SARANG SUNDER, V. KUMAR, and YI ZHAO*

In this study, the authors propose a flexible framework to assess customer lifetime value (CLV) in the consumer packaged goods (CPG) context. They address the substantive and modeling challenges that arise in this setting, namely, (1) multiple discreteness, (2) brand switching, and (3) budget-constrained consumption. Using a Bayesian estimation, the authors are also able to infer the consumer's latent budgetary constraint using only transaction information, thus enabling managers to understand the customer's budgetary constraint without having to survey or depend on aggregate measures of budget constraints. Using the proposed framework, CPG manufacturers can assess CLV at the focal brand level as well as at the category level, a departure from CLV literature, which has mostly been firm centric. The authors implement the proposed model on panel data in the carbonated beverages category and showcase the benefits of the proposed model over simpler heuristics and conventional CLV approaches. Finally, they conduct two policy simulations describing the role of the budget constraint on CLV, as well as the asymmetric effects of pricing in this setting, and develop managerial insights in this context.

Keywords: customer relationship management, Bayesian estimation, consumer packaged goods, multiple discreteness, customer lifetime value

Online Supplement: http://dx.doi.org/10.1509/jmr.14.0641

Measuring the Lifetime Value of a Customer in the Consumer Packaged Goods Industry

The customer centricity paradigm has long been documented as one of the most important tenets of effective marketing in today's dynamic environment. With the advent of technology and customer relationship management (CRM), there is an explosion of disaggregate and granular customer data (transactional as well as survey) available to firms. In the past decade, customer lifetime value (CLV) has emerged as an effective metric for CRM. Strategies for CRM that are developed from CLV modeling have led to positive financial gains in business-to-business (B2B) as well as business-to-consumer (B2C) settings. Because the CLV metric is heavily dependent on customer relationships and transaction data, it has mostly been implemented in relationship-marketing settings. However, the concepts of CLV and customer-centric marketing are applicable in traditionally product-centric industries such as consumer packaged goods (CPG) as well. In fact, the implementation of CLV in the CPG setting is one of the explicitly stated objectives of the Marketing Accountability Standards Board.[1]

*Sarang Sunder is Assistant Professor of Marketing, Neeley School of Business, Texas Christian University (e-mail: s.sunder@tcu.edu). V. Kumar (corresponding author) is Regents Professor, Richard and Susan Lenny Distinguished Chair, Professor of Marketing, and Executive Director of the Center for Excellence in Brand and Customer Management, J. Mack Robinson College of Business, Georgia State University (e-mail: vk@gsu.edu). Yi Zhao is Assistant Professor of Marketing, J. Mack Robinson College of Business, Georgia State University (e-mail: yizhao@gsu.edu). V. Kumar has also been honored as the Chang Jiang Scholar, Huazhong University of Science and Technology; TIAS Fellow, Texas A&M University; and ISB Senior Fellow, Indian School of Business. The authors are grateful to the review team for helpful suggestions during the revision process. They thank the Marketing Accountability Standards Board for providing access to the data, and they give special thanks to the executives of a major carbonated soda manufacturer for reviewing the results and offering valuable insights. They also thank the participants of the 2014 Winter Marketing Educators Conference and the ISMS 36th Marketing Science Conference and the audience in more than ten universities worldwide for providing valuable suggestions on the study. The authors are grateful to Renu for copyediting an earlier version of the manuscript. This article is based on the first author's doctoral dissertation. Donald Lehmann served as associate editor for this article.

[1]http://www.themasb.org/projects/underway/.

ISSN: 0022-2437 (print)
1547-7193 (electronic)

Journal of Marketing Research
Vol. LIII (December 2016), 901–921
DOI: 10.1509/jmr.14.0641

However, traditional marketing (especially in the CPG context) has focused on reaching out to consumers through mass marketing and delivering standardized products/services. While this has worked in the past, it may no longer be sustainable in a dynamic and digitally connected marketing environment. Although traditional aggregate metrics (market share, sales volume, revenue, etc.), which are commonly used in the CPG context to assess brand performance, convey important information about the product/brand and can be readily calculated, they do not provide us with the complete picture. While these metrics give managers an indication of the health of the brand and serve as an aggregate proxy for performance, they do not provide any information regarding which customers grew and which ones did not.

Further, flow-based metrics (market share, brand sales, etc.) are very sensitive to extraneous shocks (Yoo, Hanssens, and Kim 2011) and ignore the heterogeneity present among households. The CLV metric presents stability according to consumer purchase behavior, which is long-term focused and forward-looking in nature. Many CPG firms are investing heavily in innovations in CRM that would move them closer to a CLV-based approach to decision making. While there have been several case studies and white papers that hint at the need for customer centricity in the CPG industry (e.g., Kimberly-Clark's Huggies brand[2]), to our knowledge, there has not been an academic study that provides a robust methodology for assessing CLV in the CPG industry. Through this research, we hope to provide the first step in applying customer valuation and customer-centric marketing in the CPG industry.

