



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

## Modeling Indivisible Demand

Sanghak Lee, Greg M. Allenby

To cite this article:

Sanghak Lee, Greg M. Allenby (2014) Modeling Indivisible Demand. Marketing Science 33(3):364-381. <https://doi.org/10.1287/mksc.2013.0829>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2014, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Modeling Indivisible Demand

Sanghak Lee

Tippie College of Business, University of Iowa, Iowa City, Iowa 52242, [sanghak-lee@uiowa.edu](mailto:sanghak-lee@uiowa.edu)

Greg M. Allenby

Fisher College of Business, Ohio State University, Columbus, Ohio 43210, [allenby.1@osu.edu](mailto:allenby.1@osu.edu)

Disaggregate demand in the marketplace exists on a grid determined by the package sizes offered by manufacturers and retailers. Although consumers may want to purchase a continuous-valued amount of a product, realized purchases are constrained by available packages. This constraint might not be problematic for high-volume demand, but it is potentially troubling when demand is small. Despite the prevalence of packaging constraints on choice, economic models of choice have been slow to deal with their effects on parameter estimates and policy implications. In this paper we propose a general framework for dealing with indivisible demand in economic models of choice, and we show how to estimate model parameters using Bayesian methods. Analyses of simulated data and a scanner-panel data set of yogurt purchases indicate that ignoring packaging constraints can bias parameter estimates and measures of model fit, which results in the inaccurate measures of metrics such as price elasticity and compensating value. We also show that a portion of nonpurchase in the data (e.g., 2.27% for Yoplait Original) reflects the restriction of indivisibility, not the lack of preference. The importance of demand indivisibility is also highlighted by the counterfactual study where the removal of the smallest package size (i.e., 4 oz) mainly results in nonpurchase in the yogurt category instead of switching to larger package sizes.

**Keywords:** direct utility model; Bayesian error augmentation

**History:** Received: March 28, 2012; accepted: October 14, 2013; Preyas Desai served as the editor-in-chief and Puneet Manchanda served as associate editor for this article. Published online in *Articles in Advance* January 10, 2014.

## 1. Introduction

Marketplace demand is inherently indivisible. Demand for products in nearly every product category is constrained to lie on a grid defined by the array of available offerings and package sizes. A consumer who wants to purchase eight eggs cannot do so, nor can the consumer wanting to purchase one and a half cups of milk. The constraints imposed by packaging are also present in categories where consumers want less of an attribute, such as an automotive tire with a 30,000-mile tread life. The effects of indivisibility are most severe when demand is small, leading some consumers to not purchase in the category because the smallest available unit is too large. Demand for higher volume and aggregated purchases is less constrained as consumers can purchase across multiple package sizes to get close to their desired demand quantity.

Despite the prevalence of indivisible demand data, models of purchase quantity have generally not addressed its effect on parameter estimates, purchase incidence, and consumer welfare calculations. Exceptions include Small and Rosen (1981), who investigate the effect of simple discrete choice models on aggregate welfare calculations and theoretical work establishing the existence of equilibria in the presence of indivisible goods (Beviá et al. 1999,

Danilov et al. 2001). In marketing, Kalyanam and Putler (1997) propose an indivisible alternatives model where each package size is treated as a separate choice alternative, and only one package is purchased at a time. For perfect substitutes, it is also possible to engage in a simple search along each demand axis to identify the point of utility maximization (Allenby et al. 2004, Arora et al. 1998). However, when demand data exhibit multiple discreteness—i.e., when multiple variants of a product are simultaneously purchased—accommodating indivisibility is more difficult.

Economic models of demand built on direct and indirect utility specifications rely on first-order conditions to associate model parameters to the observed data. In a direct utility model, first-order conditions are expressed in terms of the Kuhn–Tucker conditions of constrained maximization (Chintagunta and Nair 2011). For indirect utility models, Roy's identity is used to derive demand expressions that serve as the basis for parameter estimation. The model of Hanemann (1984) is a special case where a discrete choice model is combined with a conditional demand model whose form comes from the indirect utility function. All three approaches rely on the assumption that the observed purchase quantities reflect the

point at which constrained utility is maximized. These models lead to a likelihood specified as a mixture of densities for the inside goods with positive demand and probability masses for the corner solutions where demand is zero (Kim et al. 2002, Song and Chintagunta 2007). The density contribution to the likelihood uses one realization of the error term, ignoring the possibility that other realizations may also lead to the same demand quantities because of indivisibility constraints.

In this paper, we propose a general method of dealing with the indivisibility of demand data. Consumer utility and its associated budget constraint are assumed continuous, and indivisibility enters our model as a constraint in the decision space. Observed demand does not indicate the exact point but rather the grid point at which feasible and constrained utility is maximized. Incorporating the constraint of indivisibility is challenging because the evaluation of the data likelihood requires integration of the joint density of error terms in regions dictated by the available offerings. We use the concavity property of utility functions to identify the appropriate region of integration for computing the likelihood, and we develop a variant of Bayesian data augmentation (Tanner and Wong 1987) to simplify parameter estimation.

We use a simulation study to demonstrate the importance of dealing with data indivisibility, and a scanner-panel data set is used to explore practical implications of incorrectly assuming a continuous demand space. The simulation study provides evidence of a systematic bias in parameter estimates. It also shows agreement between the continuity assumption and our model for indivisibility when the grid size is small. We then apply our model to a scanner-panel data set of yogurt purchases. Ignoring indivisibility results in a downward bias in parameter estimates, which in turn distorts policy implications. The continuous approach overestimates own-price elasticities as a result of the underestimation of satiation effect, whereas underestimation of compensating values is caused by the underestimation of baseline utility. The proposed discrete approach allows us to infer that a portion (e.g., 2.27% for Yoplait Original) of nonpurchase in our data stems from the restriction of indivisible demand, not from the lack of preference. We also find that the removal of the smallest package size in the category leads to nonpurchase of the product instead of switching to larger package sizes.

In the next section, we provide relevant marketing and economics literature to summarize how the issue of data indivisibility has been dealt with. Then, we develop our method for dealing with discrete package sizes, where consumers maximize their utility by choosing a grid point subject to a budgetary

allotment. In §4, we use a simulation study to illustrate the importance of dealing with demand indivisibility versus assuming that it is continuously available. Section 5 extends our method to a case where the subutility function of an outside good is nonlinear, resulting in an irregular region of integration. Our approach to dealing with indivisible demand is applied to scanner-panel data in §6, where comparison is made to standard analysis. Section 7 discusses implications of our model and analysis, and concluding remarks are offered in §8.

## 2. Literature Review

Demand indivisibility, which we also refer to as discreteness, is often addressed by employing a discrete statistical distribution whose domain is consistent with the observed data. Examples include the work of Anderson and Simester (2004), who assume that the number of units ordered by a customer from future catalogs follows a Poisson distribution, and Manchanda et al. (2004), who adopt a negative binomial distribution to account for the number of new prescriptions written by a physician. Alternatively, statistical models can be developed for discrete data by censoring realizations of a continuous model, as in cut-point models used in customer satisfaction research (Bradlow and Zaslavsky 1999, Gupta 1988, Rossi et al. 2001). Copulas have also been proposed as a way of describing discrete multivariate data (Danaher and Smith 2011). Although these models are useful for describing discrete demand and consumer responses, they are limited in their ability to relate these data to an underlying process of strategic behavior in which consumers are choosing from among an array of offerings (Chintagunta et al. 2006).

Structural models for consumer demand, in which consumers are thought to be goal directed, have not been as successful at dealing with indivisible demand. Consumers are assumed to be utility maximizers subject to budget and possibly other constraints (Satomura et al. 2011), and first-order conditions are used to associate observed demand with utility parameters (Hanemann 1984, Wales and Woodland 1983). There are two approaches for deriving this association. The first involves the direct utility function in which first-order (i.e., Kuhn–Tucker) conditions are used to associate observed demand with constrained utility maximization (Bhat 2005, 2008; Kim et al. 2002). The second approach derives the data likelihood using Roy's identity to associate observed demand with derivatives of the indirect utility function (Chiang 1991, Mehta 2007, Song and Chintagunta 2007). Both assume the existence of a continuous demand space. An exception is the work of Dubé (2004), who employs a univariate grid search procedure for predictions but does not incorporate indivisibility in parameter estimation.

Although demand indivisibility has not been explicitly incorporated in either direct or indirect utility models, there has been some literature that recognizes the presence of indivisibility and conduct a sensitivity analysis. Nair et al. (2005) check the robustness of their model by comparing the expected conditional quantity under discrete and continuous cases, concluding that the difference between the two cases is very small at an aggregate level. Kim et al. (2002) address the issue of integer constraints by changing the limits of integration in the evaluation of the likelihood for a corner solution, acknowledging the chance of systematic bias in the parameter estimates. These sensitivity analyses are insightful, yet the discreteness constraint is not structurally incorporated into the model specification and estimation.

Kuriyama and Hanemann (2006) proposes an integer programming approach in which the local search algorithm and the greedy method are used to handle the integral nature of recreation behavior data. Although their approach is similar to ours in that the optimality conditions are derived from comparisons with the adjacent grid points, there are two major differences. First, Kuriyama and Hanemann (2006) handle zeros as if they arise from the continuous decision space, whereas discreteness is taken care of for positive quantities. This inconsistency in the data-generating mechanism leads to the incomplete likelihood of data because the probabilities of all possible outcomes do not add up to one. Second, the approach used by Kuriyama and Hanemann (2006) relies on the comparison with the adjacent neighboring points, which is problematic when an objective function is not additively separable. We deal with the indivisibility that is present for both zero and positive quantities, and we propose an estimation method that extends the boundary of the local neighborhood to a degree appropriate for the objective function.

### 3. Model Development

#### 3.1. Direct Utility and Constraints

We assume that a consumer considers  $n$  goods in his or her purchase decision together with an outside good  $x_0$  and has a concave utility function for good  $i$  defined over a continuous nonnegative domain (i.e., marginal utility is continuous, positive, and decreasing in  $x$ ). For better exposition of our method, we first present our model with two simplifying assumptions: (i) We assume linear utility for the outside good, implying the value of the outside good does not satiate. (ii) We also assume that utility from the  $n$  goods is additively separable. These assumptions are commonly made in models of consumer choice and simplify the integration of the likelihood as explained below. We extend our method to a more general case

where additive separability is relaxed in §5. A simple additively separable model of utility is

$$U(x_{1t}, \dots, x_{nt}, x_{0t}) = \sum_{i=1}^n u_i(x_{it}) + \alpha_0 x_{0t}, \quad (1)$$

where  $x_{it}$  indicates the purchase quantity of good  $i$  at trip  $t$ . Consumers maximize utility subject to three sets of constraints: (i) nonnegativity constraints, (ii) budgetary constraints, and (iii) discreteness imposed by available package sizes. The first and most obvious constraint is nonnegativity of purchase quantity:

$$x_{it} \geq 0 \quad \forall i \in \{1, \dots, n, 0\}. \quad (2)$$

The second constraint comes from a budgetary allotment ( $M$ ), which implies the maximum amount of dollar spending for the inside goods in a given shopping trip. This constraint is typically of a linear form assuming that a unit price ( $p_{it}$ ) is constant over various purchase quantities ( $x_{it}$ ):

$$\sum_{i=1}^n p_{it} x_{it} + x_{0t} \leq M. \quad (3)$$

The third constraint reflects the discrete nature of offerings in a marketplace, implying that a consumer has to purchase an integer number of units:

$$x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, \dots, n\}. \quad (4)$$

The discreteness constraint is not applied to the outside good because its units are unspecified. The consumer's choice decision is therefore formulated as

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & U(x_{1t}, \dots, x_{nt}, x_{0t}) = \sum_{i=1}^n u_i(x_{it}) + \alpha_0 x_{0t} \\ \text{s.t.} \quad & \sum_{i=1}^n p_{it} x_{it} + x_{0t} \leq M, \\ & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, \dots, n\}, \\ & x_{0t} \geq 0. \end{aligned} \quad (5)$$

#### 3.2. A Discrete Likelihood for Indivisible Data

Although our methodology is general enough to be applied to any utility function, we begin with a simple form<sup>1</sup> for the purpose of exposition and consider the general case later (see §3.6):

$$u_i(x_{it}) = \frac{\alpha_i e^{\epsilon_{it}}}{\gamma_i} \log(\gamma_i s_i x_{it} + 1), \quad (6)$$

<sup>1</sup> Bhat (2008) indicates that the subutility specification in (6) corresponds to a special case of a generalized variant of the translated constant elasticity of substitution (CES) utility function ( $U(x) = \sum_{k=1}^K (\gamma_k / \alpha_k) \psi_k \{(x_k / \gamma_k + 1)^{\alpha_k} - 1\}$ ), when  $\alpha_k \rightarrow 0$ . He also notes that the flexibility of the utility function is not sacrificed much in this limiting case because "it is possible to closely approximate a subutility function profile based on a combination of  $\gamma_k$  and  $\alpha_k$  values with a subutility profile based solely on  $\gamma_k$  or  $\alpha_k$ " (p. 280).

where  $s_i$  indicates an observed unit volume for good  $i$  and  $\varepsilon_{it}$  is an idiosyncratic error term that is known to a decision maker but not revealed to a researcher;  $\alpha_i$  represents the baseline utility (i.e., the marginal utility at the point of zero purchase), and  $\gamma_i$  allows for flexibility in degree of satiation (i.e., a higher  $\gamma_i$  implies a greater rate of satiation for good  $i$ ) (see Bhat 2008).

