3B2v8.06a/w (Dec 5 2003).51c +model | ECONOM : 2822 | Prod.Type:FTP pp.1–23(col.fig.:1,2) | ED:Jayasudha PAGN:Mishra SCAN:

ARTICLE IN PRESS

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Product attributes and models of multiple discreteness

Jaehwan Kim[a,*], Greg M. Allenby[b], Peter E. Rossi[c]

[a]Korea University School of Business, Korea University, 5 Anam, Sungbuk-gu, Seoul 136-701, Korea
[b]Fisher College of Business, Ohio State University, 2100 Neil Avenue, Columbus, OH 403210, USA
[c]Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA

Abstract

Demand for product characteristics is examined within the context of models that allow for both corner and interior solutions corresponding to zero and non-zero demand. Product attribute information is associated with marginal utility and curvature (satiation) parameters of various utility functions. Empirical applications demonstrate the need for incorporating characteristics in a fairly general way. We also compare our approach to an ideal point and pure Lancasterian versions of our nonlinear utility model. The data support our model over either the ideal point or Lancasterian variants.

JEL classification: M3; D1; C11

Keywords: Characteristic model; Non-homothetic utility; Multiple discreteness; Product attributes

1. Introduction

The profusion of disaggregate data on consumer demand obtained either from market place observation or surveys has simulated a great deal of work on models with discrete components. Multinomial choice models have been, by far, the most popular models used with disaggregate data. However, these choice models ignore the quantity aspects of demand and can only be applied to sets of goods for which demand is mutually exclusive,

*Corresponding author. Tel.: +82 2 3290 2603.
E-mail addresses: jbayes@korea.ec.kr (J. Kim), allenby1@osu.edu (G.M. Allenby), peter.rossi@ChicagoGsb.edu (P.E. Rossi).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.05.020

i.e. only one good is purchased on each occasion. Consumers are often observed to purchase or select multiple goods on the same occasion while revealing a demand of zero for the vast majority of the available offerings. This data requires a model with a mixture of corner and interior solutions. We may also require that our model be derived from a valid utility function to facilitate policy analysis.

Kim et al. (2002) offered a utility-based model of demand along with a practical method of conducting likelihood-based inference for this model. However, the simple model in Kim et al. is lacking several important features. In marketing applications, there are typically a very large number of product offerings with a wide variety of product attributes. In order to make policy statements about optimal product assortment or design of new products, it is important to allow for product characteristics or attributes to enter the utility function. The purpose of this paper is to consider a number of extensions of the Kim et al. model of demand to incorporate product attribute information. As the basic utility model is nonlinear and allows for satiation or diminishing marginal utility, we will have several different ways of incorporating product attributes—both to influence the level of marginal utility afforded by a product offering as well as to influence the rate of satiation.

Once product characteristics are considered as drivers of utility, it is natural to consider a characteristics approach to demand such as that offered by Lancaster (1966) (see also Berry and Pakes, 2002). In the Lancasterian approach, demand is defined over the level of characteristics provided by a given bundle of demanded products, rather than over the products themselves. In the characteristics space, we should not always assume that marginal utility is strictly increasing (decreasing) in characteristics and might consider the "ideal point" alternative in which consumers have an ideal level of product characteristics, any deviation from which will result in lower utility. We compare our extended model of demand with the ideal point and Lancasterian approaches.

Kim et al. (2002) consider the demand for different varieties of yogurt for which the set of characteristics is at least as large as the number of product offerings. In this paper, we consider two other data sets, which have products with well-defined characteristics. We have created a data set of demand for various salty snacks via field experimentation and also report on results using "volumetric" conjoint data in which respondents not only choose between alternative offerings but indicate the quantity demanded. These new datasets illustrate the importance of the model extensions. We find that variants of our extended model outperform various ideal point and Lancasterian specifications. Using the conjoint data, we compare our utility-based approach to a reduced form Poisson regression and find that our approach has superior predictive performance.

2. The demand model and alternative parameterizations

The standard choice model is derived from a linear utility specification. The linear utility specification gives rise to a corner solution in which only one product is purchased on any one-purchase occasion. In some cases, an outside alternative, or "no purchase" option, is included to allow for consumers to have a base or reference level of utility. This model assumes that all products are perfect substitutes, and results in one alternative with non-zero demand.

The "multiple discreteness" phenomena in which two or more (but not all) product offerings are purchased reveals that the product offerings are not viewed as perfect substitutes. We can relax the assumption of perfect substitutability by allowing the utility function to be additive but nonlinear, $u(x) = \sum_{j=1}^{J} u_j(x_j)$. The additive model of utility assumes that all products are substitutes but the extent of this substitutability may vary across different pairs of offerings. In addition, the non-linearity gives rise to satiation or diminishing marginal utility. It is entirely possible that different products have differing rates of satiation. In models with the outside good, we clearly will need this possibility. The rate of satiation for the composite outside good is almost certainly lower than the inside goods, and in some cases a reasonable approximation might be that the outside good has constant marginal utility. Here, we do not consider models which allow for complementarity (see Gentzkow, 2005) as most disaggregate data is available on classes of products which are substitutes.

Standard non-linear utility specifications often result in strictly interior solutions in which all goods are demanded. While this might be appropriate for broad classes of goods such as food and housing, it is not appropriate for disaggregate demand modeling. In order to achieve the possibility of corner solutions, we translate the utility function so that marginal utility is finite at the axes. Perhaps, the simplest model which can exhibit a mixture of corner and interior solutions as well as satiation is the model of Kim et al. (2002):

$$u(x|\psi, \alpha, \gamma) = \sum_{j=1}^{J} \psi_j (x_j + \gamma_j)^{\alpha_j}. \tag{1}$$

This utility function is valid if $\psi_j > 0$ and $0 < \alpha_j \leqslant 1$. The $\{\gamma_j\}$ parameters serve to translate the utility function. Typically, these parameters are not estimated but are set to 1 for identification reasons. Henceforth, we will use the value of 1 for these parameters. The marginal utility for the jth product is given by

$$u_j(x_j) = \psi_j \alpha_j (x_j + 1)^{\alpha_j - 1}. \tag{2}$$

(2) shows that both ψ and α influence marginal utility. However, we often think of the ψ parameters as influencing the "baseline" level of marginal utility, in the sense of scaling the profile of marginal utility up or down. The α parameter governs the rate of satiation. We note that we could reparameterize this model as

$$u_j(x_j) = \beta_j (x_j + 1)^{\delta_j}, \tag{3}$$

as in Rossi et al. (2005). The advantage of this parameterization is that the baseline level of marginal utility is uniquely associated with β_j while satiation is uniquely associated with δ_j. We show below that this parameterization is useful for studying the effects of product characteristics on satiation while holding constant the baseline utility.

With a large number of products, the utility function in (1) may be over-parameterized with different baseline utility parameters and satiation parameters for each product. It would seem natural then to project these parameters on to product attributes or characteristics. The proliferation of parameters problem is made even worse if one allows for individual specific consumer parameters. For most data sets, we will not be able to estimate a model in which all the parameters in (1) are consumer and product specific. Some judgment will have to be exercised in the utility parameterization. We can use model comparison methods to choose between alternative parameterizations, but it is naïve to expect that we will be able to estimate the full model with consumer specific parameters.

In linear utility (choice) models, it is a common practice to introduce product-specific intercepts. In situations with a large number of products, product attributes can be introduced instead of these intercepts (e.g. Fader and Hardie, 1996; Berry et al., 1995) to produce a more parsimonious model. Underlying this practice is a fundamental issue of how to view the space over which utility is defined. One view is that the space is of finite dimension that does not increase with the number of products. Product offerings are defined by their location in characteristics space. In this world, there are no truly new products in the sense of products, which include a new attribute valued by consumers. "New" products are simply different points in the characteristic space. As the number of products increases, this space becomes more densely packed with products, increasing the average level of product substitutability. This is a reasonable view for product categories where offerings are very similar and differ in terms of simple repackaging and slight enhancements.