In order to assess CLV in the CPG industry, we need to build a model that accurately captures consumers' purchase behavior in this setting. The implementation of a CLV-based marketing paradigm in CPG firms faces several challenges, including (1) the problem of multiple discreteness (wherein consumers purchase more than one brand in the same occasion), (2) heavy brand switching, and (3) the budget-constrained nature of CPG purchases. First, CPG consumers[3] do not always purchase a single brand in a given month. Due to the relatively lower (relative to relationship-driven CLV contexts) costs of switching in the CPG industry, variety-seeking consumers tend to try various brands within the same shopping period, thus leading to multiple discreteness in CPG consumption, which has been documented in the literature (Allender et al. 2013; Dubé 2004; Richards, Gómez, and Pofahl 2012). This multibrand purchase in the same month leads to violations of typical discrete choice models, which are commonly used in conventional CLV models. This presents the first challenge, wherein in order to accurately capture the consumption patterns, the CLV model needs to account for multiple discreteness.

Second, given the low cost of switching, we need to explicitly account for brand-switching and competing-brand effects in the CPG context. Previous research has highlighted the importance of accounting for brand switching in CPG markets, especially in situations of low product differentiation (Van Oest 2005). However, conventional CLV models that rely on internal company data often ignore the role of competition and brand switching. Extant CLV models that do account for brand switching rely heavily on survey data that describe either customers' actual switching (Rust, Lemon, and Zeithaml 2004) or share-of-wallet (SOW) information (Kumar and Shah 2009). The collection of survey data, while viable in a business setting in which relationships are clearly defined, becomes very challenging in the CPG context due to scale and cost issues associated with appending panel data with survey information.

Third, past studies have suggested that households keep track of budgets, especially in the CPG setting (Antonides, Manon de Groot, and Van Raaij 2011; Heath and Soll 1996; Stilley, Inman, and Wakefield 2010), and try to maintain spending within a target maximum level to maintain control over consumption or spending (Gilboa, Postlewaite, and Schmeidler 2010). Mental budgets act as a form of self-control to ensure that consumers stay within their spending limits. That is, consumers have unobserved limits on the number of dollars they are willing to allocate toward a particular category, which includes the product category as well as outside substitutes. The budget constraint would then encompass all the dollars allocated toward this overall spending category.[4] Much of the past research in the area of mental budgeting has relied on some form of survey data (Du and Kamakura 2008; Stilley, Inman, and Wakefield 2010). Because collecting and appending survey data in the CPG setting are difficult, it becomes necessary to infer this information using readily available transaction data. This issue is further underscored when addressing CLV in the CPG setting because managers need to know not only what the CLV of each customer is but also the customer's maximum budget allocation within the category. Knowledge of the limits of a customer's spend (budget constraints) helps managers avoid overspending on customers who have a low ceiling and underspending on customers who have a high ceiling. In the implementation described in this study, we assume that the budget constraint acts at the product category level. This assumption is based on prior literature on constrained utility models in microeconomics as well as in marketing. In fact, several researchers have applied similar frameworks in various business settings, such as carbonated soft drinks (Dubé 2004), yogurt (Kim, Allenby, and Rossi 2002), fresh produce (Richards, Gómez, and Pofahl 2012), salty snacks (Kim, Allenby, and Rossi 2007), and ice cream (Allender et al. 2013). Furthermore, we note that the budget constraint captures the "cost" of consumption in the utility model and enables us to directly address the issue of multiple discreteness. The budget constraint could be related to monetary, space, or consumption restrictions. In this study, we are interested in

[2]http://www.nielsen.com/content/dam/corporate/us/en/public%20factsheets/Case%20Studies/CaseStudy-KimberlyClark-ROI.pdf.

[3]In this study, we use "consumer," "customer," and "household" interchangeably. Our model is implemented at the household level, but we note that the model is flexible to be estimated at the consumer level if the data are available.

[4]While it is possible that the consumer sets overall budget limits at the trip level, it is expected that these limits "trickle" down to the category level (which includes the focal product category and outside substitutes). Incorporating trip-level (as opposed to category-level) budget constraints is beyond the scope of this study because it would require multicategory purchase data and would complicate the model further.

monetary constraints at the household level, but we note that the constraint could be extended to space/inventory or consumption maxima (Satomura, Kim, and Allenby 2011). Given the aforementioned challenges, our main research objectives are as follows:

- Getting a long-term, customer-centric view of the CPG customer: How can we model the consumer's CLV in a CPG setting?
- Explicitly account for multiple discreteness and heavy brand switching: How can we leverage scanner panel data in the CPG industry to explicitly consider brand switching and account for the issue of multiple discreteness when modeling CLV?
- Understanding the budgetary constraint: How can we infer the customer's budget constraint at the individual level?
- Policy simulations in CLV modeling: How can firms use a structural approach to assess CLV in the CPG setting and eventually conduct relevant counterfactuals without having to conduct expensive studies in the field?