A consumer's decision problem in (5) can be equivalently reexpressed by substituting the binding budget constraint<sup>2</sup> for  $x_{0t}$  in the utility specification as follows:

$$\begin{aligned} \underset{x_t}{\text{maximize}} \quad & U^*(x_{1t}, \dots, x_{nt}) \\ &= \sum_{i=1}^n \frac{\alpha_i e^{\varepsilon_{it}}}{\gamma_i} \log(\gamma_i s_i x_{it} + 1) + \alpha_0 \left( M - \sum_{i=1}^n p_{it} x_{it} \right) \\ \text{s.t.} \quad & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, \dots, n\}, \\ & M - \sum_{i=1}^n p_{it} x_{it} \geq 0. \end{aligned} \quad (7)$$

While  $U$  is an unconstrained utility function of inside goods and an outside good,  $U^*$  is a constrained utility function of inside goods only satisfying the budget constraint in (3).  $U^*$  is a concave function and its derivative with respect to  $x_{it}$  is given by

$$\frac{\partial U^*}{\partial x_{it}} = \frac{\alpha_i e^{\varepsilon_{it}} s_i}{\gamma_i s_i x_{it} + 1} - \alpha_0 p_{it}. \quad (8)$$

Note that the derivative of  $U^*$  with respect to  $x_{it}$  is a function of  $x_{it}$  only. This is due to the assumption of an additively separable utility function with linear subutility for the outside good, and implies that we can consider one good at a time to specify the sufficient conditions for an observed purchase quantity  $(x_{1t}^*, \dots, x_{nt}^*)$  to be optimal:

$$\begin{aligned} U^*(x_{it}^*, x_{-it}^*) &\geq \max\{U^*(x_{it}^* + \Delta_i, x_{-it}^*) | \\ &\quad (x_{it}^* + \Delta_i, x_{-it}^*) \in F\}_{\Delta_i \in [-1, 1]} \quad \forall i \in \{1, \dots, n\}, \end{aligned}$$

where

$$F \equiv \left\{ (x_{1t}, \dots, x_{nt}) \left| M - \sum_{i=1}^n p_{it} x_{it} \geq 0, x_{it} \in \{0, 1, \dots\} \right. \right\}. \quad (9)$$

Here,  $x_{-it}^*$  indicates a vector of purchase quantities of all the other inside goods. Equation (9) is true because utility functions are required to be concave in each element of the quantity vector  $x_t$ . The concavity property is also present when a linear budget constraint is substituted for the outside good to obtain  $U^*$  and in the presence of a restricted solution space because of

packaging constraints. Thus,  $U^*$  has a unique maximum value in the neighborhood of observed demand, taking its greatest realized value on the grid at the point of the observed data  $x_t^*$ . The concavity of the utility function allows us to express optimality conditions in terms of relationships to neighboring grid points instead of all feasible points. Substituting the utility expression in (6) into (9) yields bounds on error realizations that are consistent with observed demand  $(x_{1t}^*, \dots, x_{nt}^*)$  being utility maximizing subject to budget and packaging, or indivisibility, constraints:

$$\begin{aligned} lb_{it} &< \varepsilon_{it} < ub_{it} \quad \forall i \in \{1, \dots, n\}, \quad \text{where} \\ lb_{it} &= \log\left(\frac{\alpha_0 p_{it} \gamma_i}{\alpha_i}\right) - \log\left(\log\left(\frac{\gamma_i s_i x_{it}^* + 1}{\gamma_i s_i (x_{it}^* - 1) + 1}\right)\right), \\ ub_{it} &= \log\left(\frac{\alpha_0 p_{it} \gamma_i}{\alpha_i}\right) - \log\left(\log\left(\frac{\gamma_i s_i (x_{it}^* + 1) + 1}{\gamma_i s_i x_{it}^* + 1}\right)\right). \end{aligned} \quad (10)$$

In the case  $x_{it}^* = 0$ , we only need to compare  $U^*(x_{it}^*, x_{-it}^*)$  with  $U^*(x_{it}^* + 1, x_{-it}^*)$ , implying that there is no lower boundary (i.e.,  $lb_{it} = -\infty$ ). Similarly, if adding one more unit of good  $i$  violates the budget constraint (i.e.,  $p_{it}(x_{it}^* + 1) + \sum_{j=1, j \neq i}^n p_{jt} x_{jt}^* > M$ ),  $\varepsilon_{it}$  is bounded only from below (i.e.,  $ub_{it} = \infty$ ).

If we assume the errors are independent and follow an independent and identically distributed probability distribution  $f$ , the likelihood can be computed by integrating the joint density of all the errors over the region specified in (10):

$$L^{(d)}(x_{1t}^*, \dots, x_{nt}^*) = \prod_{i=1}^n \left\{ \int_{lb_{it}}^{ub_{it}} f(\varepsilon_{it}) d\varepsilon_{it} \right\}. \quad (11)$$

The presence of nonlinear utility for the outside good, or a utility function involving multiplicative factors reflecting complementary goods, gives rise to a well defined but more complicated set of restrictions. The intersection of the set of these restrictions forms the region of integration in (11).

### 3.3. Comparison to a Continuous Likelihood

A continuous likelihood is less flexible than a discrete likelihood because it assumes that interior points exactly satisfy the Kuhn–Tucker conditions. Assuming the outside good is positive, the likelihood for an observed purchase quantity vector  $(x_{1t}^*, \dots, x_{nt}^*)$  is given by

$$L^{(c)}(x_{1t}^*, \dots, x_{nt}^*) = \prod_{i=1}^n \{\ell(x_{it}^*)\}, \quad (12)$$

$$\ell(x_{it}^*) = \begin{cases} f(\nu_{it}) \times |J_{\varepsilon_{it} \rightarrow x_{it}}| & \text{if } x_{it}^* > 0, \\ \int_{-\infty}^{\nu_{it}} f(\varepsilon_{it}) d\varepsilon_{it} & \text{if } x_{it}^* = 0, \end{cases}$$

$$\text{where } \nu_{it} = \log\left(\frac{\alpha_0 p_{it} (\gamma_i s_i x_{it}^* + 1)}{\alpha_i s_i}\right),$$

$$J_{\varepsilon_{it} \rightarrow x_{it}} = \frac{\gamma_i s_i}{\gamma_i s_i x_{it}^* + 1}.$$

<sup>2</sup> At an optimal condition, the budget constraint in (3) is always binding because the marginal utility of an outside good is greater than 0.

There are two important differences in the likelihood (12) relative to the discrete likelihood in (10) and (11). First, the discrete likelihood produces a greater probability mass for a zero (i.e., a corner) than that of the continuous likelihood given the same set of parameters. When  $x_{it}^* = 0$ , the upper boundary for  $\varepsilon_{it}$  in the discrete likelihood is equal to  $\log(\alpha_0 p_{it}/\alpha_i) - \log(\log(\gamma_i s_i + 1)/\gamma_i)$ , which is greater than that for the continuous likelihood  $\log(\alpha_0 p_{it}/\alpha_i) - \log s_i$ . In other words, the continuous likelihood requires smaller baseline parameters (i.e.,  $\alpha_i$ ) to produce the same corner probabilities. Baseline parameters will therefore suffer from a downward bias when zeros in the data are generated from the discrete data-generating mechanism, but parameters are estimated with the continuous likelihood in (12). This is true for any concave utility function, and the proof is provided in Appendix A. Intuitively, the difference in the estimates of the baseline parameters reflects the different interpretations of nonpurchase (i.e., zero) from the two approaches. A model with a continuous decision space assumes that a consumer does not purchase a product because she does not like it. In contrast, when indivisibility restrictions are present, the interpretation of nonpurchase is that a consumer might not like it enough to buy one unit. This difference can be substantial depending on the unit's size.

The second difference is that model fit statistics are not strictly comparable across the two likelihoods because (12) associates one error realization with each interior point while the discrete likelihood in (10) and (11) associates many error realizations. An adjustment is therefore required to make the likelihoods from the two approaches comparable. We compute the discrete likelihood based on the parameter values that are estimated from the continuous approach. Because discreteness is actually present in the data, the continuous estimates should be viewed as approximations. We refer to fit statistics resulting from the evaluation of the discrete likelihood as the “true likelihood” below.

### 3.4. Identification

Random utility models rely on the theory of constrained utility maximization for rationalizing observed demand (e.g., purchase quantities). Error terms are included in the utility specification to explain variation in purchase quantities stemming from unobserved factors. The simple utility function in (6) allows us to shed light on the identification of the two parameters (i.e.,  $\alpha_i$  and  $\gamma_i$ ) through the discrete likelihood given in (10). First,  $\alpha_i$  is informed largely by the ratio of zeros in data. The lower bound is  $-\infty$  when  $x_{it}^* = 0$  while it is a finite value that depends on  $\alpha_i$  when  $x_{it}^* > 0$ . Higher ratio of zeros in data results in a smaller estimate of  $\alpha_i$  because a smaller  $\alpha_i$  increases the upper boundary in (10),

translating to the greater likelihood of zeros with the lower boundary being  $-\infty$ . Second, the variations in positive purchase quantities help identify  $\gamma_i$ . While the influence of  $\alpha_i$  is indifferent for different positive quantities in determining the upper and lower boundaries in (10), different purchase quantities result in different likelihoods when  $\gamma_i$  is positive (i.e., both  $\partial lb_{it}/\partial x_{it}$  and  $\partial ub_{it}/\partial x_{it}$  are dependent on  $\gamma_i$ , but not on  $\alpha_i$ ).

Indivisibility constraints imposed on decision variables have an influence on the estimation of utility parameters because the range of latent utilities (or equivalently the range of errors) is differently determined according to the presence of indivisibility constraints. In §3.4, we provide theoretical evidence of the downward bias in the estimates of baseline parameters when ignoring the indivisibility constraints. The magnitude of the bias depends on the difference between the upper boundaries of the error region. The upper boundary of the discrete likelihood is greater than that of the continuous likelihood, and the magnitude of the difference is given by  $\log(\gamma_i s_i / \log(\gamma_i s_i + 1))$ , which is a strictly increasing function with respect to  $\gamma_i s_i$ . Therefore, the bias caused by ignoring demand indivisibility is severe when  $\gamma_i s_i$  is large, whereas it is relatively weak with smaller  $\gamma_i s_i$ . This implies that handling demand indivisibility is more important when dealing with larger package sizes or when dealing with products of strong satiation, as manifested by infrequent observation of large purchase quantities.