An opposing view holds that product offerings are relatively unique and not well described by a low-dimensional space of attributes. That is, the dimension of the characteristics space is so much greater than the number of offerings that it is pointless to try to represent offerings in terms of their characteristics. This view is consistent with the notion that product attributes interact to produce a unique taste, or feel, that cannot be easily replicated. An implication of this view is that new products can be introduced that have attributes not available in existing products, and that these products create new sources of demand as they satisfy unmet consumer needs.

The challenge in comparing these views is that they cannot exist simultaneously. It is not possible to hold the view that a consumer's preference for an offering has both unique and common (attribute-related) components without restricting the nature of the unique component. If left unrestricted, the unique component saturates, or spans, the space of offerings and no additional preference information is available to understand the common component. This is similar to attempting to estimate both observation-specific intercepts and slope coefficients in a linear model. Assumptions used to identify such analysis may involve requiring the unique and common spaces to be orthogonal to each other, or that the unique component be restricted to be the same across consumers (e.g. BLP). In the analysis reported below, we investigate various projections of product offerings onto the characteristics space. We compare the fit of these characteristics-based models with models that have separate utility parameters for each product offering. Clearly, the more flexible models that do not project on characteristics space will have better in-sample fit. Thus, a critical issue we investigate is the degree to which various projections of the offerings onto the characteristics space degrades the fit of the model.

2.1. Introducing product characteristics

The logical approach to study the role of product attributes is to relate them to the baseline utility parameters in Eq. (1) (note that consumer characteristics are not entered directly into the utility function but, instead can be used to drive the heterogeneity distribution). Since the baseline utility parameters must be constrained to be positive, we relate the log of the baseline parameters to characteristics.

$$\ln(\psi_j) \equiv \psi_j^* = \sum_{k=1}^{K} \beta_k c_{j,k}. \tag{4}$$

Here there are K characteristics and $c_{j,k}$ is the level of characteristic k in product offering j.

While linear characteristics models have been popular in the economics literature (e.g. Berry et al., 1995; Berry and Pakes, 2002), there is no necessary reason why characteristics should enter linearly or even monotonically. For many applications, an ideal point model (e.g. Kamakura and Srivastava, 1986) might be more reasonable. That is, we expect that there is an "ideal" level of the characteristic for a given consumer. If the actual level of the characteristic deviates from the ideal level we expect, then marginal utility should decline.

$$\ln(\psi_j) \equiv \psi_j^* = -\sum_{k=1}^{K} \gamma_k |c_{j,k} - \theta_k|. \tag{5}$$

θ_k is the "ideal" point or optimal level of characteristic k. The negative sign in front of the summation ensures that lower levels of marginal utility are associated with characteristics further away from the ideal point. The idea point model creates a nonlinear relationship between the characteristics and the baseline utility parameters. One could simply postulate that the function in (3) is quadratic in characteristics.

$$\ln(\psi_j) \equiv \psi_j^* = \sum_{k=1}^{K} (\beta_k c_{j,k} + \tau_k c_{j,k}^2). \tag{6}$$

The problem with either the ideal point (5) or quadratic (6) forms is that they introduce a large number of parameters. The advantage is that the dimension of the parameter space is fixed at the number of characteristics and not the number of products.

While making the baseline utility parameters a function of characteristics is a natural extension of the characteristics models in the linear utility literature, characteristics might also influence the rate of satiation parameters. For example, different types of product packaging may facilitate higher rates of consumption or easy of storage as in beverage categories where large bottles are available along with packages of smaller containers such as six-packs. Certain combinations of product attributes may facilitate alternative uses of the product and this might induce a demand for larger quantities. For example, plain yogurt can be consumed directly or used in food preparation. Thus, for some applications, we will want to drive the satiation parameters as a function of product characteristics. The satiation parameters are identified by the quantity decision conditional on choice so that projecting the satiation parameters on a smaller set of characteristics can help identify these parameters, particularly in situations with a large number of product offerings. We can relate the log-odds transform of the satiation parameters to product characteristics.

$$\ln\left(\frac{\alpha_j}{1-\alpha_j}\right) = \sum_{k=1}^{K} \phi_k c_{j,k}. \tag{7}$$

2.2. The Lancasterian approach

As demonstrated above, it is straightforward to project the utility function parameters onto the observed characteristics of products. If the characteristics space is small and unchanging, this may prove to be a useful simplification. The Lancaster approach is even more restricted. Utility is only defined over the total amount of each characteristic obtained by the purchase of a bundle of goods. In linear utility models, the Lancaster approach is equivalent to simply projecting product intercepts on product characteristics. However, with non-linear models, the Lancasterian approach requires a different formulation of the utility structure as utility depends only on the aggregate quantities of characteristics consumed. We can write a variant of the Lancaster model using our translated power utility function. Define W as a $K \times J$ matrix with $w_{k,j}$ giving the level of characteristic k provided by product offering j. The total amount of characteristics obtained by a vector x of products is $z = Wx$. We can define utility over z using our specification to obtain a "Lancasterian" formulation of our model.

$$u(x) = \sum_{k=1}^{K} \omega_k \left(\sum_{j=1}^{J} w_{kj} x_j + 1 \right)^{\alpha_k}. \tag{8}$$

(8) is fundamentally a different model than our direct utility applied to product consumption and with the baseline or satiation parameters projected onto characteristics (as in (4) or (7)). Clearly, if there are interactions between characteristics the model in (8) may not be adequate. In our empirical analysis, we consider the demand for salty snacks such as potato chips or Doritos. Here there may be an interaction between cheese flavor and saltiness not captured in the additive model in (8). Although the basic model in (1) is additive in consumption of the products it is not additive in characteristics. In many product categories, the essence of product design will include consideration of interactions between characteristics or ideal points. Chan (2003) applies a model similar to (8) to data on purchase of soft drinks.

3. Statistical specification and estimation

Since disaggregate data is most often available as a panel of cross-sectional units observed over time, we formulate our statistical specification in two parts: (1) the "within" unit likelihood function and (2) a model of heterogeneity or "across" unit variation in parameters. This is what is termed a hierarchical model (for further discussion see Chapter 5 of Rossi et al., 2005). This joint model is estimated using a hybrid MCMC method.

3.1. Within unit likelihood

The likelihood for a given consumer's set of demands is derived by assuming a joint distribution for the vector of shocks to marginal utility for each product offering. These errors can be viewed as omitted characteristics influencing the marginal utility of each product offering. Some, c.f. BLP, recognize that there are unobserved characteristics that might influence demand and incorporate a uni-dimensional unobserved characteristic into linear utility specifications. The assumption that unobserved characteristics are uni-dimensional implies that the unobserved characteristics represent some sort of vertical quality difference between products. It may also be the case that the unobserved characteristics are multivariate, which would result in correlated utility errors.

We employ a likelihood-based method rather than a non-likelihood-based method, method of moments, as it facilitates inference regarding of individual unit-level demand parameters as well as common or "population" parameters. In addition, a likelihood-based approach has the advantage of imposing a relatively strict discipline in the sense that we rule out models with zero likelihood. For example, we could apply a standard multinomial model of product choice using a method of moments estimation technique, even though this model has zero likelihood for any dataset which contains a vector of demand with two or more non-zero elements. We do recognize that our particular parmeterization of utility may not be correctly specified. For this reason, we engage in extensive comparisons of alternative formulations in our empirical work. We believe that this, at least to some degree, insulates us from model specification.