We implement a structural model of multiple discrete purchases on scanner panel transaction data spanning three years. We demonstrate the predictive power of our approach relative to conventional CLV modeling approaches and also highlight its advantages over simpler heuristics (usage, market share, etc.). Specifically, we show that the proposed model outperforms conventional models (e.g., multivariate probit-then-quantity approach) in terms of predictive power and customer base classification. Furthermore, we also compare our approach of customer valuation with simpler heuristics commonly used in the CPG industry. We show that the proposed model outperforms simpler heuristics in terms of classifying customers into high, medium, and low segments, thus underscoring the need for a sophisticated model of consumer behavior to assess CLV in the CPG setting. Overall, we demonstrate the advantages that the proposed framework presents over conventional approaches (both heuristic-based and model-based) in measuring CLV in the CPG setting. Specifically, we show that the proposed model outperforms simpler heuristics/models in predicting future purchases as well as in CLV segmentation. In addition, we compute individual CLV and segment the customers into high, medium, and low CLV segments. At the segment level, we provide insights into each CPG brand's share of CLV and discuss the implications for each brand. The use of a structural model in this setting allows us to conduct theoretically grounded policy simulations that can generate valuable insights to managers in the CPG context. To this effect, we conduct two policy simulations that are managerially relevant. First, we simulate the effect of changes to the budget constraint on CLV. We find that, on average, a reduction in the budget constraint leads to a greater effect on CLV than a gain in budget. We show that this effect is heterogeneous, that is, the magnitude of the effect is different depending on the CLV segment. Second, we study the own effects and cross-effects of price on quantity consumed. We find that the effects are nonsymmetric for increases and decreases in price, indicating nonlinear price elasticities. Furthermore, as we highlight, this effect is also heterogeneous across CLV segments. These policy simulations have major managerial implications: they indicate that (1) budget constraints significantly and nonlinearly affect the CLV of a CPG customer, and (2) price elasticity is nonlinear and asymmetric across various CLV segments.

The remainder of this article is organized in the following manner. In the next section, we discuss the related marketing literature in the areas of CLV and multiple discreteness modeling and outline our contributions. Next, we provide a brief description of the data used in the empirical application and present evidence of multiple discreteness in the data. Then, we develop the structural model of multiple discreteness, discuss the operationalization of the budget parameter, and derive the likelihood. Within this section, we also elaborate on the Bayesian estimation procedure used to recover the parameters. Next, we elaborate on the findings from the study and compare our model with conventional CLV models. In the subsequent section, we compute the CLV and conduct managerially relevant counterfactuals or policy simulations that could aid CPG manufacturers in understanding CLV in the CPG setting. Finally, we highlight the key academic and managerial implications of the proposed approach and conclude with limitations and future research directions.

LITERATURE GAP

CLV Modeling

Competition in CLV modeling. Consumers make choices relative to competing brands/firms/offerings in the marketplace, so it is important to evaluate the importance of competition in CLV modeling, especially in the CPG context. Researchers have tried to mitigate this issue by including survey-based measures of the customers' SOW to control for competitive effects. However, this approach has two shortfalls. First, it is difficult for the researcher to collect survey data for the entire customer base and maintain the database for the entire transaction history of the customer. Second, the SOW metric does not explicitly incorporate competition into the choice framework of the customer because it is used more as a control variable. Rust, Lemon, and Zeithaml (2004) use a Markov switching matrix to account for the customer's brand-switching tendencies. However, this method considers only the customer's switching behavior but not simultaneous purchasing behavior (i.e., purchasing from multiple brands at the same time). Furthermore, their approach relies heavily on the data gathered from large-scale surveys of customers. This may prove impractical in the CPG setting due to the cost structures associated with data collection and inherent reporting biases within the survey data. The lack of consumption and other marketing-related data has proven to be very difficult to fill, especially in a CLV setting. However, the rise of cooperative databases, wherein data sets across multiple firms are pooled by third-party vendors, has enabled researchers to have a clearer view of the customer (Liu, Pancras, and Houtz 2015). Our approach to handling competition follows a similar idea. Leveraging data from third-party vendors such as Nielsen/IRI, we directly include competition within the consumer's utility and implement a unified CLV model on transaction data from scanner panel data.