### 3.5. The Discrete Choice Model

A discrete choice model, where only one choice alternative is allowed to be chosen, can be viewed as a special case of our method. Suppose that a respondent is given a choice set with  $n$  choice alternatives and asked to choose one that is most preferred. This task can be formally represented in the constrained utility maximization framework as follows:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & U(x_{1t}, \dots, x_{nt}, x_{0t}) = \sum_{i=1}^n u_i(x_{it}) + \alpha_0 x_{0t} \\ \text{s.t.} \quad & \sum_{i=1}^n p_{it} x_{it} + x_{0t} \leq M, \\ & x_{it} \in \{0, 1\} \quad \forall i \in \{1, \dots, n\}, \\ & x_{0t} \geq 0, \\ & \sum_{i=1}^n x_{it} = 1. \end{aligned} \tag{13}$$

Note that a quantity restriction is additionally introduced, reflecting that one alternative must be chosen with its quantity being one.<sup>3</sup> There exist only  $n$

<sup>3</sup> No choice option can be naturally accommodated in this framework by modifying the quantity constraint as  $\sum_{i=1}^n x_{it} \leq 1$ .

feasible points, and the sufficient conditions for optimality are derived by comparing the utilities at these feasible points. For example, the choice probability of  $i$ th alternative is given by

$$\begin{aligned}\Pr(x_{it}^* = 1) &= \Pr\left(i = \arg\max_{k \in \{1, \dots, n\}} (u_k(1) + \alpha_0(M - p_{kt}))\right) \\ &= \Pr(u_i(1) - \alpha_0 p_{it} \geq u_j(1) - \alpha_0 p_{jt}, \forall j \neq i) \\ &= \Pr(\alpha_i e^{\varepsilon_{it}} - \alpha_0 p_{it} \geq \alpha_j e^{\varepsilon_{jt}} - \alpha_0 p_{jt}, \forall j \neq i). \quad (14)\end{aligned}$$

Because the degree of satiation cannot be identified in a discrete choice model as a result of the absence of quantity variations, the subutility for the  $i$ th alternative (i.e.,  $u_i(1)$ ) can be represented by  $\alpha_i e^{\varepsilon_{it}}$  without loss of generality. The inequalities on the bottom equation specify the error region for computing the model likelihood.

### 3.6. General Case

A consumer's decision problem in (7) can be rewritten for any concave subutility function,  $u_i(x_{it} | \varepsilon_{it})$ , as follows:

$$\begin{aligned}\underset{x_i}{\text{maximize}} \quad & U^*(x_{1t}, \dots, x_{nt}) \\ &= \sum_{i=1}^n u_i(x_{it} | \varepsilon_{it}) + \alpha_0 \left( M - \sum_{i=1}^n p_{it} x_{it} \right) \\ \text{s.t.} \quad & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, \dots, n\}, \\ & M - \sum_{i=1}^n p_{it} x_{it} \geq 0.\end{aligned} \quad (15)$$

Equation (9) should be satisfied for an observed quantity to be optimal, which defines the region in an error space that rationalizes the goal-directed behavior of a consumer:

$$\begin{aligned}lb_{it} &< \varepsilon_{it} < ub_{it} \quad \forall i \in \{1, \dots, n\}, \quad \text{where} \\ lb_{it} \text{ satisfies } & u_i(x_{it}^* | lb_{it}) - u_i(x_{it}^* - 1 | lb_{it}) = \alpha_0 p_{it}, \\ ub_{it} \text{ satisfies } & u_i(x_{it}^* | ub_{it}) - u_i(x_{it}^* + 1 | ub_{it}) = -\alpha_0 p_{it}.\end{aligned} \quad (16)$$

The same rules are applied in the limiting cases. When  $x_{it}^* = 0$ ,  $lb_{it} = -\infty$ . If  $p_{it}(x_{it}^* + 1) + \sum_{j=1, j \neq i}^n p_{jt} x_{jt}^* > M$ ,  $ub_{it} = \infty$ . The model likelihood is computed by integrating the joint density of the errors over the specified region in (16).

## 4. Simulation Study

A simulation study is used to investigate the effect of employing a continuous likelihood when demand data are constrained to lie on a discrete grid. Data are generated by searching across the grid for the point that maximizes utility among a feasible set defined by a budgetary allotment and unit prices. In the

simulation study, we assume the following decision problem:

$$\begin{aligned}\underset{x_t}{\text{maximize}} \quad & U(x_{1t}, x_{2t}, x_{0t}) \\ &= \frac{1.0 e^{\varepsilon_{1t}}}{1.0} \log(1.0 s_1 x_{1t} + 1) \\ &\quad + \frac{0.5 e^{\varepsilon_{2t}}}{0.7} \log(0.7 s_2 x_{2t} + 1) + x_{0t} \quad (17) \\ \text{s.t.} \quad & p_{1t} x_{1t} + p_{2t} x_{2t} + x_{0t} \leq 50, \\ & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, 2\}, \\ & x_{0t} \geq 0,\end{aligned}$$

where the pack sizes are set to 6 ( $s_1 = s_2 = 6$ ), and  $\varepsilon_{it} \sim \text{Normal}(0, 1)$ ,  $p_{it} \sim \text{Uniform}(1, 3) \quad \forall i \in \{1, 2\}$ . In the simulation, we assume that the error term  $\varepsilon_{it}$  is known to the decision maker but not to the analyst. We allow for consumer heterogeneity by using a random-effect specification for 100 consumers; 100 observations are simulated for each respondent by searching over all the possible grid points under the constraint and finding the utility maximizer. Once simulated, the same data set is used to calibrate five different models, one in which the true package size (6) is assumed, one in which a continuous decision space is assumed, and the others in which smaller pack sizes (i.e., 3, 2, 1) are imposed. Table 1 summarizes the simulated data and the estimation results based on the five different models.

Panel (a) in Table 1 reports the average number of interior and corner solutions across the 100 observations, and panel (b) reports the estimated means of the random-effect distributions along with the true values. Parameter estimates are based on the mean of the posterior distribution using a diffuse but proper Normal prior with mean zero and variance 100. The "Model fit" column reports the log-likelihood of the data based on the different assumptions of the unit grid as well as the true log-likelihood of the observed data. Parameter estimates from each model are used to compute the joint probability of the observed data under the assumption of the true grid size (i.e.,  $s_1 = s_2 = 6$ ), which we name the "true log-likelihood."

The results of this simulation study show the following. First, parameter estimates are recovered to within the precision indicated by the posterior distribution only for the unit grid equal to six, which is the value used to generate the data. When the unit grid is assumed to be less than six, parameter estimates exhibit a downward bias that becomes increasingly severe as the unit grid diminishes. Second, parameter estimates for the continuous likelihood agree with the discrete estimates when the unit grid is small. This is expected because the continuous likelihood can be viewed as the mathematical limit of the discrete likelihood as the unit grid approaches

**Table 1** Simulation Study

(a) Simulated data							
		$x_1 > 0, x_2 > 0$	$x_1 > 0, x_2 = 0$	$x_1 = 0, x_2 > 0$	$x_1 = 0, x_2 = 0$	Total	
Avg. no. of observations		16.93	34.17	15.13	33.77	100	
(b) Estimation results							
Approach	Unit grid	Parameter estimates				Model fit	
		$\hat{\alpha}_1^* (= 0.000)$	$\hat{\alpha}_2^* (= -0.693)$	$\hat{\gamma}_1^* (= 0.000)$	$\hat{\gamma}_2^* (= -0.357)$	LL	True LL
Discrete	6	0.015 (0.055)	-0.685 (0.058)	-0.011 (0.068)	-0.328 (0.079)	-20,560	-20,560
	3	-0.797 (0.050)	-1.419 (0.049)	-1.311 (0.047)	-1.648 (0.049)	-29,234	-21,484
	2	-0.920 (0.049)	-1.512 (0.049)	-1.514 (0.042)	-1.835 (0.046)	-33,229	-21,852
	1	-1.022 (0.046)	-1.591 (0.048)	-1.687 (0.043)	-1.993 (0.043)	-39,559	-22,232
Continuous		-1.108 (0.046)	-1.661 (0.046)	-1.843 (0.042)	-2.131 (0.044)	-25,170	-22,630

Notes. Posterior standard deviations are given in parentheses.  $\alpha_i^* = \log(\alpha_i)$ ,  $\gamma_i^* = \log(\gamma_i)$ . LL, log-likelihood.

zero. Third, the true log-likelihood statistic also converges to that of the continuous model, and fourth, this measure of model fit correctly indicates that the true data-generating mechanism has a unit grid of 6.0. These results indicate that data discreteness is an important problem and that the proposed likelihood offers a viable solution.

## 5. Irregular Regions of Integration

The assumption of additive separability for the inside goods and linear utility for the outside good simplifies the evaluation of the likelihood in two ways: (i) it reduces the number of comparisons to  $2n$  (see Equation (9)), and (ii) each comparison provides a separate inequality for each error term. This results in a rectangular region of integration that is easy to calculate. However, this computational benefit comes with the cost of a potentially simplistic utility function and the budgetary allotment playing no role in the likelihood expression. That is, the optimal quantity for good  $i$  is not influenced by the price of good  $j$ , so long as the total expenditure does not exceed the budget (see §3.3 of Chandukala et al. 2007 for more discussion). In this section, we provide a solution for the more general case where the region of integration is irregular.

### 5.1. Conditions for Optimality

Suppose that utility for the outside good is a nonlinear function of  $x_{0t}$  (i.e.,  $u_0(x_{0t}) = \alpha_0 \log(x_{0t})$ ). The budget constraint can be used to substitute for the outside good, yielding a utility function written in terms of the inside goods:

$$U^*(x_{1t}, \dots, x_{nt}) = \sum_{i=1}^n \frac{\alpha_i e^{\varepsilon_{it}}}{\gamma_i} \log(\gamma_i s_i x_{it} + 1) + \alpha_0 \log\left(M - \sum_{i=1}^n p_{it} x_{it}\right). \quad (18)$$

$U^*$  is a concave function, and its derivative with respect to  $x_{it}$  is now a function of  $(x_{1t}, \dots, x_{nt})$ :

$$\frac{\partial U^*}{\partial x_{it}} = \frac{\alpha_i e^{\varepsilon_{it}} s_i}{\gamma_i s_i x_{it} + 1} - \frac{\alpha_0 p_{it}}{M - \sum_{i=1}^n p_{it} x_{it}}. \quad (19)$$

Because the derivative of  $U^*$  with respect to  $x_{it}$  depends on the other decision variables (i.e.,  $x_{-it}$ ), Equation (9) becomes necessary, though not sufficient, conditions for optimality. That is, satisfying Equation (9) does not guarantee that  $(x_{it}^*, x_{-it}^*)$  are the solution of the constrained maximization problem. Figure 1 provides an example that illustrates why the conditions in Equation (9) are not sufficient for optimality. Note that Equation (9) does not compare the optimal quantity (i.e.,  $(x_1^* = 3, x_2^* = 2)$ ) with the second best combination, which is  $(x_1 = 4, x_2 = 1)$  in this example.