Our approach to deriving the likelihood is to use the Kuhn–Tucker conditions from the utility maximization problem to derive the distribution of quantity demanded. We only briefly summarize our approach, for details see Kim et al. (2002). The Kuhn–Tucker conditions from maximimization of (1) w.r.t the standard budget constraint are:

$$\frac{\partial u(\mathbf{x}^*; \boldsymbol{\psi}, \boldsymbol{\alpha})}{\partial x_j} = \lambda p_j \quad \text{if } x_j^* > 0, \tag{9}$$

$$\frac{\partial u(\mathbf{x}^*; \boldsymbol{\psi}, \boldsymbol{\alpha})}{\partial x_j} \leqslant \lambda p_j \quad \text{if } x_j^* = 0, \tag{10}$$

where λ is a Lagrange multiplier. We introduce a multiplicative error (ε) into the expression for the marginal utility in a manner similar to that encountered in discrete choice models:

$$\frac{\partial u(\mathbf{x}^*; \boldsymbol{\psi}, \boldsymbol{\alpha})}{\partial x_j} \mathrm{e}^{\varepsilon_j} = \lambda p_j.$$

Rearranging terms and taking logs of both sides gives:

$$V_j + \varepsilon_j = \ln \lambda \quad \text{if } x_j^* > 0, \tag{11}$$

$$V_j + \varepsilon_j < \ln \lambda \quad \text{if } x_j^* = 0, \tag{12}$$

where

$$V_j = \ln\left[\frac{\partial u(x^*; \psi, \alpha)}{\partial x_j}\right] - \ln(p_j).$$

Eqs. (11) and (12) are similar to those encountered with standard choice models. The term, V_j, is the log of marginal utility divided by price. In a traditional discrete choice model, log marginal utility is assumed constant and does not depend on the quantity demanded. In our model, the marginal utility depends on quantity because of the parameter, α.

Our goal is to derive the distribution of observed demand, x^*. This is done by assuming an error distribution for ε and employing the Kuhn–Tucker conditions in Eqs. (11) and (12) to obtain the distribution of x^* using change-of-variable calculus. In our analysis and model development, we assume that ε is normally distributed. The presence of the budgetary constraint introduces a singularity in mapping from ε to x^* because the observed demand multiplied by price must add up to the budget amount, $p'x = y$. Hence the dimension of x^* is one less than the dimension of ε. Without loss of generality, we re-label the alternatives such that the zeroth alternative, formally the outside good, is always chosen, and consider the Kuhn–Tucker conditions for the mapping $h_j(x^*, p) = V_0 - V_j$ for $j = 1, .., J$:

$$v_j = h_j(x^*, p) \quad \text{if } x_j^* > 0, \tag{13}$$

$$v_j < h_j(x^*, p) \quad \text{if } x_j^* = 0, \tag{14}$$

where $v_j = \varepsilon_j - \varepsilon_0$. The likelihood function of the data is a mixture of density ordinates (Eq. (13)) and point masses (Eq. (14)) corresponding to nonzero and zero demand, respectively. Assuming that the first n of J alternatives, in addition to the zeroth alternative, has non zero demand:

$$\begin{aligned} &P(x_i^* > 0 \text{ and } x_j^* = 0; i = 1, \ldots, n \text{ and } j = n+1, \ldots, J) \\ &\quad = \int_{-\infty}^{J} \cdots \int_{-\infty}^{n+1} \phi(h_1, \ldots, h_n, v_{n+1}, \ldots, v_J | \mathbf{\Omega}) |\mathbf{J}| \mathrm{d}v_{n+1} \cdots \mathrm{d}v_J, \end{aligned} \tag{15}$$

where $\phi(\cdot)$ is normal density, $h_j = h_j(x^*; p)$, Ω is the covariance matrix of the differenced errors with element ε_j–ε_0, and J is the Jacobian of the transformation from ε_j–ε_0 to x^*:

$$J_{ij} = \frac{\partial h_{i+1}(x^*; p)}{\partial x_{j+1}} \quad i, j = 1, \ldots, n.$$

The likelihood for the Lancasterian model in (8) can be derived in a similar fashion. However, the Jacobian elements for the products that have positive demand are different and given by

$$J_{ij} = \frac{(\alpha_0 - 1)}{(x_0 + \gamma)}(-p_{i+1}) - \frac{1}{\kappa_{i+1}} \left(\sum_k \omega_k \alpha_k (\alpha_k - 1) w_{k,i+1} w_{k,j+1} (z_k + \gamma)^{\alpha_k - 2} \right) \tag{16}$$

where $\kappa_j = \sum_k \omega_k \alpha_k w_{k,j} (z_k + \gamma)^{\alpha_k - 1}$ and $z_k = \sum_{j=1}^J w_{k,j} x_j$.

We emphasize that the Kuhn–Tucker are written in the offering space, not the characteristics space, relating observed demand (x^*) and observed prices (p) to primitive error assumptions (ε) through the first-order conditions. One could consider a version of the Lancasterian model that would assume that the errors, or demand shocks, originate from the characteristics space, leading to first-order conditions with a more complicated latent density (ϕ) and regions of integration. The advantage of specifying the error distribution in the offering space is that it leads to a simpler likelihood specification involving observed prices and represents a less restrictive model in the sense that the error distribution is of higher dimension than the number of characteristics ($K < J$).

3.2. Heterogeneity

We recognize that consumers have different preferences and these differences are not well captured by observable consumer characteristics. For this reason, we employ a distribution of parameters across consumers in the panel. As mentioned above, even if we use a fairly standard distribution such as a joint normal distribution, we cannot hope to make all of the parameters in models (1), (4)–(7) vary across consumers. Our approach is to concentrate on a subset of parameters and make only this subset vary across consumers to achieve a compromise between flexibility and parsimony. In many, but not all situations, we view the curvature parameters (α) of the utility function as the most difficult to estimate as these require variation in the quantity purchased. For this reason, we will start by considering specifications in which either the baseline parameters are heterogeneous or the parameters relating characteristics to the baseline are heterogeneous.

$$\ln \psi_h \sim N\left(\bar{\psi}, \Sigma_\psi\right). \tag{17}$$

Here h indexes the consumer. For model (3) in which characteristics are related linearly to the log baseline parameters, $\ln \psi = C\beta$, we use the model

$$\beta_h \sim N\left(\bar{\beta}, \Sigma_\beta\right). \tag{18}$$

For the ideal point model in (5), we make the ideal points vary across consumers:

$$\ln \theta_h \sim N\left(\bar{\theta}, \Sigma_\theta\right). \tag{19}$$

4. Empirical applications

We will consider two different datasets to illustrate how characteristics information can be incorporated into our model. The first data set is the result of a field experiment in which subjects were observed to purchase different varieties of salty snacks. This category of products has well-defined characteristics, which the manufacturer manipulates to create ‘‘new’’ products. In this application, we will compare variants of our model with ideal point and Lancasterian models. In the second application, we apply our methods to conjoint data on the demand for different packages of the same variety of a fruit drink. Since the product is the same across the offerings and only the packaging changes, this application emphasizes the importance of characteristics explaining satiation. An advantage of using these datasets to study the relationship between characteristics and demand is that confounding effects due to factors such as product price endogeneity, couponing and product stockouts are absent.

4.1. Application 1: driving baseline utility with product characteristics

This data set was generated by a field experiment involving undergraduate students from a large Midwestern university. Students were screened for inclusion in the experiment if they reported that they frequently purchased offerings in the product category—salty snacks. Each week, the students were allocated a \$2.00 budget and asked to make purchases from among eight varieties of Doritos brand corn chips. The offerings were priced at \$0.33, allowing the students to select up to 6 bags each week. The regular price of the corn chips was \$0.99. The students were told that any unused budget allocation would be paid in cash at the end of the experiment. By offering the chips at reduced prices, we hoped to induce higher levels of consumption, which might provide information about satiation. Students were instructed to purchase the chips for their own consumption, not for the consumption of others.

