_{ht} * \varepsilon_{ht}(D_{ht}, \Theta_0) \mid Z_{ht}\} = \vec{0}_I.$$

For estimation, we use the sample analog of these conditional moments:

$$g(D_{HT}, \Theta) = \frac{1}{HT} \sum_{h=1}^{H} \sum_{t=1}^{T_h} Z_{ht} * \varepsilon_{ht}(D_{ht}, \Theta),$$
(17)

where $D_{HT} \equiv (D'_{1T1}, \ldots, D'_{HT_H})$ denotes the matrix containing all of the household/trip information for the sample of H households, each making T_h shopping trips, and $T = 1/H \sum_{h=1}^{H} T_h$. We search for the value Θ_{GMM} that minimizes the function J_{HT} given by:

$$J_{HT}(\Theta) = [g(D_{HT}, \Theta)]' W_{HT} [g(D_{HT}, \Theta)].$$
(18)

For the weight matrix, W_{HT}, we use the inverse of the asymptotic variance of g which gives asymptotically efficient estimates under certain regularity conditions (Hansen 1982). The computation of W is discussed below. This framework gives estimates with the following asymptotic distribution:

$$\sqrt{N}(\Theta_{GMM} - \Theta_0) \implies N(0, \Xi)$$
(19)

$$\Xi = \left(\text{plim}\left\{\frac{dg(D_{ht}, \Theta_0)}{d\Theta}\right\} W \text{plim}\left\{\frac{dg(D_{ht}, \Theta_0)}{d\Theta}\right\}'\right)^{-1}.$$
(20)

Appendix B: Estimation of the Weight Matrix, W

Hansen (1982) shows that, under certain regularity conditions, the efficient weighting matrix W_{HT} is the inverse of S, the variance of the sample moments. The variance of the moments has the following form:

$$S = \lim_{H, T \to \infty} HT \cdot E\{E([g(D_{HT}, \Theta_0)][g(D_{HT}, \Theta_0)]' \mid D_{HT})\}$$

$$= \lim_{H, T \to \infty} \frac{1}{HT} \sum_{h=1}^{H} \sum_{t=1}^{T_h} \sum_{k=1}^{T_h} E\left[\left(1 + \frac{1}{R}\right) Z_{ht} \Omega_{tk} Z'_{hk}\right],$$
(21)

where R is the number of simulation draws. Similar to the discussion in McFadden (1989), the added simulation "noise" will not affect the consistency of the estimator, but it will reduce the efficiency by a factor of $(1 + 1/R)$. As $R \to \infty$, the estimator approaches asymptotic efficiency (we use 30 draws).

In the current context, S must account for the panel structure of the data. To deal with the cross-sectional aspect of the data, we include several state variables, such as temperature and seasonal dummies, to capture contemporaneous aggregate demand shocks that could affect households in a similar fashion. Despite these controls, we may still find dependence in the shocks. The source of these shocks could be systematic measurement error. For instance, household-specific reporting errors in the scanning process or shopping at nonscanner stores could generate unobserved serial dependence. Alternatively, unobserved heterogeneity not captured by the current model specification might also generate dependence. Yet another source of error consists of unobserved local factors, such as the presence of a convenience store in which consumers purchase CSDs. One might expect this latter source of error to be common across households in close proximity, potentially generating cross-sectional dependence.

To estimate the matrix S, we index each observation by both time and household. We use Conley's (1999) nonparametric, positive, semidefinite covariance estimator, which is analogous to Newey and West's (1987) spectral time-series estimator. This approach allows us to correct both for serial dependence over time, as well as dependence across geographic space (e.g., households that are physically close to one another may have correlated shocks). Given a consistent estimate $\widehat{\Theta}$ and a predetermined time L and inter-household distance B, beyond which the dependence in the shocks dies out, the estimator for S is:

$$\widehat{S}_{HT} = \frac{1}{HT} \sum_{h,l=1}^{H} \sum_{t=1}^{L} \sum_{k=t+1}^{T_h} \omega(|t-k|, d_{hl})$$
$$\cdot [M(D_{h,k}, \widehat{\Theta}) M(D_{l,k-t}, \widehat{\Theta})' + M(D_{h,k-t}, \widehat{\Theta}) M(D_{l,k}, \widehat{\Theta})']$$
$$- \frac{1}{HT} \sum_{h=1}^{H} \sum_{k=1}^{T_h} M(D_{h,k}, \widehat{\Theta}) M(D_{h,k}, \widehat{\Theta})',$$
(22)