Solving a nonlinear integer programming is generally known to be nondeterministic polynomial-time

**Figure 1** Local Comparison of Utility

		$x_1$					
		0	1	2	3	4	5
$x_2$	0	12.664	15.204	16.381	16.941	17.097	16.915
	1	14.164	16.636	17.726	18.172	<b>18.174</b>	17.767
	2	14.733	17.118	18.094	<b>18.386</b>	18.163	17.401
	3	14.861	17.133	17.955	18.022	17.444	16.034
	4	14.683	16.799	17.397	17.109	15.883	12.891

Notes.  $U^*(x_1, x_2) = 4.5 \log(x_1 + 1) + 3 \log(x_2 + 1) + 5.5 \log(10 - x_1 - x_2)$ . The numbers in cells indicate the utility value of the corresponding pair.



hard (NP-hard) (Wolsey 1998). This implies that the comparisons with all the feasible points are required to guarantee the optimality of the observed purchase quantity. The sufficient conditions for optimality can be formally written as follows:

$$U^*(x_{1t}^*, \dots, x_{nt}^*) \geq \max \left\{ U^*(x_{1t}, \dots, x_{nt}) \left| M - \sum_{i=1}^n p_{it} x_{it} \geq 0, \right. \right. \\ \left. \left. x_{it} \in \{0, 1, \dots\} \right\}. \quad (20)$$

However, it is often computationally too burdensome because the number of feasible points increases exponentially as the number of choice alternatives increases. Therefore, we propose an estimation method where only the adjacent feasible points are compared, instead of the full comparisons:

$$U^*(x_{1t}^*, \dots, x_{nt}^*) \\ \geq \max \{ U^*(x_{1t}^* + \Delta_1, \dots, x_{nt}^* + \Delta_n) \mid \\ (x_{1t}^* + \Delta_1, \dots, x_{nt}^* + \Delta_n) \in F \}_{\Delta_i \in \{-1, 0, 1\}}. \quad (21)$$

Although exploring the adjacent points is likely to be sufficient for optimality because of the concavity of the objective function in (18), there exist other cases where it does not guarantee the optimality (i.e., Equation (21) becomes the necessary rather than sufficient conditions). The concavity of the problem allows us to conduct the robustness check of the estimation results by increasing the scope of the adjacent points (e.g.,  $\Delta_i \in \{-2, -1, 0, 1, 2\}$ ).

The proposed local comparisons in (21) complicate the likelihood computation not only because the number of inequalities increases but also because each inequality contains multiple error terms that are nonlinearly associated. We develop a variant of Bayesian data augmentation (Tanner and Wong 1987), termed “error” augmentation (Zeithammer and Lenk 2006), to avoid the complex integration for computing the likelihood.

## 5.2. Estimation by Bayesian Error Augmentation

Bayesian error augmentation works by considering the error realization as the augmented variable instead of some variable such as  $y$  in a regression model. Probit models are frequently estimated with data augmentation by generating the latent values of utility (see Rossi et al. 2005). Given the utility values, model parameter estimation involves analysis of a standard linear regression model.

In our model, we treat the following variables as augmented:

$$z_{it} = \alpha_i e^{\varepsilon_{it}} \quad \forall i \in \{1, \dots, n\}, \quad \forall t \in \{1, \dots, T\}, \\ \text{or} \quad z_{it}^* = \alpha_i^* + \varepsilon_{it}, \quad (22)$$

where  $z_{it}^* = \log(z_{it})$  and  $\alpha_i^* = \log(\alpha_i)$ . We note that conditional on  $\alpha_i$ , the augmented variables  $z_{it}$  and  $\varepsilon_{it}$  contain the same information and are deterministically related. Given either  $\{z_{jt}\}_{j \neq i}$  or  $\{\varepsilon_{jt}\}_{j \neq i}$ , the inequalities in (21) reduce to a one-dimension region defined by the intersection of all the constraints containing  $z_{it}$  or  $\varepsilon_{it}$ . Each realization of  $\varepsilon_{it}$  either does or does not identify the observed demand vector  $x^*$  as the point of constrained utility maximization. The likelihood in error augmentation is therefore equal to one or zero, and draws of  $\varepsilon_{it}$  can be generated as a truncated draw from its prior (i.e., Normal) distribution. Values of  $z_{it}^*$  are obtained by adding  $\alpha_i^*$ . A simple example is provided in Appendix B to show the detailed procedure for error augmentation.

Then, conditioning on the values of  $\{z_{it}^*\}_{t=1, \dots, T}$  just drawn, a simple regression model can be used to generate draws of  $\alpha_i^*$ . The draw of  $\alpha_i^*$  is unconstrained because the truncation is induced by the optimality conditions in (21), and given that  $\{z_{it}^*\}_{t=1, \dots, T}$ , no further truncation is needed to estimate  $\alpha_i^*$ . The reason is the same as in a Bayesian estimation of a probit model using data augmentation: latent utilities are truncated by likelihood, but once they are obtained, a standard regression model can be used to obtain estimates of the model parameters. Details of the estimation procedure are provided in Appendix C. Two simulation studies are used to verify the procedure in Appendix D.

Although the example in Equations (18) and (19) demonstrates a particular case where the additive separability is relaxed by using a nonlinear subutility for an outside good, it does not limit the generality of our proposed method. Our method can be applied to any concave utility function where additive separability is relaxed. For example, a demand model for complementary goods requires a superadditive utility structure where the utility of joint purchase is greater than the total utilities of the separate purchases (see Lee et al. 2013 for more discussion), and an error augmentation is a viable solution for estimating the model.

## 6. Empirical Analysis

### 6.1. Data

We applied our methodology to the yogurt category in the IRI household panel data set described in Bronnenberg et al. (2008). The top 12 brands<sup>4</sup> are chosen based on their market shares; different flavors within the same brand are not distinguished in our analysis. Although each brand is offered in multiple pack sizes, we focus on the pack sizes whose market share

<sup>4</sup> Breyers, Breyers Creme Savers, Colombo Classic, Colombo Light, Dannon Fruit on the Bottom, Dannon Light & Fit, private label, Stonyfield Farm, Yoplait Light, Yoplait Original, Yoplait Thick and Creamy, and Yoplait Whips!.

**Table 2** Choice Alternatives

ID	Brand	Pack size (oz)									Total
		4	6	8	14	18	22.5	24	26.4	32	
B	Breyers			0.049							0.049
BCS	Breyers Creme Savers			0.039				<0.001			0.039
CC	Colombo Classic			0.070						<0.001	0.070
CL	Colombo Light			0.071							0.071
DFB	Dannon Fruit on the Bottom		0.051					<0.001			0.051
DLF	Dannon Light & Fit		0.102	0.004				<0.001		0.002	0.108
PL	Private label		0.007	0.090	0.001	0.001	0.001		0.001	0.005	0.106
SF	Stonyfield Farm		0.035	0.006						0.008	0.049
YL	Yoplait Light		0.033								0.033
YO	Yoplait Original		0.092					0.001			0.093
YTC	Yoplait Thick and Creamy		0.040					<0.001			0.040
YW	Yoplait Whips!	0.055									0.055
Total		0.055	0.360	0.329	0.001	0.001	0.001	0.001	0.001	0.015	0.774

*Notes.* Numbers indicate the market shares of the corresponding items. A cell is left blank when the corresponding package is not available.

is at least greater than 1% to deal with manageable number of choice alternatives in our analysis. Consequently, our data set contains 12 choice alternatives that account for 72.6% of the total purchases in the original data. Table 2 summarizes the pack size and market share of each alternative in our data set.

A total of 218 households are included in our data set such that each household shows the minimal amount of yogurt purchases during the two-year sample period (i.e., at least 10 purchase occasions) and relies heavily on the selected alternatives (i.e., at least 90% of the total yogurt purchases). The data contain the trips where multiple as well as single varieties are purchased and the trips where none of the inside goods is purchased.

Table 3 presents the descriptive statistics of our data. Households, on average, make 135.5 shopping trips during the two-year period, and 22.1% of the trips contain yogurt in the shopping basket. Conditional on the purchase occasion, an average of 5.5 units are purchased in a single shopping trip, which

comprises an average of 1.16 different brands purchased. Approximately 97.7% of the observed data are zeros, indicating great potential of obtaining biased estimates from assuming a continuous likelihood. In our analysis, we use 90% of the data for the model calibration, leaving 10% for prediction.

An outside good is introduced in the model specification to explain the variations in the total dollar spending for the inside goods. Here, the outside good is defined as the remaining dollars that are saved rather than spent on the inside goods. The subutility of the outside good is assumed to be linear, which implies that the value of saved money does not satiate. The linear subutility specification for the outside good does not allow us to infer the budgetary allotment, but it enables us to compute the exact likelihood, which decreases the computational burden of estimation. We also estimated the model with a nonlinear subutility for the outside good, and we find results consistent with those presented here.

**Table 3** Data Description

Characteristic	B	BCS	CC	CL	DFB	DLF	PL	SF	YL	YO	YTC	YW
Unit price (USD)	0.73	0.75	0.84	0.83	0.7	0.7	0.68	0.82	0.72	0.72	0.72	0.72
Unit volume (oz)	8	8	8	8	6	6	8	6	6	6	6	4
Price per volume	0.09	0.09	0.10	0.10	0.12	0.12	0.09	0.14	0.12	0.12	0.12	0.18
Purchase incidence	352	365	964	841	492	1,026	846	132	462	1,163	425	530
Purchase quantity	1,483	1,654	4,682	4,507	2,464	5,316	4,186	497	1,684	6,260	1,642	2,459
Available	28,980	26,981	28,866	22,105	28,160	28,271	22,263	26,883	27,479	28,666	28,615	28,336
Zero	28,628	26,616	27,902	21,264	27,668	27,245	21,417	26,751	27,017	27,503	28,190	27,806
One	29	20	98	63	30	49	32	13	53	44	45	59
Two	67	78	165	126	86	142	103	27	132	161	100	96
Three	51	38	110	92	49	104	159	36	69	103	62	43
Four and above	205	229	591	560	327	731	552	56	208	855	218	332

*Note.* The total number of shopping trips included in the data is 29,535.

**Table 4** Estimation Results

(a) Model-fit comparison												
	In-sample							Predictive				
		LL				True LL			LL			True LL
Discrete		−37,439				−37,439			−4,053			−4,053
Continuous		−37,637				−38,476			−4,160			−4,063
(b) Estimates by discrete approach												
	B	BCS	CC	CL	DFB	DLF	PL	SF	YL	YO	YTC	YW
Baseline ( $\bar{\alpha}_i^*$ )	−3.10 (0.10)	−3.56 (0.13)	−2.47 (0.10)	−2.35 (0.11)	−2.71 (0.10)	−2.39 (0.10)	−2.91 (0.13)	−3.00 (0.16)	−2.85 (0.12)	−2.30 (0.11)	−3.16 (0.16)	−2.37 (0.12)
Satiation ( $\bar{\gamma}_i^*$ )	−3.79 (0.21)	−3.24 (0.22)	−3.86 (0.15)	−3.92 (0.18)	−3.79 (0.13)	−3.86 (0.16)	−3.75 (0.16)	−3.40 (0.21)	−3.30 (0.23)	−3.78 (0.13)	−3.4 (0.30)	−3.05 (0.19)
Inventory ( $\bar{\lambda}^*$ )	−3.71	(0.24)										
(c) Estimates by continuous approach												
	B	BCS	CC	CL	DFB	DLF	PL	SF	YL	YO	YTC	YW
Baseline ( $\bar{\alpha}_i^*$ )	−3.23 (0.12)	−3.70 (0.14)	−2.55 (0.09)	−2.46 (0.10)	−2.8 (0.10)	−2.49 (0.10)	−2.97 (0.11)	−3.04 (0.11)	−3.03 (0.12)	−2.36 (0.10)	−3.40 (0.12)	−2.52 (0.13)
Satiation ( $\bar{\gamma}_i^*$ )	−4.40 (0.14)	−4.05 (0.22)	−4.27 (0.13)	−4.24 (0.15)	−4.03 (0.13)	−4.16 (0.14)	−4.28 (0.14)	−4.04 (0.23)	−3.89 (0.18)	−4.05 (0.10)	−3.91 (0.23)	−3.4 (0.20)
Inventory ( $\bar{\lambda}^*$ )	−3.49	(0.24)										

Note. In panels (b) and (c), posterior standard deviations are given in parentheses.