Product characteristic information was provided by the manufacturer of Doritos brand corn chips, Frito Lay. Table 1 displays a list of the offerings and associated characteristics. The characteristics are disguised for proprietary purposes, but reflect summary taste

Table 1
Varieties and characteristics

Variety	Characteristic					
	C_1	C_2	C_3	C_4	C_5	C_6
Nacho Cheesier 3D	1.00	0.00	0.00	4.00	4.00	2.00
Spicier Nacho	0.00	0.00	1.00	3.00	3.83	4.25
Cooler Ranch 3D	1.00	3.17	0.00	0.00	4.17	4.00
Baja Picante	0.00	0.00	1.33	0.88	3.33	5.50
Jalapeno Cheddar 3D	1.00	0.00	1.83	4.25	4.17	5.00
Nacho Cheesier	0.00	0.00	0.00	4.00	4.00	2.00
Cooler Ranch	0.00	3.17	0.00	0.00	4.17	4.00
Sonic Sour Cream	0.00	0.67	0.00	3.38	3.83	0.00

Table 2

Variety	Purchase incidence	Total purchase quantity
(a) Purchase incidence and quantity		
Nacho cheesier 3D	168	224
Spicier nacho	177	262
Cooler ranch 3D	188	231
Baja picante	180	235
Jalapeno cheddar 3D	190	295
Nacho cheesier	244	446
Cooler ranch	235	338
Sonic sour cream	218	277

Variety	Observations	Single item purchase	Multiple items purchase
(b) Frequency of single-item and multi-item purchase for multiple unit purchases			
Nacho cheesier 3D	168	—	168 (1.00)
Spicier nacho	177	4 (.02)	173 (0.98)
Cooler ranch 3D	188	—	188 (1.00)
Baja picante	180	—	180 (1.00)
Jalapeno cheddar 3D	190	2 (.01)	188 (0.99)
Nacho cheesier	244	6 (.02)	238 (0.98)
Cooler ranch	235	—	235 (1.00)
Sonic sour cream	218	—	218 (1.00)

characteristics such as "citrus", "red pepper" and "treated corn" that are meaningful to the manufacturer.

The experiment was conducted over a seven-week period, resulting in a total of 634 observations for the 101 subjects. The data for each purchase occasion is comprised of a vector of purchase quantities of each of the eight Dorito corn chip varieties, and the quantity of the outside good that was set equal to the unspent budget allocation.

Summary statistics of the data are reported in Table 2. Seven percent of the observations were corner solutions in which only one of the eight varieties was selected, and 93% of the observations were interior. The varieties were purchased from 244 to 168 times, with an expected demand of more than one package. The data indicate that there is no one dominant product offering or characteristic. The varieties with the highest and lowest incidence are both "Nacho Chessier", but differ in terms of their shape (flat versus 3-dimensional), indicating that preferences are related to the characteristics of the offerings.

Five models were fit to the data. The first model serves as a benchmark and does not attempt to relate product characteristics to the parameter ψ in Eq. (1). Consumer heterogeneity is modeled with a random-effects specification on the ψ parameter. The remaining models specify ψ as a function of product characteristics. Since the number of characteristics (6) is less than the number of offerings (8), the fit of these models is expected to be worse than the first model, which specifies ψ as unrestricted. Models 2–4 relate the product characteristics to the ψ parameter using (2), with heterogeneity of characteristic importance weights (β) as in (18). Model 2 allows for a unique curvature parameter, α, for each choice alternative. Model 3 restricts the curvature parameter to be the same for all alternatives, including the outside good. Model 4 allows α to be different for the outside good.

The fifth model relates ψ to the characteristics using an ideal point specification similar to (5) with α different for each choice alternative. The ideal point specification in (5) includes not only K ideal point parameters (θ) but also different weights for each deviation from the idea point. If the ideal points are consumer specific, then we already have introduced considerable flexibility into the model. To add the γ weights for each deviation would be asking too much from this data.

The log marginal density (e.g., Newton and Raftery, 1994) for models 1–5 is reported in Table 3. The fit statistics indicate that model 4 is preferred characteristics model, with a linear relationship between the product characteristics and utility parameter ψ. This model fits better than any of the other characteristics models, including the ideal point model. Moreover, assuming that the curvature parameter α is the same for all flavors, but different for the outside good, is preferred. As discussed below, this does not imply that the rate of satiation is identical for all brands. The rate of satiation is the second derivative of the utility function, which is a function of ψ (or β) and α.

Even though there is no natural way to nest the Lancasterian and our standard models, we can use marginal likelihood to compare non-nested models (see Chapter 6, Rossi et al., 2005). The characteristics model has a marginal likelihood of −5426 showing that this model is nowhere in the ballpark of our other models which incorporate characteristics in

Table 3
Model fit

Model	Specification	Log marginal density[a]
1	$\boldsymbol{\psi}_h^* \sim N(\bar{\boldsymbol{\psi}}^*, \boldsymbol{\Sigma}_\psi)$; α_j unique for each flavor	−3767
2	$\psi_j^* = \Sigma_k \beta_k c_{jk}$; $\boldsymbol{\beta}_h \sim N(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_\beta)$; α_j unique for each flavor	−3820
3	$\psi_j * = \Sigma_k \beta_k c_{jk}$; $\boldsymbol{\beta}_h \sim N(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_\beta)$; α_j common	−3814
4	$\psi_j * = \Sigma_k \beta_k c_{jk}$; $\boldsymbol{\beta}_h \sim N(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_\beta)$; α_j, $j = 1$(inside), 2(outside)	−3810
5	$\psi_j * = \Sigma_k \lvert c_{jk} - \theta_k \rvert$; $\boldsymbol{\theta}_h^* \sim N(\bar{\boldsymbol{\theta}}^*, \boldsymbol{\Sigma}_{\theta *})$; α_j unique for each flavor	−3905
Lancasterian model	$u(x) = \sum_{k=1}^{K} \omega_k \left(\sum_{j=1}^{J} w_{kj} x_j + 1 \right)^{\alpha_k}$	−5426

[a]Computed using the importance sampling method of Newton and Raftery (1994).

the utility weights but do not specify utility over the level of a characteristic directly. The last row in Table 3 provides the log-marginal density for the Lancasterian Model and shows that it has a dramatically poorer fit and posterior probability than any of our parameterizations in models 1–5.

Parameter estimates for model 4, the best fitting characteristics model, are reported in Table 4. Reported in the upper portion of the table is the mean and covariance matrix of β, the coefficients that relate the product characteristics to the parameter ψ. Estimates of the curvature parameters, $\alpha^* = \ln(\alpha/1-\alpha)$, are reported in the lower portion of the table. The mean of the random-effects distribution is reported in the upper left portion, and the covariance matrix in the upper right portion of the table. The upper triangular region of latter reports correlations rather than covariances. In general, the parameters are estimated precisely, nearly all have substantial posterior mass away from zero.

The mean of the random-effects distribution has both positive and negative elements. A positive element indicates that the majority of respondents in the survey prefer to have more of the characteristic than less, while a negative coefficient indicates that a majority would prefer to have less of the characteristic. The most favorable characteristic, on average, is C_4 with a coefficient of 0.56, and the most disliked characteristic is C_5 with a coefficient of -1.57. Heterogeneity around these mean levels is large, with some respondents favoring and some respondents disliking each of the characteristics. The most diverse preferences are for characteristic C_3 with a variance of 2.55. In addition, there are a number of large negative covariances, with respondents favoring either C_3 or C_6 but not both, and either C_4 or C_5, but not both.