where d_{hl} is the Euclidean distance between households h and l, and $M^s(D_{h,k}, \widehat{\Theta}) = Z_{hk} * \varepsilon_{hk}(D_{hk}, \widehat{\Theta})$. We use the Bartlett weight for $\omega(t, d)$ to assign decreasing weight to the correlation between households' purchases as they grow further apart in time and space:

$$\omega(|t-k|, d_{hl}) = \begin{cases} \left(1 - \frac{|t-k|}{1+L}, \text{if } t \leq L\right) \left(1 - \frac{d_{hl}}{1+B}, \text{if } d_{hl} \leq B\right) \\ 0, \quad \text{else} \end{cases}.$$
(23)

Appendix C: "Quantity-Then-Brand Choice" Model

In this section, we describe a combined quantity-and-brand choice model that accounts for the multiple discreteness observed in the soft drink purchases. The model consists of two components. The first component, the Poisson quantity

regression, estimates the total number of items in the shopping basket on a given trip. The total units purchased are distributed among the various brands according to a multinomial distribution. The approach is similar to Dillon and Gupta (1996), although they use cross-sectional data and they do not include the no-purchase option.

More formally, we assume the total number of items in household $h's$ basket during trip t, $\widetilde{Q}_{ht} = \sum_{i=1}^{I} Q_{hit}$, is distributed Poisson with mean, λ_h, a function of household characteristics. Thus, category purchases for an individual can be written as:

$$P(\widetilde{Q}_{ht} = n_{ht}; \lambda) = \frac{\lambda_h^{n_{ht}} e^{-\lambda_h}}{n_{ht}!}, \qquad n_{ht} = 0, 1, \ldots.$$

To capture the brand choice aspect of demand, we assume consumer choices are generated by a multinomial distribution:

$$P(Q_{ht1} = n_{ht1}, \ldots, Q_{htI} = n_{htI}; n_h, p_h) = \frac{n_{ht}!}{\prod_i n_{hit}!} \prod_i p_{hit}^{n_{hit}},$$

where p_{hit} is the probability that consumer h chooses product i and $n_{ht} = \sum_{i=1}^{I} n_{hit}$ are the total units purchased on the trip. We model the brand choice probability, p_{hit}, as a multinomial logit:

$$p_{hit} = \frac{\exp(X_{it}\beta_h)}{1 + \sum_{k=1}^{I} \exp(X_{kt}\beta_h)},$$

where X_{it} is a vector of product i attributes including, marketing mix, and β_h is a vector of household $h's$ response parameters. To capture heterogeneity across households, we model the parameters, β_h, as normally distributed random variables:

$$\beta_h = \bar{\beta} + D'_h \mu + \Omega \nu_h,$$

where $\bar{\beta}$ is the mean response across households, Ω is a diagonal matrix whose elements are standard deviations, and ν_h is a vector of i.i.d. standard normal deviates. As in the proposed model, we capture correlations in product valuations using random tastes for product attributes.

Assembling the components of the model, we can write the conditional likelihood for household $h's$ category purchase vector on trip t:

$$L_{ht|\beta} = \left(\frac{\lambda_h^{n_{ht}} e^{-\lambda_h}}{n_{ht}!} \right) \left(\frac{n_{ht}!}{\prod_i n_{hit}!} \prod_i p_{hit}^{n_{hit}} \right).$$

The conditional likelihood of household $h's$ entire purchase history is then computed as $L_{h|\beta} = \prod_t L_{ht|\beta}$. Accounting for the uncertainty in household $h's$ response parameters, we obtain the unconditional likelihood: $L_h = \int \prod_t L_{ht|\beta(\nu)} \phi(\nu) d\nu$, where $\phi(\cdot)$ is the probability density function of a standard normal random variable. To estimate the model parameters λ, $\bar{\beta}$, and Ω, we use simulated maximum likelihood (see, e.g., Erdem 1996).

Note that this alternative multinomial specification shares many similarities to the proposed model of §2. For instance, unlike standard brand choice models, we explicitly use the information contained in the quantity purchases. A subtle difference in this model is that the total number of items in the shopping basket is determined by the Poisson. In the proposed model of §2, the Poisson distribution determines the number of consumption occasions, placing an upper bound on the number of brands chosen. However, the actual nunber of units of each of the selected brands is not constrained by the Poisson.

References

Allenby, Greg M., Thomas S. Shively, Sha Yang, Mark J. Garrett. 2001. A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. *Marketing Science* **23** 95–108.