## 6.2. Inventory

Because zeros in data have different implications when discreteness is accounted for, it is important to pay careful attention to the potential sources that may result in the observation of zeros. We introduce an additional component to our model to capture the influence of inventory on purchase decision because the presence of inventory decreases the purchase utility as a result of the satiation effect, resulting in a nonpurchase decision. We use an exponential decay model for inventory depletion. In this model a fixed proportion of goods is consumed in a unit period (Ailawadi et al. 2007):

$$\eta_t^x = \lambda^{(t-1, t)} \left( \eta_{t-1}^x + \sum_{i=1}^n s_i^x x_{i, t-1} \right), \quad (23)$$

where  $(t-1, t)$  denotes the time interval between the two shopping trips, and  $\lambda$  indicates the rate of inventory for category  $X$ , which lies between 0 and 1. The impact of inventory on a consumer's decision is modeled such that it decreases the utility of an additional purchase as a result of the satiation effect:

$$u_i(x_{it}) = \frac{\alpha_i e^{s_{it}}}{\gamma_i} \log(\gamma_i s_i x_{it} + \eta_t^x + 1). \quad (24)$$

The rate of inventory (i.e.,  $\lambda$ ) is identified by the time-series observations of purchase quantities and inter-purchase time.

## 6.3. Estimates

The model parameters are estimated both by our discrete likelihood and by the standard continuous likelihood. Heterogeneity across households is incorporated by a random-effect specification for household parameters:

$$\theta_h = \{\alpha_{1h}^*, \dots, \alpha_{12h}^*, \gamma_{1h}^*, \dots, \gamma_{12h}^*, \lambda_h^*\} \sim \text{Normal}(\bar{\theta}, V_\theta), \quad (25)$$

where  $\alpha_{ih} = \exp(\alpha_{ih}^*)$ ,  $\gamma_{ih} = \exp(\gamma_{ih}^*)$ , and  $\lambda_h = 1/(1 + \exp(-\lambda_h^*))$ . We use a Bayesian Markov chain Monte Carlo (MCMC) method for estimation with a conjugate but relatively diffuse prior distribution for the hyper-parameters.<sup>5</sup> Sixty-thousand iterations of the chain were used to generate parameter estimates, with the first 30,000 draws discarded as burn-in. Table 4 displays the estimation results, comparing the two methods.

The discrete likelihood outperforms the continuous likelihood in both in-sample and predictive fit to the data. Although the parameter estimates from the two likelihoods show an ordinal consistency, the parameters are underestimated with the continuous likelihood due to the proliferation of zeros in the data. This illustrates that assuming continuously available demand can be costly when zeros datum values are common. Comparison between the two sets of parameter estimates show two patterns of biases that stem

<sup>5</sup>  $\bar{\theta} \mid V_\theta \sim N(0, V_\theta \otimes 100I)$ ,  $V_\theta \sim \text{IW}(10, 10I)$ .

from the ignorance of the data indivisibility. First, the baseline utility is underestimated in the continuous approach to account for the zeros in the data, some of which are due to indivisible demand (i.e., do not like it enough) and others due to lack of preference (i.e., do not like it, period). The slope of the subutility at zero quantity is steeper in the discrete approach, implying that an observation of nonpurchase in data represents lower level of preference in the continuous approach. This is consistent with our theoretical prediction discussed in §3.3. Second, the degree of satiation is underestimated when the discreteness of the data is not accounted for. This can be viewed as a by-product of the bias of the baseline utility. As the continuous approach underestimates the baseline parameter (i.e.,  $\alpha_i$ ) to rationalize zeros, a smaller  $\gamma_i$  is required to increase the log-likelihood for the positive quantities (see Equation (12)). In other words, because the continuous approach infers a lower level of preference from nonpurchase, it rationalizes the positive quantities by attributing to a lower degree of satiation.

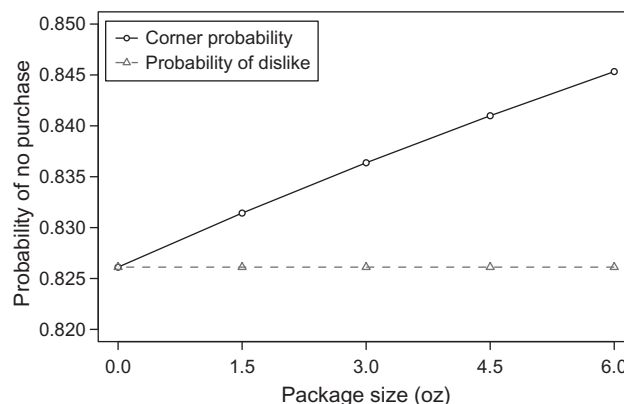
## 7. Discussion

### 7.1. Decomposition of Corner Probability

When a good is indivisible, a nonpurchase decision can arise from two different reasons: (i) a consumer does not like it, or (ii) a consumer likes it but not enough to buy one unit of it. The former simply reflects the lack of preference, whereas the latter can be adjusted if a smaller package is available. In this section, we explore how package sizes affect purchase decisions by investigating the probability of nonpurchase. Based on the full posterior distribution of individual-level parameters from the discrete model, we compute the probability of nonpurchase (i.e., corner probability) with different package sizes. These values can then be averaged across all the households to produce average corner probabilities. We assume that per-volume prices remain constant in this investigation.

Figure 2 displays the relationship between the nonpurchase probability and package size for a subset of households who prefer Yoplait Original yogurt. Sixty-three households were selected for analysis if they purchased Yoplait Original at least five times during the last two years.<sup>6</sup> When the actual package size of six ounces is assumed, the probability of no purchase is estimated to be 0.845, primarily because of the general tendency of households to not purchase yogurt on every shopping trip. This corner probability can be decomposed into two factors, i.e., one as a result of a lack of demand for Yoplait Original and the other a restriction on demand due to packaging. Although

**Figure 2** Impact of Package Size on Corner Probability for Yoplait Original



the majority of the nonpurchase probability (0.825) is driven by a lack of demand for the brand, there exists a nonnegligible probability (0.020) that is caused by packaging. This difference is represented by the gap between the two lines in Figure 2. As the container size decreases, the corner probability decreases and converges to the nonpurchase probability in a continuous decision space of small package sizes.

### 7.2. Price Elasticity and Compensating Value

Modeling data indivisibility is important because it yields consistent estimates of model parameters, which are the primitives for producing important metrics such as price elasticity and compensating value. In this section, we compare these metrics estimated from the discrete approach with those from the continuous approach and show how the different parameter estimates translate into the differences in price elasticity and compensating value.

Own-price elasticities are computed by comparing the expected demand under its regular price with one in which the price is reduced by 10%. Because the maximum dollar spending for the inside goods is \$20.00 in our data, we set the budgetary allotment to be \$25.00 to guarantee that the quantity of the outside good is always positive.<sup>7</sup> Panel (a) in Table 5 presents the own-price elasticities of each brand and compares the elasticities from the two approaches. The own-price elasticities from the two approaches show an ordinal consistency, indicating that the demand for the private label brand is the least elastic to price

<sup>6</sup> The households who seldom purchase the brand are not included in this analysis because their choice probabilities are not affected.

<sup>7</sup> The budgetary allotment cannot be estimated because it does not affect the model likelihood when the subutility of the outside good is linear. However, we need to set the budgetary allotment for our counterfactual studies. Because the choice of \$25.00 is arbitrary, we conducted a sensitivity analysis in which we vary the budgetary allotment to see how sensitive our counterfactual results are. We found that our main results, as for the differences between the discrete and continuous approach, are robust across various budget levels.

**Table 5** Price Elasticity and Compensating Value

Approach	B	BCS	CC	CL	DFB	DLF	PL	SF	YL	YO	YTC	YW
(a) Price elasticity												
Discrete	2.83 (0.26)	2.48 (0.18)	2.42 (0.12)	2.40 (0.12)	2.71 (0.19)	2.32 (0.10)	2.25 (0.10)	3.36 (0.49)	2.58 (0.19)	2.29 (0.10)	2.64 (0.21)	2.33 (0.15)
Continuous	2.97 (0.35)	2.59 (0.20)	2.58 (0.14)	2.47 (0.14)	2.81 (0.20)	2.40 (0.12)	2.35 (0.12)	3.54 (0.61)	2.72 (0.19)	2.37 (0.11)	2.76 (0.21)	2.48 (0.16)
(b) Compensating value												
Discrete	4.87 (0.65)	7.69 (1.19)	21.70 (1.85)	28.78 (2.54)	8.74 (0.95)	26.18 (2.54)	22.71 (2.07)	1.71 (0.43)	7.36 (1.08)	32.38 (2.50)	6.44 (0.84)	12.79 (1.43)
Continuous	4.52 (0.70)	6.92 (0.83)	18.38 (1.58)	23.81 (2.25)	7.69 (1.03)	22.92 (2.05)	19.78 (1.87)	1.55 (0.37)	6.77 (1.02)	27.48 (2.21)	5.81 (0.73)	10.82 (1.32)

Note. Posterior standard deviations are given in parentheses.

whereas that of the Stonyfield Farm is most price elastic. However, the own-price elasticities based on the continuous approach are always greater than those of the discrete approach because of biases in the parameter estimates of consumers' utility. In particular, the underestimation of the satiation effects in the continuous approach results in the overestimation of the own-price elasticities.

Compensating value measures the value of each brand by computing the amount of money a consumer would need to reach its initial utility after removing a brand from the product line. It reflects multiple facets of a consumer's utility (e.g., baseline utility, satiation) and provides useful information for assortment decisions. The compensating value of  $i$ th brand for household  $h$  is numerically computed by first evaluating the indirect utility under the full assortment condition, then equating it to the indirect utility under the new condition where brand  $i$  is removed and a budget is increased by  $CV_h^{(i)}$  for compensation:

$$V_h(p, M) = V_h^{(i)}(p, M + CV_h^{(i)}),$$

where

$$V_h(p, M) = \max_x U(x | \theta_h) \quad \text{s.t. } p'x = M; \quad (26)$$

$$V_h^{(i)}(p, M) = \max_x U(x | \theta_h) \quad \text{s.t. } p'x = M \text{ and } x_i = 0.$$

Panel (b) in Table 5 shows the per-trip compensating value of each brand for a household. On average, deletion of one brand is equivalent to the aggregate

compensation across the 218 households of \$32.58 per week based on the discrete approach, and Yoplait Original shows the greatest value, which is expected based on its large market share in the data. However, when these metrics are computed using a continuous likelihood, the compensating value for each brand is consistently underestimated.

In sum, the analyses of price elasticity and compensating value demonstrate the importance of modeling indivisible demand. Misrepresenting consumer behavior by ignoring packaging constraint in the decision space leads to biased estimates of preference and satiation, which in turn distorts managerial metrics such as price elasticities and compensating values.

### 7.3. Changes in Demand Upon the Removal of a Pack Size

Because our model of indivisible demand explicitly handles the discreteness constraints imposed by available pack sizes, it allows us to predict the changes in share-of-wallet when one of available pack sizes is removed. Share-of-wallet calculates the percentage of a consumer's budget that is allocated to each brand and is a more suitable measure than market share when dealing with multiple pack sizes whose prices are also different. Among the three available pack sizes (4, 6, and 8 oz) in our data, we choose the smallest pack size (4 oz) as the target of our counterfactual study because there is only one brand (i.e., Yoplait Whips!) offered in 4 oz. The impact of removing Yoplait Whips! on the share-of-wallet is summarized in Table 6.