Our model of product characteristics, preference and satiation is useful for understanding the impact of changes in the levels of characteristics on expected demand quantities. These quantities are determined by the marginal utility consumers derive from the consumption of varieties relative to the marginal utility of consuming the outside good. However, as shown in (2), both the ψ and α parameters influence marginal utility as well as the curvature of the utility function or the extent of diminishing returns. In order to

Table 4
Parameter estimates (posterior standard deviation)

Characteristics	Mean (std.dev)	**Covariance***Correlation*					
		C_1	C_2	C_3	C_4	C_5	C_6
(a) $\bar{\beta}$ and $\sum_\beta$							
C_1	−0.40 (.07)	1.40 (.25)	*−.29*	*.37*	*−.40*	*.31*	*−.38*
C_2	.62 (.10)	−.22 (.12)	.41 (.10)	*−.02*	*.49*	*−.50*	*.15*
C_3	−0.43 (.17)	.71 (.2)	−.02 (.17)	2.55 (.51)	*−.62*	*.62*	*−.81*
C_4	0.56 (.09)	−.32 (.12)	.21 (.09)	−.68 (.19)	.46 (.10)	*−.82*	*.68*
C_5	−1.57 (.18)	.30 (.14)	−.26 (.10)	.80 (.23)	−.45 (.12)	.65 (.16)	*−.71*
C_6	.24 (.07)	−.29 (.10)	.06 (.07)	−.83 (.19)	.30 (.08)	−.37 (.10)	.42 (.08)

Variety	Mean (stand. dev.)
(b) α^*	
Inside-good	−2.78 (.20)
Outside-good	−4.15 (.37)

Table 5
Expected change in marginal utility

Characteristic (k)	$\partial U'j/\partial c_k$				
Variety (j)	C_2	C_3	C_4	C_5	C_6
Nacho cheesier 3D	—	—	0.52	−1.20	0.27
Spicier nacho	—	−0.43	0.62	−1.39	0.36
Cooler ranch 3D	0.53	—	—	−1.06	0.13
Baja picante	—	−0.04	0.39	−1.10	0.21
Jalapeno cheddar 3D	—	0.22	0.39	−1.14	0.10
Nacho cheesier	—	—	0.77	−1.55	0.46
Cooler ranch	0.78	—	—	−1.48	0.33
Sonic sour cream	0.62	—	0.48	−1.17	—

facilitate interpretation of our model parameters, we explore the impact of changes in the level of characteristics on the marginal utility and compute the elasticity of demand with respect to each characteristic.

Table 5 reports the expected change in the marginal utility for changes in the levels of the characteristics. An entry of zero is entered in the table for characteristics that are not present in the variety (see Table 1), and for the first characteristic (C_1) that is a dummy variable that represents 3D versus a flat chip. To compute these expectations, we integrate over the posterior distribution of model parameters.

An increase in the characteristics C_2, C_4 and C_6 lead to increases in marginal utility and, consequently, demand for the varieties. An increase in characteristics C_3 and C_5 leads to a reduction in demand for nearly all the varieties offered. The exception is characteristic C_3 for Jalapeno Cheddar 3D where the change in the gradient is positive, not negative. Recall from Table 4 that the mean of the random-effects distribution is positive for C_2, C_4 and C_6, reflecting general preference for more of these characteristics in the sample. The mean for C_3 and C_5 is estimated to be negative, indicate that most respondents do not prefer more of the attribute. The results in Table 5 illustrate that the log–linear relationship between the characteristics and the utility parameter ψ, when coupled with heterogeneity in preferences, results in a flexible model specification capable of representing a complex pattern of tastes.

We can also estimate the change in expected demand by comparing the quantity, x^*, that maximizes consumer utility for various specifications of product characteristics. Table 6 reports the elasticity of demand with respect to each of the characteristics. The entries in the table are computed by changing each characteristic by one percent and computing expected demand by summing over consumers and purchase occasions. Increases in characteristic C_5 lead to large reductions in demand for all varieties, while an increase in the other characteristics leads to a nearly uniform increase in demand.

4.1.1. A comparison to the Lancasterian approach

As discussed in Section 2, we can formulate a Lancasterian version of our model of demand by postulating utility over characteristics using our translated nonlinear utility function (8). We use all characteristics except C_1 (this is binary) in our Lancasterian variant. We accommodate consumer heterogeneity by specifying that the vector of baseline characteristic utility weights is distributed log-normally over consumers,

Table 6
Elasticity of demand with respect to characteristics

Characteristic (k)	$\partial \ln x_j^* / \partial \ln C_k$				
Variety (j)	C_2	C_3	C_4	C_5	C_6
Nacho cheesier 3D	—	—	4.19	−11.37	1.04
Spicier nacho	—	−2.04	5.54	−16.05	5.97
Cooler ranch 3D	4.86	—	—	−13.98	0.67
Baja picante	—	0.60	0.43	−7.17	1.75
Jalapeno cheddar 3D	—	1.72	2.51	−8.51	−0.15
Nacho cheesier	—	—	6.31	−11.66	2.13
Cooler ranch	4.69	—	—	−11.00	2.27
Sonic sour cream	0.96	—	3.17	−9.26	—

Table 7
Parameter estimates for lancasterian model

Characteristic	Mean (std. dev)	**Covariance** *Correlation*					
		C_1	C_2	C_3	C_4	C_5	C_6
(a) $\bar{\omega}^*$ and $\Sigma_{\omega *}$							
C_1	−3.39 (.03)	.32 (.05)	*.05*	*.70*	*.62*	*−.30*	*.56*
C_2	−3.12 (.13)	−.03 (.08)	1.40 (.28)	*−.03*	*.23*	*−.14*	*.54*
C_3	−3.41 (.11)	.48 (.10)	−.05 (.16)	1.43 (.29)	*.62*	*−.27*	*.60*
C_4	−2.65 (.10)	.41 (.09)	.32 (.18)	.87 (.18)	1.36 (.21)	*−.30*	*.74*
C_5	−4.65 (.09)	−.10 (.05)	−.10 (.15)	−.19 (.10)	−.21 (.11)	.35 (.08)	*−.26*
C_6	−3.35 (.15)	.38 (.12)	.78 (.26)	.88 (.24)	1.05 (.22)	−.19 (.13)	1.50 (.36)
(b) Satiation parameter:							
$\alpha = .003$							
(.003)							

$\ln \omega = \omega^* \sim N(\bar{\omega}^*, \Sigma_{\omega^*})$. Chan (2003) assumes that these parameters have a scalar covariance matrix and uses a method of moments estimation approach. The method of moments approach does not allow for comparison of alternative models as our Bayesian approach does. In addition, we can fit a full covariance structure at no greater computation cost than a scalar structure.

Table 7 provides the results of our fit of the Lancasterian model. The second column provides our inferences about the average value of the log of the utility weight parameter, ω^*. The characteristic utility weights are all precisely estimated but small in magnitude (recall that the weights are the exponential of the log parameters presented in Table 7). This means that there is a detectable relationship but it is of little substantive significance. Examination of the covariance matrix and associated correlations shows large differences between consumers in their preferences for C_4 and C_6. In addition, there are high intercorrelations between characteristics. It is interesting that there are some very large negative as well as positive correlations. Consumers who like C_1 dislike C_5 while those who like 4 also tend to like 6. The assumption of a scalar covariance structure would be unreasonable for this data set.

4.2. Application 2: product characteristics and satiation

As noted above, product characteristics can drive either the baseline utility parameters or the satiation parameters. In some situations, the same basic product is packaged in different sizes and/or containers. In these situations, it may be useful to bring package characteristics into the satiation parameters. We now consider data from a conjoint survey in which consumers were asked to choose and indicate quantity demand from alternative product offerings of the same product (fruit juice) in different package sizes and form. It should be emphasized that a model with satiation is required to make volume forecasts for situations in which consumers purchase more than one unit of the good.