Anderson, E., N. Vilcassim. 2001. Structural demand models for retail category pricing. Working paper, University of Chicago, Chicago, IL.

Ben-Akiva, M., S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. The MIT Press, Cambridge, MA.

Berry, S. 1994. Estimating discrete-choice models of product differentiation. *Rand J. Econom.* **25** 242–262.

Berto Villas-Boas, Sofia. 2001. Vertical contracts between manufacturers and retailers: An empirical analysis. Working paper, University of California, Berkeley, CA.

Chiang, J. 1991. The simultaneous approach to the whether, what, and how much to buy questions. *Marketing Sci.* **10** 297–315.

Chintagunta, P. 1992. Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Sci.* **11** 386–407.

Chintagunta, P. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Sci.* **12** 184–208.

Chintagunta, P, J. P. Dubé, V. Singh. 2002. Balancing profitability and customer value: An application to zone-pricing by a supermarket chain. Working paper, University of Chicago, Chicago, IL.

Chintagunta, P., E. Kyriazidou, J. Perktold. 2001. Panel data analysis of household brand choices. *J. Econometrics* **103** 111–153.

Conley, T. G. 1999. GMM estimation with cross-sectional dependence. *J. Econometrics* **92**(1) 1–45.

Dillon, W. R., S. Gupta. 1996. A segment-level model of category volume and brand choice. *Marketing Sci.* **15** 38–59.

Erdem, T. 1996. A dynamic analysis of market structure based on panel data. *Marketing Sci.* **16** 359–378.

Erdem, T., M. P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Sci.* **15** 1–20.

Erdem, T., S. Imai, M. P. Keane. 2002. A model of consumer brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing Econom.* **1**(1) 5–64.

Fader, P. S., B. G. S. Hardie. 1996. Modeling consumer choice among SKUs. *J. Marketing Res.* **33** 442–452.

Gönül, F. F. 1999. Estimating price expectations in the OTC medicine market: An application of dynamic stochastic discrete choice models to scanner panel data. *J. Econom.* **89** 41–56.

Gönül, F. F., K. Srinivasan. 1996. Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Sci.* **15** 262–279.

Guadagni, P. M., J. D. C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2** 203–238.

Hansen, L. P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* **50** 1029–1054.

Harlam, B. A., L. M. Lodish. 1995. Modeling consumers' choices of multiple items. *J. Marketing Res.* **32**(November) 404–418.

Hauser, J. R., B. Wernerfelt. 1990. An evaluation cost model of consideration sets. *J. Consumer Res.* **16** 393–408.

Hendel, I. 1999. Estimating multiple-discrete choice models: An application to computerization returns. *Rev. Econom. Stud.* **66** 423–446.

Kalyanam, K., D. S. Putler. 1997. Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Sci.* **16** 166–181.

Kalyanaram, G., R. S. Winer. 1995. Empirical generalizations from reference price research. *Marketing Sci.* **14** 161–169.

Keane, M. P. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15** 310–327.

Krishnamurthi, L., S. P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Sci.* **7** 1–20.

Manchada, P., A. Ansari, S. Gupta. 1999. The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Sci.* **18** 95–114.

McAlister, L. 1982. A dynamic attribute satiation model of variety-seeking behavior. *J. Consumer Res.* **9** 141–150.

McCulloch, R., P. E. Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *J. Econometrics* **64** 207–240.

McFadden, D. 1981. Econometric models of probabilistic choice. C. Manski, D. McFadden, eds. *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.

McFadden, D. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57** 995–1026.

Montgomery, A. L. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Sci.* **16** 315–338.

Newey, W. K., K. D. West. 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708.

Pakes, A., D. Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* **57** 1027–1057.

Rossi, P. E., G. Allenby. 1993. A Bayesian approach to estimating household parameters. *J. Marketing Res.* **30** 171–182.

Seetharaman, P. B. 1999. Modeling serial correlation in household brand choices using a multinomial probit model of brand choice. Working paper, Washington University, St. Louis, MO.

Simonson, I. 1990. The effect of purchase quantity and timing on variety-seeking behavior. *J. Marketing Res.* **27** 150–162.

Smith, R. J. 1992. Non-nested tests for competing models estimated by generalized method of moments. *Econometrica* **60** 973–980.

Walsh, J. W. 1995. Flexibility in consumer purchasing for uncertain future tastes. *Marketing Sci.* **14** 148–165.

Winer, R. S. 1985. A price vector model of demand for consumer durables: Preliminary developments. *Marketing Sci.* **4** 74–90.