**Table 6** Changes in Share-of-Wallet upon Removal of the 4 oz Item

Characteristic	B	BCS	CC	CL	DFB	DLF	PL	SF	YL	YO	YTC	YW	Outside good
Unit volume (oz)	8	8	8	8	6	6	8	6	6	6	6	4	n/a
YW is present	0.353	0.477	1.140	1.523	0.578	1.222	1.049	0.150	0.476	1.600	0.447	0.662	90.322
YW is removed	0.354	0.477	1.141	1.524	0.579	1.222	1.050	0.150	0.477	1.603	0.448	0.000	90.976
Changes in share-of-wallet	0.001	0.000	0.001	0.001	0.001	0.000	0.001	0.000	0.001	0.003	0.001	−0.662	0.654

Note. Numbers indicate the share-of-wallet (in percent) of the corresponding items.

When Yoplait Whips! is present, it accounts for about 0.66% of the total budget of consumers. However, when it is removed from the marketplace, the corresponding share-of-wallet mainly contributes to the outside good, increasing the wallet share for the outside good from 90.32% to 90.98%, whereas the influence is smaller for the other yogurt brands of larger pack sizes. Consumers who prefer the 4 oz Yoplait Whips! decide not to purchase any yogurt when their preferred pack size is not available.

#### 7.4. Magnitude of the Impact of Incorporating Indivisibility

The existence and direction of the bias in the estimates of continuous approach are proven theoretically in §3.4 and are demonstrated empirically in both our simulation study (in §4) and our empirical analysis (in §6). The magnitude of the bias, however, depends on characteristics of data. The effect is largest at the corners where nonpurchase of an item is falsely construed as someone not liking an offering enough to buy it as opposed to not liking it enough to buy the first unit, defined by the package size of the offering. We find in our analysis that strictly interior data, where none of the choice alternatives has zero demand, does not lead to much bias of parameters as a result of data indivisibility. Similarly, data that reside strictly at the corners (e.g., discrete choices) do not lead to biases because discrete multinomial data are not informative about rates of satiation, and these data only support models with linear indifference curves. Indivisibility effects are largest when the data comprise a mixture of corner and interior solutions.

The data analyzed in our empirical analysis above are dominated by no-choice trips to the store in which none of the brands was purchased. This resulted in a small amount of bias in the estimated model coefficients from the continuous likelihood when compared with the estimates from the indivisible likelihood. The Web appendix (available as supplemental material at <http://dx.doi.org/10.1287/mksc.2013.0829>) describes an application where the bias is more pronounced. Data for this analysis focus on just one brand (6 oz Yoplait yogurt) with different flavors treated as different choice alternatives. In this data set, 92% of the observed data are zeros compared with 98% in our empirical study above. We find that the greater presence of interior solutions leads to greater bias in estimated parameters and larger estimates of nonpurchase probability as a result of the effects of packaging (0.08 versus 0.02).

## 8. Concluding Remarks

This paper proposes a general approach to dealing with indivisible demand by using inequality restrictions in the model likelihood. Assuming additively

separable utility and linear subutility for the outside good simplifies the region of integration of the model likelihood. The region becomes irregular whenever the utility function is not additively separable or when utility for the outside good is specified nonlinearly. Bayesian error augmentation is shown to provide a viable solution for estimation when the likelihood has regular or irregular boundaries. Simulation studies are used to validate our algorithm and compare it with estimates obtained from a continuous likelihood that assumes that constrained utility is exactly maximized at the points of observed demand. The simulations show that baseline utilities as well as the degree of satiation are underestimated from the continuous likelihood, particularly in the presence of corner solutions.

We find support for our model using a scanner-panel data set of yogurt purchases. Explicitly recognizing the indivisibility of demand leads to better in-sample and predictive fits to the data. The same pattern of biases in the parameter estimates are observed in the continuous approach as is shown in the simulation study. We show that the biased parameter estimates in the continuous approach lead to the overestimation of own-price elasticities and the underestimation of compensating values. Our analysis also indicates that a portion of nonpurchase in the data (e.g., 2.27% for Yoplait Original) is due to the effects of packaging.

We believe that dealing with indivisible demand requires a model of direct utility maximization and likelihood-based estimation. Models based on indirect utility, where Roy's identity is used to associate first-order conditions with the observed data, assume that the indirect utility is continuous and differentiable. The presence of demand discreteness is inconsistent with the notion that attainable utility is continuous. As a result, it is not possible to use derivatives to obtain demand equations. Estimation based on the dual formulation of cost minimization is similarly unviable (Inoue 2009). Moment-based estimation techniques such as generalized method of moments (GMM) (Greene 2002, Chapter 18) also require modification in the presence of inequality constraints (Pakes et al. 2011). We leave the comparison of our method to GMM for future research. Finally, our approach can be applied to other utility functions, including models of complementarity and forward-looking behavior. Lee et al. (2013) show an application of our method in the context of complementary demand where the joint utility is superadditive. Models of forward-looking behavior lead to complex regions of integration in the likelihood for which Bayesian augmentation may be the only viable solution.

Our analysis of the yogurt data provides estimates of the effect of indivisibility on demand and compensating value estimates in one empirical setting. Additional replication is useful, particularly in categories where aggregation across flavors and varieties are not necessary to control for the number of offerings. Another application is in the development of models for volumetric conjoint analysis in which packaging constraints are explicitly explored. Finally, non-purchases account for almost 98% of our data, and it would be useful to more fully explore the effects of indivisibility in models of conditional demand.

### Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/mksc.2013.0829>.

### Appendix A. Probability Mass for Zero (Discrete vs. Continuous)

A consumer's utility function is given as follows (the time subscript is suppressed for simplicity):

$$U^*(x_1, \dots, x_n) = \sum_{i=1}^n u_i(x_i) + \alpha_0 \left( M - \sum_{i=1}^n p_i x_i \right).$$

When  $x_i^* = 0$  in the observation (i.e.,  $(x_i^* = 0, x_{-i}^*)$ ), the probability mass for  $x_i^* = 0$  is differentially computed in the following two methods:

#### 1. Discrete approach:

$$\begin{aligned} U^*(x_i^* = 0, x_{-i}^*) &\geq U^*(x_i^* = 1, x_{-i}^*) \\ \Leftrightarrow u_i(x_i^* = 0) + \sum_{j=1, j \neq i}^n u_j(x_j^*) + \alpha_0 \left( M - \sum_{j=1, j \neq i}^n p_j x_j^* \right) \\ &\geq u_i(x_i^* = 1) + \sum_{j=1, j \neq i}^n u_j(x_j^*) + \alpha_0 \left( M - p_i - \sum_{j=1, j \neq i}^n p_j x_j^* \right) \\ \Leftrightarrow u_i(x_i^* = 0) &\geq u_i(x_i^* = 1) - \alpha_0 p_i. \end{aligned}$$

Therefore, the corresponding region in error space is

$$D_{\varepsilon_i} = \{\varepsilon_i \mid u_i(x_i^* = 1, \varepsilon_i) - u_i(x_i^* = 0, \varepsilon_i) \leq \alpha_0 p_i\}.$$

#### 2. Continuous approach:

$$\left. \frac{\partial U^*}{\partial x_i} \right|_{x_i=x_i^*=0} = \left. \frac{\partial u_i(x_i)}{\partial x_i} \right|_{x_i=x_i^*=0} - \alpha_0 p_i \leq 0.$$

Therefore, the corresponding region in error space is

$$C_{\varepsilon_i} = \left\{ \varepsilon_i \mid \left. \frac{\partial u_i(x_i, \varepsilon_i)}{\partial x_i} \right|_{x_i=x_i^*=0} \leq \alpha_0 p_i \right\}.$$

Since  $u_i(x_i)$  is concave,  $(u_i(x_i^* = 1) - u_i(x_i^* = 0))/(1 - 0) < (\partial u_i(x_i)/\partial x_i)|_{x_i=x_i^*=0}$  by definition. Therefore,  $\varepsilon_i \in C_{\varepsilon_i} \rightarrow \varepsilon_i \in D_{\varepsilon_i}$  (i.e.,  $C_{\varepsilon_i} \subset D_{\varepsilon_i}$ ). For any concave utility function, the discrete likelihood produces a greater probability mass for a zero than that of the continuous likelihood given the same set of parameters.

### Appendix B. An Example for Error Augmentation

Suppose we observe  $(x_1^* = 1, x_2^* = 0)$  as the solution for the following constrained utility maximization problem:

$$\begin{aligned} \underset{x_t}{\text{maximize}} \quad & U^*(x_{1t}, x_{2t}) = \frac{\alpha_1 e^{\varepsilon_{1t}}}{\gamma_1} \log(\gamma_1 s_1 x_{1t} + 1) + \frac{\alpha_2 e^{\varepsilon_{2t}}}{\gamma_2} \\ & \cdot \log(\gamma_2 s_2 x_{2t} + 1) + \log(M - p_{1t} x_{1t} - p_{2t} x_{2t}) \\ \text{s.t.} \quad & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, 2\}, \\ & M - p_{1t} x_{1t} - p_{2t} x_{2t} \geq 0. \end{aligned}$$

Inequality conditions in (21) should be satisfied for the observed demand  $(x_1^* = 1, x_2^* = 0)$  to be the optimal solution, which translates to the following  $9 (= 3^2)$  comparisons:

- (c1)  $U^*(1, 0) \geq U^*(1 - 1, 0 - 1) = U^*(0, -1)$ ,
- (c2)  $U^*(1, 0) \geq U^*(1, 0 - 1) = U^*(1, -1)$ ,
- (c3)  $U^*(1, 0) \geq U^*(1 + 1, 0 - 1) = U^*(2, -1)$ ,
- (c4)  $U^*(1, 0) \geq U^*(1 - 1, 0) = U^*(0, 0)$ ,
- (c5)  $U^*(1, 0) \geq U^*(1, 0)$ ,
- (c6)  $U^*(1, 0) \geq U^*(1 + 1, 0) = U^*(2, 0)$ ,
- (c7)  $U^*(1, 0) \geq U^*(1 - 1, 0 + 1) = U^*(0, 1)$ ,
- (c8)  $U^*(1, 0) \geq U^*(1, 0 + 1) = U^*(1, 1)$ ,
- (c9)  $U^*(1, 0) \geq U^*(1 + 1, 0 + 1) = U^*(2, 1)$ .

Because the decision variables cannot take negative values, (c1), (c2), (c3) are not valid comparisons, and thus they can be disregarded. And (c5) is always satisfied. Therefore, we have five conditions to be satisfied (i.e., (c4), (c6), (c7), (c8), and (c9)). Substitution of the utility function defined above in these conditions produce the following inequalities:

- (c4)  $U^*(1, 0) \geq U^*(0, 0)$   
 $\Leftrightarrow \frac{\alpha_1 e^{\varepsilon_{1t}}}{\gamma_1} \log \frac{\gamma_1 s_1 + 1}{1} + \log \frac{M - p_{1t}}{M} \geq 0,$
- (c6)  $U^*(1, 0) \geq U^*(2, 0)$   
 $\Leftrightarrow \frac{\alpha_1 e^{\varepsilon_{1t}}}{\gamma_1} \log \frac{\gamma_1 s_1 + 1}{2\gamma_1 s_1 + 1} + \log \frac{M - p_{1t}}{M - 2p_{1t}} \geq 0,$
- (c7)  $U^*(1, 0) \geq U^*(0, 1)$   
 $\Leftrightarrow \frac{\alpha_1 e^{\varepsilon_{1t}}}{\gamma_1} \log \frac{\gamma_1 s_1 + 1}{1} + \frac{\alpha_2 e^{\varepsilon_{2t}}}{\gamma_2} \log \frac{1}{\gamma_2 s_2 + 1}$   
 $+ \log \frac{M - p_{1t}}{M - p_{2t}} \geq 0,$
- (c8)  $U^*(1, 0) \geq U^*(1, 1)$   
 $\Leftrightarrow \frac{\alpha_2 e^{\varepsilon_{2t}}}{\gamma_2} \log \frac{1}{\gamma_2 s_2 + 1} + \log \frac{M - p_{1t}}{M - p_{1t} - p_{2t}} \geq 0,$
- (c9)  $U^*(1, 0) \geq U^*(2, 1)$   
 $\Leftrightarrow \frac{\alpha_1 e^{\varepsilon_{1t}}}{\gamma_1} \log \frac{\gamma_1 s_1 + 1}{2\gamma_1 s_1 + 1} + \frac{\alpha_2 e^{\varepsilon_{2t}}}{\gamma_2} \log \frac{1}{\gamma_2 s_2 + 1}$   
 $+ \log \frac{M - p_{1t}}{M - 2p_{1t} - p_{2t}} \geq 0.$

The intersection of these inequalities defines the region in the error space  $(\varepsilon_{1t}, \varepsilon_{2t})$ , where the observed demand is the optimal solution for the constrained maximization problem. However, because it is difficult to integrate the joint distribution of the errors over the identified region, we propose a Bayesian error augmentation method as a viable solution.