We employ the alternative parameterization of our model (Eq. (3)) to isolate the effects of packaging on satiation: $u_j(x_j) = \beta_j(x_j + 1)^{\delta_j}$. The β_j coefficients are constrained to be equal across choice alternatives so that baseline preference is equal, and an outside good is introduced to allow for a no-purchase response by the respondent. The corresponding utility function is

$$u(x) = \sum_{j=1}^{J} \frac{1}{\alpha_j}(x_j + \gamma)^{\alpha_j} + \psi_0 \frac{1}{\alpha_0}(x_0 + \gamma)^{\alpha_0}. \tag{20}$$

Here x_0 represents the outside good, $\alpha = 1 - \delta$, and we restrict $\psi_j = \psi = 1$ for $j = 1, \ldots, J$ for identification purposes. We do not introduce separate ψ parameters for each of the inside goods as in our application, each inside good is the same juice product, simply packaged differently. As in (7), we allow package attributes to drive a log-odds transformation of the satiation parameters.

The data were obtained as part of a study by a beverage company seeking to identify optimal package configurations. The manufacture anticipated that the major retailer chains would allow access to sufficient shelf space so that two stock keeping units could be offered. A survey was developed by the beverage company and administered to respondents from Baltimore, Chicago, Los Angeles and Florida, comprising a nationally representative sample ($n = 289$).

Table 8 presents the packaging attributes under study, along with the dummy variable coding associated with the attribute-levels. The packaging attributes (and levels) are package type (can versus bottle), container volume (12 ounce, $\frac{1}{2}$l, or 24 ounce) and package size (4, 6 and 12 pack). An interesting issue in volumetric analysis is in determining the appropriate metric to measure volume, corresponding to the quantity "x" in Eqs. (1)–(3) above. The quantity "x", for example, could refer to total fluid ounces, the number of containers, or the number of packages. If a respondent selects two 4-packs of 24 oz. bottles, quantity could be coded as 192 ounces, or 8 bottles, or 2 packages. A purpose of volumetric conjoint analysis, and conjoint analysis in general, is to identify the attribute-levels that lead to greatest utility and demand for an offering. Since the utility function in Eq. (20) is not linear in the parameters, the scale of observed demand will lead to different model fits and policy implications. The analysis presented below explores this issue with alternative models that interpret the observed demand data in different ways.

Table 8
Attribute-level coding

Attribute name/level	z_1	z_2	z_3	z_4	z_5
(A) Type (z_1)					
A1. Can	1				
A2. Plastic bottle	0				
(B) Container volume (z_2, z_3)					
B1. 12 oz		0	0		
B2. 1/2 litter		1	0		
B3. 24 oz		0	1		
(C) Package size (z_4, z_5)					
C1. 4 pack				0	0
C2. 6 pack				1	0
C3. 12 pack				0	1

Each respondent was exposed to 17 choice scenarios comprised of two choice offerings and a no-buy option. For each scenario, the respondent instructions read ‘‘If you were at your grocery store and these were the only BRAND products available, READ A THEN B.’’

A	B	Neither
4 pack of 24 ounce bottles $2.49	12 pack of 12 ounce cans for $3.29	Not buying a beverage at this time.

The respondent was then asked to record the number of packages of A and B they would buy at this time, corresponding to the quantity demanded (x^*). Across the 17 choice scenarios, eight different package configurations were investigated. Table 2 provides a list of the unique configurations of the packages. Price was varied across all the choice scenarios so that none of the product descriptions appeared twice.

The 4898 observations that comprise the dataset have the following characteristics. Respondents selected the ‘‘no-purchase’’ option 20% of the time, and indicated that they wanted both offerings (A and B) 20% of the time. For those occasions in which at least two units were demanded, nearly 50% of the responses correspond to demand for both offerings A and B. Moreover, the reported demand for these multiple offerings is distributed both within and across respondents, involving all of the product configurations in Table 9. Thus, packaging characteristics, i.e., type, container volume and package size, appear to play a significant role in consumer demand. For this reason, total number of fluid ounces is not sufficient to describe the attributes of the good demanded and we must allow for utility to be influenced by package type.

Three alternative model specifications are investigated, differing in terms of how demand (x) is measured. The models are summarized in Table 10. Consider, for example, a respondent that reports that they want two 4-packs of 12-ounce cans. This response can be coded in terms of the total fluid ounces (96 oz.), the number of containers (eight) or the

Table 9
Choice offering profiles

Configuration	Attribute				
	Can (z_1)	$\frac{1}{2}$l (z_2)	24 ounce (z_3)	6 pack (z_4)	12 pack (z_5)
1	0	1	0	0	0
2	0	0	1	0	0
3	1	0	0	1	0
4	0	0	0	1	0
5	0	1	0	1	0
6	0	0	1	1	0
7	1	0	0	0	1
8	0	0	0	0	1

Note: Default coding is a 4-pack of 12 ounce bottles.

Table 10
Alternative model specifications

Model	Measurement of demand (×) for 2 four-packs of 12 oz cans	Attributes affecting satiation (α)	Log marginal density
Fluid	96 ounces	Type (can vs. bottle), container volume, Package size	−14,873.8
Container	Eight cans	Type, container volume, package size	−11,997.7
Package	Two packages	Type, container volume, package size	−10,067.9

number of packages (two) demanded, associating the remaining characteristics (e.g., container size) with the rate of satiation (α). The first model, named “fluid”, assumes that demand is measured in terms of the total fluid ounces contained in the packages (e.g., $x = 96$). The other characteristics of the offering are used as explanatory variables for satiation. The second model, named “container”, assumes that demand is measured in terms of number of containers (e.g., $x = 8$), and the third model, “package” measures demand in terms of the number of packages (e.g., $x = 2$). By measuring demand in these three ways, we can check which conceptualization of volumetric response is most consistent with the responses in the survey. Reported on the right side of Table 3 is the log marginal density, a Bayesian measure of model fit (e.g., Newton and Raftery, 1994). The fit statistic indicates that the package model fits the data best.

Table 11 provides parameter estimates for the “package” model. Reported is the mean and covariance matrix of the distribution of heterogeneity. Posterior standard deviations are reported in parentheses. Positive parameter estimates of ϕ in Eq. (7) lead to values of the satiation parameter, α, closer to one, and negative estimates lead to values of α closer to zero. Values of α closer to one indicate less curvature in the utility function, implying less satiation, while value of α closer to zero indicate greater satiation. As satiation increases, optimal demand (x^*) is smaller holding all else constant.

The parameter estimates indicate that, on average, the can and $\frac{1}{2}$l attributes lead to more satiation, while the 6 and 12 pack parameters lead to less satiation, holding fixed the other

Table 11
Parameter estimates (posterior standard deviation) for volumetric model

Attribute (z)	Parameter[a]	Mean	**Covariance***Correlation*						
			ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ψ_0^*	α_0^*
Can	ϕ_1	−1.40 (.05)	.87 (.11)	*.10*	*.10*	*.10*	*.08*	*.13*	*.05*
$\frac{1}{2}$1	ϕ_2	−1.20 (.23)	.09 (.07)	1.04 (.09)	*.23*	*.23*	*.22*	*.16*	*−.01*
24 ounce	ϕ_3	0.22 (.31)	.10 (.07)	.26 (.02)	1.23 (.10)	*.23*	*.23*	*.16*	*−.09*
6 pack	ϕ_4	1.42 (.27)	.10 (.07)	.25 (.02)	.27 (.02)	1.22 (.10)	*.23*	*.17*	*.05*
12 pack	ϕ_5	3.65 (.22)	.07 (.07)	.22 (.02)	.24 (.02)	.23 (.02)	.92 (.08)	*.17*	*−.01*
Outside-good	ψ_0*	−0.85 (.05)	.09 (.06)	.12 (.01)	.13 (.02)	.13 (.01)	.12 (.01)	.56 (.06)	*.09*
Outside-good	α_0*	0.01 (.08)	.05 (.08)	−.01 (.08)	−.10 (.08)	.05 (.08)	−.01 (.07)	.07 (.06)	1.15 (.17)

[a]Part-worth (ϕ) are relative to a 4-pack of 12 ounce bottles.

Table 12
Parameter estimates (posterior standard deviation) for poisson regression model

Attribute (z)	Parameter	Mean (μ)	**Covariance***Correlation*						
			ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
Intercept	ϕ_0	.83 (.10)	.60 (.15)	*.02*	*−.17*	*−.16*	*−.05*	*.30*	*−.56*
Can	ϕ_1	−.45 (.07)	.02 (.11)	.80 (.16)	*.13*	*.20*	*−.14*	*−.10*	*−.16*
$\frac{1}{2}$1	ϕ_2	−.27 (.08)	−.13 (.08)	.12 (.08)	.97 (.12)	*.88*	*−.27*	*−.25*	*−.05*
24 ounce	ϕ_3	−.14 (.09)	−.13 (.10)	.20 (.09)	.94 (.13)	1.19 (.20)	*−.21*	*−.16*	*−.12*
6 pack	ϕ_4	.35 (.05)	−.02 (.04)	−.05 (.06)	−.12 (.06)	−.10 (.07)	.19 (.03)	*.54*	−.22
12 pack	ϕ_5	.70 (.09)	.17 (.08)	−.07 (.11)	−.18 (.10)	−.12 (.12)	.17 (.05)	.52 (.10)	−.55
Price	ϕ_6	−.55 (.04)	−.17 (.06)	−.06 (.05)	−.02 (.04)	−.05 (.05)	−.04 (.02)	−.16 (.04)	.16 (.03)

attributes. The covariance matrix of random-effects is primarily diagonal with approximately unit variances, indicating that there exists substantial heterogeneity in the parameter estimates.

Table 12 reports estimates for a heterogeneous Poisson model fit to the quantity data. The Poisson model differs from the proposed model in that the likelihood comprises point mass probabilities, whereas the likelihood for the proposed model is a mixture of point masses corresponding to corner solutions and densities corresponding to interior solutions. Hence, likelihood values and corresponding log marginal density estimates are not comparable. Below we report on a predictive comparison using mean absolute deviations as a measure of predictive fit. The Poisson model is specified as

$$y_{h,j,t} \sim \text{Poisson}(\lambda_{hjt}),$$

$$\ln(\lambda_{hjt}) = \phi_h' z_{hjt}.$$

$y_{h,j,t}$ is the quantity of jth offering($j = 1, 2$) that respondent h chose at time t, and z_{hjt} is the vector of the attributes and the price for offering j. The coefficients ϕ_h are distributed multivariate normal. The algebraic signs of the estimates for the Poisson are generally in agreement with those of the proposed volumetric model.

Table 13
Predictive results

Model	Predictive mean absolute deviation
Fluid	0.93
Container	0.71
Package	0.69
Poisson	0.84

Table 13 reports the results of predictive tests on holdout samples. Two observations per respondent were reserved for holdout prediction. We investigate the predictive performance of four models: the three models reported in Table 10, plus a heterogeneous Poisson model fit to the quantity data. The results indicate that the Package model has the best predictive fit. The predictive performance of the Container model is slightly worse. The Poisson and Fluid models have substantively worse predictive performance.

5. Optimal product offerings

We illustrate the implications of the model (given in (20)) by considering the assortment selection problem faced by the manufacturer. As discussed above, the manufacturer originally initiated the present study to determine which package configurations would result in greatest expected profit. In addition, the manufacturer was constrained by retailers to offering just two different packages (labeled "A" and "B" in the tables and figures). Cost information regarding the contribution margins, and suggested retail prices, were provided by the manufacturer, and analysis using alternative model specifications proceeded by identifying expected profits for each of the 28 different combinations (i.e., eight choose two) of offerings. Expected profit for a given set of prices is computed by integrating the profit function with respect to the distribution of model parameters.

$$E[\text{Profit}|p] = \int \left[\sum_j x_j^*(\boldsymbol{\theta}_h)(p_j - MC_j) \right] \pi(\boldsymbol{\theta}_h|\text{data})\text{d}\boldsymbol{\theta}_h, \tag{21}$$

where θ_h is the parameter vector for respondent h, x^* is quantity vector associated with the utility maximizing solution, and MC_j is the marginal cost of offering j.

Table 14 presents the expected profit for each of the 28 possible product line configurations, sorted from most to least profitable. The eight alternative product profiles are provided in Table 9. The most profitable combination of offerings is offering "3" in conjunction with offering "4", where the former is a 6-pack of 12-ounce cans, and the latter is a 6-pack of 12-ounce bottles. In general, the most profitable combinations include configuration "4" in combination with the other offers.

Fig. 1 compares the expected profit computed from the proposed volumetric model to one from a heterogeneous Poisson model by plotting the expected profit for each of the 28 package combinations. The offer combinations are listed below the figure, and refer to the profiles listed in Table 2. The volume equivalent of configurations one through eight in Table 2 are 68, 96, 72, 72, 101, 144, 144, and 144 ounces, respectively. The expected profits from the volumetric model are up to 40 percent higher than the Poisson model for the

Table 14
Expected profit for offering combinations

Configuration A	Configuration B	Expected profit ($)
3	4	19,391.78
1	4	18,892.00
4	5	18,867.34
4	6	18,769.87
4	8	18,766.30
2	4	18,530.10
4	7	18,493.33
1	3	17,986.18
3	5	17,908.46
3	6	17,839.38
3	8	17,709.24
2	3	17,660.43
3	7	17,291.10
1	2	17,167.10
1	5	17,148.38
1	6	16,886.05
2	5	16,755.75
1	8	16,697.11
5	8	16,610.18
2	8	16,550.12
5	6	16,521.74
2	6	16,519.39
1	7	16,210.17
5	7	16,110.73
2	7	16,068.76
6	8	15,939.69
6	7	15,379.27
7	8	14,016.42

small-volume offer combinations on the left side of the figure, and converge to the same expected profits for the larger volume combinations and the right side. The Poisson model predicts that larger volume combinations will be more profitable than offers combinations with smaller volume (i.e., the graph has positive slope). In contrast, the volumetric model predicts that the smaller volume offerings will be more profitable (i.e., negative slope).

The most profitable offer combinations predicted by the volumetric model, which include configuration "4", are plotted with a triangle in Fig. 1. While some of the most profitable offer combinations predicted by the Poisson model also include configuration "4", many do not, and relative ranks of the offer combinations can be very different across models. For example, the second most profitable combination predicted by the volumetric model comprises configurations "4" and "1". This combination is ranked 14th by the Poisson model.