Suppose we know the realized value of  $\varepsilon_{2t}$ , which satisfies (c8). Then, conditional on  $\varepsilon_{2t}$ , the inequalities (c4), (c6), (c7), and (c9) become unidimensional with respect to  $\alpha_1 e^{\varepsilon_{1t}}$ :

$$(c4) \quad U^*(1, 0) \geq U^*(0, 0)$$

$$\Leftrightarrow \alpha_1 e^{\varepsilon_{1t}} \geq \frac{-\gamma_1 \log((M - p_{1t})/M)}{\log((\gamma_1 s_1 + 1)/1)} \equiv lb_1,$$

$$(c6) \quad U^*(1, 0) \geq U^*(2, 0)$$

$$\Leftrightarrow \alpha_1 e^{\varepsilon_{1t}} \leq \frac{\gamma_1 \log((M - p_{1t})/(M - 2p_{1t}))}{-\log((\gamma_1 s_1 + 1)/(2\gamma_1 s_1 + 1))} \equiv ub_1,$$

$$(c7) \quad U^*(1, 0) \geq U^*(0, 1)$$

$$\Leftrightarrow \alpha_1 e^{\varepsilon_{1t}} \geq (\gamma_1 (-\log((M - p_{1t})/(M - p_{2t})) - ((\alpha_2/e^{\varepsilon_{2t}})/\gamma_2) \cdot \log(1/(\gamma_2 s_2 + 1))) \cdot (\log((\gamma_1 s_1 + 1)/1))^{-1} \equiv lb_2,$$

$$(c9) \quad U^*(1, 0) \geq U^*(2, 1)$$

$$\Leftrightarrow \alpha_1 e^{\varepsilon_{1t}} \leq (\gamma_1 (\log((M - p_{1t})/(M - 2p_{1t} - p_{2t})) + ((\alpha_2/e^{\varepsilon_{2t}})/\gamma_2) \log(1/(\gamma_2 s_2 + 1))) \cdot (-\log((\gamma_1 s_1 + 1)/(2\gamma_1 s_1 + 1))) \equiv ub_2.$$

The intersection of the four inequalities is summarized as

$$\max(lb_1, lb_2) \leq z_{1t} \equiv \alpha_1 e^{\varepsilon_{1t}} \leq \min(ub_1, ub_2),$$

which is equivalent to

$$\begin{aligned} \log(\max(lb_1, lb_2)) &\leq \log(z_{1t}) \\ &\equiv \log(\alpha_1) + \varepsilon_{1t} \leq \log(\min(ub_1, ub_2)). \end{aligned}$$

If the prior distribution of  $\varepsilon_{1t}$  is a Normal distribution of mean zero,  $\log(z_{1t})$  can be drawn from a truncated Normal distribution with its mean being  $\log(\alpha_i)$ , the lower boundary being  $\log(\max(lb_1, lb_2))$ , and its upper boundary being  $\log(\min(ub_1, ub_2))$ .

### Appendix C. Estimation Procedure (with Nonlinear Subutility for an Outside Good)

$$\begin{aligned} U^*(x_{1t}, \dots, x_{nt}) &= \sum_{i=1}^n \frac{z_{it}}{\gamma_i} \log(\gamma_i s_i x_{it} + 1) \\ &\quad + \alpha_0 \log \left( M - \sum_{i=1}^n p_{it} x_{it} \right), \end{aligned}$$

where  $z_{it} = \alpha_i e^{\varepsilon_{it}}$  (i.e.,  $z_{it}^* = \alpha_i^* + \varepsilon_{it}$ ),  $\varepsilon_{it} \sim \text{Normal}(0, 1)$ .

We want to infer about all the unknowns,  $\alpha_1^*, \dots, \alpha_n^*, \gamma_1^*, \dots, \gamma_n^*, M^*$ , and we introduce  $\{z_{1t}^*\}_{t=1, \dots, T}, \dots, \{z_{nt}^*\}_{t=1, \dots, T}$  as augmented variables:  $\alpha_i^*, \gamma_i^*, M^*$ , and  $z_{it}^*$  denote  $\log(\alpha_i)$ ,  $\log(\gamma_i)$ ,  $\log(M)$ , and  $\log(z_{it})$ , respectively.

The joint posterior distribution of the parameters and augmented variables is proportional to the likelihood times the prior:

$$\begin{aligned} &p(\alpha_1^*, \dots, \alpha_n^*, \gamma_1^*, \dots, \gamma_n^*, \{z_{1t}^*\}_{t=1, \dots, T}, \dots, \{z_{nt}^*\}_{t=1, \dots, T} \mid \\ &\quad \{x_{1t}\}_{t=1, \dots, T}, \dots, \{x_{nt}\}_{t=1, \dots, T}) \\ &\propto \prod_{t=1}^T \ell(x_{1t}, \dots, x_{nt} \mid z_{1t}^*, \dots, z_{nt}^*, \gamma_1^*, \dots, \gamma_n^*, M^*) \\ &\quad \times p(\{z_{1t}^*\}_{t=1, \dots, T} \mid \alpha_1^*) \times \dots \times p(\{z_{nt}^*\}_{t=1, \dots, T} \mid \alpha_n^*) \\ &\quad \times p(\alpha_1^*) \times \dots \times p(\alpha_n^*) \times p(\gamma_1^*) \times \dots \times p(\gamma_n^*) \times p(M^*). \end{aligned}$$

Given  $\{z_{1t}^*, \dots, z_{nt}^*, \gamma_1^*, \dots, \gamma_n^*, M^*\}$ , the likelihood of  $(x_{1t}, \dots, x_{nt})$  is an indicator function whose value equals 1 when all the conditions in (21) are satisfied and 0 otherwise. In addition, the  $(\alpha_1^*, \dots, \alpha_n^*)$  are conditionally independent of the data given  $(z_{1t}^*, \dots, z_{nt}^*)$ . Therefore, the full conditional distributions can be expressed as

$$\begin{aligned} &p(z_{it}^* \mid \text{else}) \\ &\quad \propto \ell(x_{1t}, \dots, x_{nt} \mid z_{1t}^*, \dots, z_{nt}^*, \gamma_1^*, \dots, \gamma_n^*, M^*) p(z_{it}^* \mid \alpha_i^*), \\ &p(\alpha_i^* \mid \text{else}) \propto p(z_{1t}^*, \dots, z_{nt}^* \mid \alpha_i^*) p(\alpha_i^*), \\ &p(\gamma_i^* \mid \text{else}) \\ &\quad \propto \prod_{t=1}^T \ell(x_{1t}, \dots, x_{nt} \mid z_{1t}^*, \dots, z_{nt}^*, \gamma_1^*, \dots, \gamma_n^*, M^*) p(\gamma_i^*), \\ &p(M^* \mid \text{else}) \propto \prod_{t=1}^T \ell(x_{1t}, \dots, x_{nt} \mid z_{1t}^*, \dots, z_{nt}^*, \gamma_1^*, \dots, \gamma_n^*, M^*) \\ &\quad \times p(M^*) \quad \forall i \in \{1, \dots, n\}, \quad \forall t \in \{1, \dots, T\}. \end{aligned}$$

The conditional distribution of  $z_{it}^*$  is a truncated Normal distribution whose upper and lower truncation points are determined by the sufficient conditions in (21) given  $\{z_{jt}^*\}_{j \neq i}$ . Given  $(z_{1t}^*, \dots, z_{nt}^*)$ ,  $\alpha_i^*$  can be generated from a simple Bayesian regression model. However, we cannot draw directly from the posterior distributions for  $\gamma_1^*, \dots, \gamma_n^*, M^*$ . Therefore, we use Metropolis-Hastings algorithm for drawing  $\gamma_1^*, \dots, \gamma_n^*, M^*$ . Consequently, we can simulate all the unknowns from the joint posterior distribution by using a Gibbs-type sampler and following a sequence of steps described below.

Step 1. Set initial values for  $\alpha_1^*, \dots, \alpha_n^*, \gamma_1^*, \dots, \gamma_n^*, M^*$ .

Step 2. Set initial values for  $\{z_{1t}^*\}_{t=1, \dots, T}, \dots, \{z_{nt}^*\}_{t=1, \dots, T}$  (by solving the Kuhn-Tucker conditions with given  $\alpha_1^*, \dots, \alpha_n^*, \gamma_1^*, \dots, \gamma_n^*, M^*$ ).

Step 3. Draw  $\{z_{it}^*\}_{t=1, \dots, T}$ :

$$z_{it}^* \mid \{z_{jt}^*\}_{j \neq i}, \alpha_i^* \sim N(\alpha_i^*, 1) \times 1(lb_{it} \leq z_{it}^* \leq ub_{it})$$

( $lb_{it}, ub_{it}$  denote the lower and upper truncation points, respectively, for  $z_{it}^*$ ).

Step 4. Draw  $\alpha_i^*$ :

$$\begin{aligned} &\alpha_i^* \mid \{z_{it}^*\}_{t=1, \dots, T}, \alpha_{i0}, \sigma_{i0} \\ &\quad \sim N \left( \left( \frac{\alpha_{i0}}{\sigma_{i0}^2} + \sum_{t=1}^T z_{it}^* \right) / \left( \frac{1}{\sigma_{i0}^2} + T \right), \left( \frac{1}{\sigma_{i0}^2} + T \right)^{-1} \right) \end{aligned}$$

(the prior distribution for  $\alpha_i^*$  is  $N(\alpha_{i0}, \sigma_{i0}^2)$ ).



Step 5. Draw  $\gamma_i^*$  using a Metropolis-Hastings algorithm with a random-walk chain.

Step 6. Draw  $M^*$  using a Metropolis-Hastings algorithm with random-walk chain.

Step 7. Repeat Steps 3–6 for all  $i \in \{1, \dots, n\}$ .

Step 8. Repeat Steps 3–7 at each iteration of the MCMC (the draws at the previous iteration become the new initial values).

## Appendix D. Simulation Studies with Nonlinear Subutility for an Outside Good

We present two simulation studies where the subutility of an outside good is nonlinearly specified. Whereas the first study clearly demonstrates the validity of our method based on a small scale simulation study (e.g., two inside goods), the second study is provided to show that the proposed method of error augmentation can be used in a more realistic setting where the scale of our empirical data is employed.

### D.1. Simulation Study with Two Inside Goods

The data are simulated from the following constrained utility maximization setting:

$$\begin{aligned} \underset{x_t}{\text{maximize}} \quad & U(x_{1t}, x_{2t}, x_{0t}) = 0.3e^{\varepsilon_{1t}} \log(s_1 x_{1t} + 1) \\ & + 0.4e^{\varepsilon_{2t}} \log(s_2 x_{2t} + 1) + 10 \log(x_{0t}) \\ \text{s.t.} \quad & p_{1t}x_{1t} + p_{2t}x_{2t} + x_{0t} \leq 50, \\ & x_{it} \in \{0, 1, 2, \dots\} \quad \forall i \in \{1, 2\}, \\ & x_{0t} \geq 0, \end{aligned} \quad (\text{D1})$$

where the pack sizes are set to 6 ( $s_1 = s_2 = 6$ ), and  $\varepsilon_{it} \sim \text{Normal}(0, 1)$ ,  $p_{it} \sim \text{Uniform}(1, 3)$ . The data contain 200 respondents and 100 observations for each respondent, and both corner and interior solutions are present (see Table D.1). Nonlinear subutility for an outside good allows us to infer about the budgetary allotment because it affects

the model likelihood (see Satomura et al. 2011). Therefore, we estimate  $M$  as well in this simulation study. Given the simulated data, we estimate the model by the proposed method using a Bayesian error augmentation and compare the results with those from the continuous approach.

Whereas the proposed estimation method recovers the true values within 95% credible intervals, the continuous approach suffers in estimating the parameters. We observe the convergence of results to that of the continuous approach as the unit grid gets smaller in the discrete approach. We also estimated the model using an approximate method that relies on a narrower set of local comparisons described in Equation (9), in which only the neighborhood points with their Euclidean distance from an optimal point being less than or equal to 1 are compared. The result confirms that it is important to extend the boundary of neighborhood points to be compared. When the boundary of local comparison is extended in our proposed method (i.e.,  $\Delta_i \in \{-2, -1, 0, 1, 2\}$ ), the estimation results remains the same, indicating that the proposed inequalities in (21) are close to the sufficient conditions. Therefore, the simulation study confirms that we can estimate models of discrete demand in which nonlinear subutility is assumed for an outside good by using the Bayesian error augmentation technique.

### D.2. Simulation Study with 12 Inside Goods

An additional simulation study is conducted to demonstrate that the proposed method of error augmentation works well in handling large-scale problems. We employ the scale of our empirical data, where 218 households make 136 shopping trips on average and make purchase decisions approximately 12 inside goods. The prices and pack sizes are also used from the empirical data such that three different pack sizes are available and offered at different prices. Table D.2 summarizes the setting for data simulation and compares

**Table D.1** Simulation Study 1 with Nonlinear Subutility for an Outside Good

(a) Simulated data						
		$x_1 > 0, x_2 > 0$	$x_1 > 0, x_2 = 0$	$x_1 = 0, x_2 > 0$	$x_1 = 0, x_2 = 0$	Total
Avg. number of obs.		43.98	17.68	27.64	10.70	100
(b) Estimation results						
Approach	Unit grid	Parameter estimates				
		$\bar{\alpha}_1^* (= -1.204)$	$\bar{\alpha}_2^* (= -0.916)$	$\bar{\gamma}_1^* (= 0.000)$	$\bar{\gamma}_2^* (= 0.000)$	$\bar{M}^* (= 3.912)$
Discrete	6	-1.196 (0.032)	-0.887 (0.033)	0.003 (0.021)	0.002 (0.021)	3.923 (0.029)
	3	-1.501 (0.034)	-1.212 (0.032)	-1.088 (0.023)	-0.983 (0.022)	3.631 (0.031)
	2	-1.622 (0.033)	-1.327 (0.033)	-1.465 (0.025)	-1.319 (0.025)	3.500 (0.033)
	1	-1.693 (0.034)	-1.381 (0.034)	-1.753 (0.024)	-1.695 (0.025)	3.415 (0.032)
Continuous	na	-1.724 (0.037)	-1.405 (0.035)	-1.876 (0.030)	-1.745 (0.031)	3.368 (0.037)
(c) Different boundary of local comparison						
Neighborhood		Parameter estimates				
Narrower set (Equation (9))		-1.056 (0.058)	-0.754 (0.054)	-0.069 (0.048)	-0.076 (0.048)	3.761 (0.015)
Extended set ( $\Delta_i \in \{-2, -1, 0, 1, 2\}$ in (21))		-1.199 (0.033)	-0.884 (0.033)	0.001 (0.020)	0.008 (0.020)	3.926 (0.028)

Notes. Posterior standard deviations are given in parentheses.  $\alpha_i^* = \log(\alpha_i)$ ,  $\gamma_i^* = \log(\gamma_i)$ , and  $M^* = \log(M)$ .

Table D.2 Simulation Study 2 with Nonlinear Subutility for an Outside Good

	Brand											
	1	2	3	4	5	6	7	8	9	10	11	12
(a) Simulation setting												
Unit price	0.73	0.75	0.84	0.83	0.7	0.7	0.68	0.82	0.72	0.72	0.72	0.72
Unit volume	8	8	8	8	6	6	8	6	6	6	6	4
True $\bar{\alpha}_i^*$	−3.100	−3.560	−2.470	−2.350	−2.710	−2.390	−2.910	−3.000	−2.850	−2.300	−3.160	−2.370
True $\bar{\gamma}_i^*$	−0.790	−0.240	−0.860	−0.920	−0.790	−0.860	−0.750	−0.400	−0.300	−0.780	−0.400	−0.050
(b) Parameter estimates by discrete approach												
$\bar{\alpha}_i^*$	−3.112 (0.055)	−3.481 (0.064)	−2.473 (0.055)	−2.404 (0.059)	−2.774 (0.055)	−2.358 (0.053)	−2.997 (0.053)	−2.998 (0.059)	−2.914 (0.060)	−2.228 (0.053)	−3.187 (0.056)	−2.404 (0.063)
$\bar{\gamma}_i^*$	−0.739 (0.057)	−0.222 (0.054)	−0.890 (0.056)	−1.011 (0.054)	−0.761 (0.057)	−0.840 (0.058)	−0.715 (0.054)	−0.417 (0.050)	−0.357 (0.058)	−0.845 (0.054)	−0.345 (0.055)	0.022 (0.050)
(c) Parameter estimates by continuous approach												
$\bar{\alpha}_i^*$	−4.048 (0.058)	−4.665 (0.065)	−3.325 (0.061)	−3.194 (0.062)	−3.556 (0.059)	−3.108 (0.058)	−3.943 (0.057)	−3.945 (0.062)	−3.893 (0.063)	−2.969 (0.058)	−4.157 (0.063)	−3.358 (0.064)
$\bar{\gamma}_i^*$	−2.803 (0.049)	−2.987 (0.091)	−2.623 (0.040)	−2.607 (0.038)	−2.517 (0.043)	−2.380 (0.039)	−2.757 (0.045)	−2.586 (0.063)	−2.545 (0.059)	−2.329 (0.037)	−2.640 (0.074)	−2.102 (0.054)

Notes. Posterior standard deviations are given in parentheses.  $\alpha_i^* = \log(a_i)$ ,  $\gamma_i^* = \log(\gamma_i)$ .

the estimation results based on the discrete likelihood with those based on the continuous likelihood.

The estimation results based on the discrete likelihood show that all the true parameters are well recovered within a 95% credible set, which confirms that the proposed method of error augmentation is a viable solution for handling data indivisibility in large-scale problems. The same bias pattern is replicated in the estimation results of the continuous approach such that both the baseline and the satiation parameters are underestimated when the indivisibility is not taken into consideration.

## References

- Ailawadi KL, Gedenk K, Lutzky C, Neslin SA (2007) Decomposition of the sales impact of promotion-induced stockpiling. *J. Marketing Res.* 44(3):450–467.
- Allenby GM, Shively TS, Yang S, Garratt MJ (2004) A choice model for packaged goods: Dealing with discrete quantities and quantity discounts. *Marketing Sci.* 23(1):95–108.
- Anderson ET, Simester DI (2004) Long-run effects of promotion depth on new versus established customers: Three field studies. *Marketing Sci.* 23(1):4–20.
- Arora N, Allenby GM, Ginter JL (1998) A hierarchical Bayes model of primary and secondary demand. *Marketing Sci.* 17(1):29–44.
- Beviá C, Quinzii M, Silva JA (1999) Buying several indivisible goods. *Math. Social Sci.* 37(1):1–23.
- Bhat CR (2005) A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions. *Transportation Res.* 39(8):679–707.
- Bhat CR (2008) The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Res.* 42(3):274–303.
- Bradlow ET, Zaslavsky AM (1999) A hierarchical latent variable model for ordinal data from a customer satisfaction survey with “no answer” responses. *J. Amer. Statist. Assoc.* 94(445):43–52.
- Bronnenberg BJ, Kruger MW, Mela CF (2008) The IRI marketing data set. *Marketing Sci.* 27(4):745–748.
- Chandukala S, Kim J, Otter T, Rossi P, Allenby G (2007) Choice models in marketing: Economic assumptions, challenges and trends. *Foundations Trends Marketing* 2(2):97–184.
- Chiang J (1991) A simultaneous approach to the whether, what, and how much to buy questions. *Marketing Sci.* 10(4):297–315.
- Chintagunta PK, Nair HS (2011) Discrete-choice models of consumer demand in marketing. *Marketing Sci.* 30(6):977–996.
- Chintagunta P, Erdem T, Rossi PE, Wedel M (2006) Structural modeling in marketing: Review and assessment. *Marketing Sci.* 25(6):604–616.
- Danaher PJ, Smith MS (2011) Modeling multivariate distributions using copulas: Application in marketing. *Marketing Sci.* 30(1):4–21.
- Danilov V, Koshevoy G, Murota K (2001) Discrete convexity and equilibria in economies with indivisible goods and money. *Math. Soc. Sci.* 41(3):251–273.
- Dubé J-P (2004) Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Sci.* 23(1):66–81.
- Greene WH (2002) *Econometric Analysis*, 5th ed. (Prentice Hall, Upper Saddle River, NJ).
- Gupta S (1988) Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* 25(4):342–355.
- Hanemann WM (1984) Discrete/continuous models of consumer demands. *Econometrica* 52(3):541–561.
- Inoue T (2009) Indivisible commodities and an equivalence theorem on the strong core. IMW Working Paper 417, Universität Bielefeld, Bielefeld, Germany.
- Kalyanam K, Putler DS (1997) Incorporating demographic variables in brand choice models: An indivisible alternatives framework. *Marketing Sci.* 16(2):166–181.
- Kim J, Allenby GM, Rossi PE (2002) Modeling consumer demand for variety. *Marketing Sci.* 21(3):229–250.
- Kuriyama K, Hanemann WM (2006) The integer programming approach to a generalized corner-solution model: An application to recreation demand. Working paper, Waseda University, Tokyo.
- Lee S, Kim J, Allenby GM (2013) A direct utility model for asymmetric complements. *Marketing Sci.* 32(3):454–470.
- Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(4):467–478.

- Mehta N (2007) Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis. *Marketing Sci.* 26(2):196–217.
- Nair H, Dubé J-P, Chintagunta P (2005) Accounting for primary and secondary demand effects with aggregate data. *Marketing Sci.* 24(3):444–460.
- Pakes A, Porter J, Ho K, Ishii J (2011) Moment inequalities and their application. Working paper, Harvard University, Cambridge, MA.
- Rossi PE, Allenby GM, McCulloch R (2005) *Bayesian Statistics and Marketing* (John Wiley & Sons, Chichester, UK).
- Rossi PE, Gilula Z, Allenby GM (2001) Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statist. Assoc.* 96(453):20–31.
- Satomura T, Kim J, Allenby GM (2011) Multiple-constraint choice models with corner and interior solutions. *Marketing Sci.* 30(3):481–490.
- Small KA, Rosen HS (1981) Applied welfare economics with discrete choice models. *Econometrica* 49(1):105–130.
- Song I, Chintagunta PK (2007) A discrete-continuous model for multicategory purchase behavior of households. *J. Marketing Res.* 44(4):595–612.
- Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82(398):528–540.
- Wales TJ, Woodland AD (1983) Estimation of consumer demand systems with binding non-negativity constraints. *J. Econometrics* 21(3):263–285.
- Wolsey LA (1998) *Integer Programming*, Wiley Series in Discrete Mathematics and Optimization, Vol. 52 (John Wiley & Sons, New York).
- Zeithammer R, Lenk P (2006) Bayesian estimation of multivariate-normal models when dimensions are absent. *Quant. Marketing Econom.* 4(3):241–265.