Fig. 2 compares the expected demand for volumetric and Poisson models, measured in fluid ounces. The two model match in their prediction for large volume combinations, but diverge greatly for the small volume combinations plotted on the left side of the figure. The expected demand from the volumetric model is up to 50 percent larger than the Poisson model for some offer combinations. By capturing the effects of substitution among

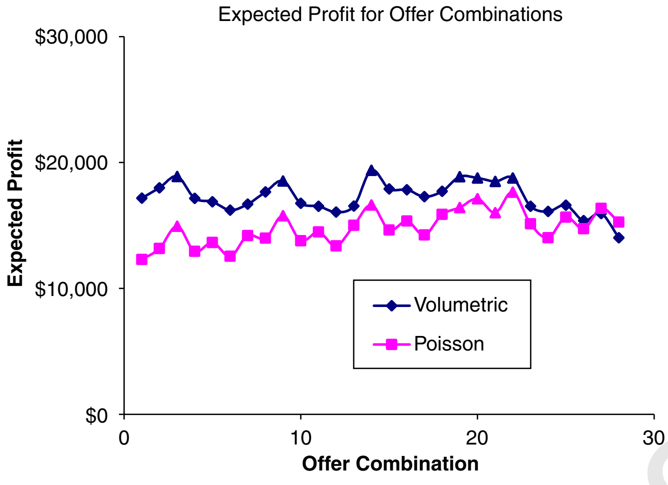

Fig. 1. Expected profit for offer combinations.

Offer Combination	1	2	**3**	4	5	6	7	8	**9**	10	11	12	13	**14**	15	16	17	18	**19**	**20**	**21**	**22**	23	24	25	26	27	28
Configuration A	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	4	4	4	4	5	5	5	6	6	7
Configuration B	2	3	4	5	6	7	8	3	4	5	6	7	8	4	5	6	7	8	5	6	7	8	6	7	8	7	8	8

Note: Points plotted with a triangle indicate the presence of configuration 4.

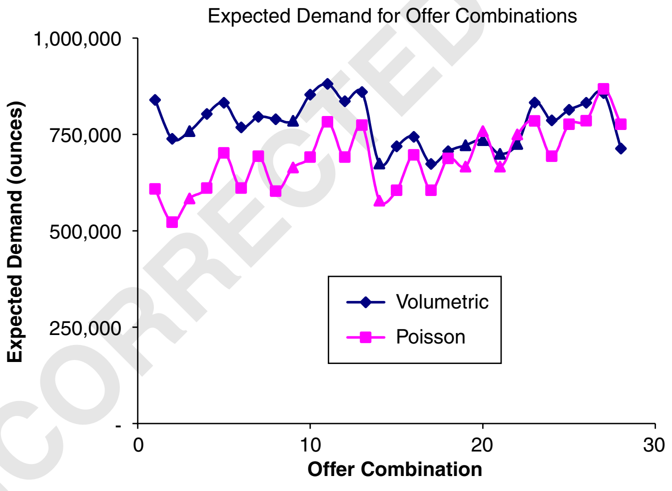

Fig. 2. Expected demand for offer combinations.

Offer Combination	1	2	**3**	4	5	6	7	8	**9**	10	11	12	13	**14**	15	16	17	18	**19**	**20**	**21**	**22**	23	24	25	26	27	28
Configuration A	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	4	4	4	4	5	5	5	6	6	7
Configuration B	2	3	4	5	6	7	8	3	4	5	6	7	8	4	5	6	7	8	5	6	7	8	6	7	8	7	8	8

Note: Points plotted with a triangle indicate the presence of configuration 4.

configurations that comprise the product line, and satiation for the volume consumed, the volumetric model predicts that smaller package sizes will result in greater demand. When these demand predictions are coupled with the contribution margins provided by the manufacturer, the volumetric model predicts the optimal offer combination to have smaller package volumes than the Poisson model.

6. Conclusions

This paper presents evidence that product attributes are associated with baseline marginal utility and satiation, and that the manner in which these associations are modeled matters. We find that the Lancastrian view of demand, where utility is related to the total level product characteristics in a given consumption bundle, is not supported in our analysis. The fit of the Lancaster model is inferior to that of models where characteristics are assumed to drive the parameters of the utility function directly. The log marginal density fit statistic is -5426 for the Lancaster model (Eq. (8)) versus -3800 for the offerings-based utility models (Eq. (1) and Table 3). This does not mean that projections on characteristics spaces are not a useful simplification of the model. We find that other models that project the baseline marginal utility parameter, ψ, onto the characteristics space as in Eq. (4) fit almost as well as the model with unique utility parameters for every product offering.

The results indicate that the Lancastrian model overly restricts the manner in which characteristics affect utility. If the space of attributes were small relative to the number of offerings, and there existed a sufficient proliferation of offerings so that consumers could find the attribute-combinations that matched their preferences, then the Lancaster model would perform well. However, as illustrated in Table 1, the combination of varieties and characteristics available for purchase will typically be sparsely populated unless the number of attributes important in determining preference is small. Unfortunately, this condition is unlikely to be true in nearly any product category. In the salty snack category, for example, producers work with thousands of variables (e.g., ingredients, temperature and other aspects of the manufacturing process) that give offerings their unique tastes, and employ highly aggregative characteristics such as “spice” or “citrus” to summarize and discuss aspects of a product’s formulation. It is doubtful, therefore, that the dimensions on which characteristics are defined can be accurately pre-specified by the data analyst.

Specifying baseline utility and satiation parameters as functions of attributes, as in Eqs. (4) and (7), flexibly projects these parameters onto the characteristics space. In our analysis of salty snacks, the dimension of the attribute space (i.e., 6) was nearly equal to the dimension of the product space (i.e., 8), and the estimated coefficients (β) resulted in a fairly minimal degradation in the fit of the demand model. This cost is offset by the benefit of understanding the influence of changes in characteristics to consumer utility and expected demand (e.g., Tables 5 and 6).

Our results also indicate that the ideal-point model (Eqs. (5) and (6)) are not supported in the data relative to simpler parameterizations. The ideal point model (Eq. (5)) requires twice the number of parameters as a model that linearly associates characteristics to baseline preference (Eq. (4)). While the concept of an ideal point is attractive conceptually, the data requirements are too severe. Our salty snack dataset comprises between six and seven observations per household. We believe that more than twice this number of observations would be needed to locate points of maximal preference at the respondent-level.

Introducing product attributes into models of demand is desirable because the results and inferences are actionable. Firms want to know the effects of product attributes for consumer welfare analysis and for predicting demand quantities. This paper examines the role of attributes within the context of an additive utility model for goods that are substitutes. A promising avenue for further research would extend our analysis to complementary products, which would involve adding interaction terms into the additive, but nonlinear, utility function.

7. Uncited reference

Hanemann, 1984.

Acknowledgements

The authors thank an anonymous referee for useful suggestions. Rossi thanks the James M. Kilts Center for Marketing at the Graduate School of Business, University of Chicago for generous support.

References

Berry, S., Pakes, A., 2002. The pure characteristics discrete choice model of differentiated products demand. Working paper, Yale University.
Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63 (4), 841–890.
Chan, T., 2003. Estimating a continuous hedonic choice model with an application to demand for soft drinks. Working paper, Washington University, St Louis.
Fader, P.S., Hardie, B.G.S., 1996. Modeling consumer choice among SKUs. Journal of Marketing Research 33, 442–452.
Gentzkow, M., 2005. Valuing new goods in a model with complementarity: online newspapers. Working paper, University of Chicago.
Hanemann, M., 1984. The discrete/continuous model of consumer demand. Econometrica 52, 541–561.
Kamakura, W.A., Srivastava, R.K., 1986. An ideal-point probabilistic choice model for heterogeneous preferences. Marketing Science 5 (3), 199–218.
Kim, J., Allenby, G.M., Rossi, P.E., 2002. Modeling consumer demand for variety. Marketing Science 21 (3), 229–250.
Lancaster, K., 1966. A new approach to consumer theory. Journal of Political Economy 74 (2), 132–157.
Newton, M.A., Raftery, A.E., 1994. Approximating Bayesian inference with the weighted likelihood boostrap. Journal of the Royal Statistical Society (B) 56, 3–48.
Rossi, P.E., Allenby, G.M., McCulloch, R., 2005. Bayesian Statistics and Marketing. Wiley, New York.