Choice, quantity, and timing modeling. Previous research on CLV modeling has mostly relied on separate specifications of choice, quantity, and timing decision models to describe customer decision making (Gupta et al. 2006; Kumar and Luo 2008). While these models have worked well in situations in which customer relationships are well defined, they may not be well suited for the CPG context. A choice-then-quantity approach forces the researchers to make explicit assumptions regarding the temporal ordering of decisions. In a CPG setting, this assumption may not hold, especially when consumers purchase more than one brand in the same purchase occasion and switching costs are relatively low. Specification of separate choice, quantity, and timing models could lead to parameter proliferation problems as well as the introduction of new random utility error terms (for each decision model) into consumer preference (Chintagunta and Nair 2011). Furthermore, a reduced-form approach of specifying joint models of multiple decisions could suffer from the Lucas critique. This is true with dynamic models, such as vector autoregression models and other multivariate time series models that are commonly used in CLV modeling. Thus, we propose a unified structural model that incorporates all of the aforementioned consumer decisions within the same utility framework, thereby avoiding the parameter proliferation problem while still modeling CLV.

In the CPG context, Yoo, Hanssens, and Kim (2011) adopt a vector autoregression modeling–based approach to model customer equity and link it to marketing investments. This approach, however, is applicable only for a one-brand-one-category setting and does not address the multiple discreteness issue that is common in CPG purchases (Allender et al. 2013; Dubé 2004; Kim, Allenby, and Rossi 2007; Richards, Gómez, and Pofahl 2012).

Models of Multiple Discreteness

In the CPG setting, consumers tend to purchase assortments of products/brands in a shopping trip, thus leading to the multiple discreteness problem. The multivariate probit model (Manchanda, Ansari, and Gupta 1999), which essentially treats the consumer choice decision as a set of correlated binary choice models, has been proposed to handle this issue without the use of a structural modeling approach. While popular in marketing literature, this approach is suboptimal when studying CLV because it does not make any conclusions regarding the quantity decision, which is critical for CLV computation. Direct utility structural models that derive demand from Karush–Kuhn–Tucker conditions have been proposed as a viable alternative for modeling multiple discreteness while taking advantage of the continuous nature of consumer purchase. Variants of these models include those proposed by Kim, Allenby, and Rossi (2002), Bhat (2008), and Satomura, Kim, and Allenby (2011), which rely on satiation to explain multiple discreteness. An alternative approach in the economics literature was proposed by Hendel (1999), who treats multiple discreteness as temporary variety-seeking behavior. This approach was later applied in marketing by Dubé (2004) to study demand for carbonated soft drinks. In the current study, we adopt a direct utility approach to structurally model multiple discreteness while accounting for variety-seeking behavior in the demand model.

This research falls within the broader areas of multiple discreteness modeling and CLV as we attempt to develop and apply a structural multiple discrete methodology to measure CLV in a CPG setting. Furthermore, unlike much of the prior work on multiple discreteness modeling, through Bayesian updating, we are able to also infer the consumer's unobserved budget constraint.[5] In the next section, we describe the data with which the empirical model was developed and implemented.

DATA

The empirical setting for the application of the proposed CLV model is the CPG industry. Specifically, we used scanner panel data for carbonated beverages, obtained from Nielsen/IRI, in our subsequent analyses. In the data, we observe at the Universal Product Code (UPC) level monthly carbonated soft drink purchases made by 40,098 consumers who were part of the Nielsen panel between July 2007 and August 2010.

Next, we describe the criteria used in preparing the data in order to develop and estimate the proposed model. First, a common challenge in modeling scanner panel data is to devise an aggregation strategy such that a tractable set of choices/alternative are used for estimation. Following Gordon, Goldfarb, and Li (2013), we aggregated UPC-level data within the category into manufacturer-level brands.[6] This yielded a data set that considered customer purchases across four major brands (Coca-Cola, Pepsi, Dr. Pepper, and private labels), which account for a cumulative 89% of market share in the overall sample. A second issue that comes up when building models using only customer-level scanner panel data is that the researcher observes price only when the consumer makes a purchase of the focal brand. In order to infer the missing price data for the other brands, we follow the heuristic outlined by Gordon, Goldfarb, and Li (2013). We imputed the price information using purchases by other consumers in the same store type in the same month. That is, when customer i did not purchase brand b at time t, we searched the database for any other customers k (where $k \neq i$) who purchased brand b at time t in a similar store type. We then computed the average of the price across the customers k to arrive at an imputed price, which we use in place of the missing information.

Model-Free Analyses

To provide a deeper understanding of the data structure and patterns, we provide visual representations of key trends in the data. First, in Figure 1, we illustrate the time trend in market share as well as price for the four major brands in the data. We can see that there are two leaders in the market, Coca-Cola and Pepsi, who each command an average of about 30% market share. Visual inspection suggests that these two brands seem to be close competitors and seem to steal market share from one another on a month-to-month basis. This idea is further supported when we study the time trend of price data. In contrast, we can see that Dr. Pepper's market share is increasing over time, as is its price. These trends indicate that there is significant competition between brands in this market, and in addition to pricing, there are several factors that could be influencing this.

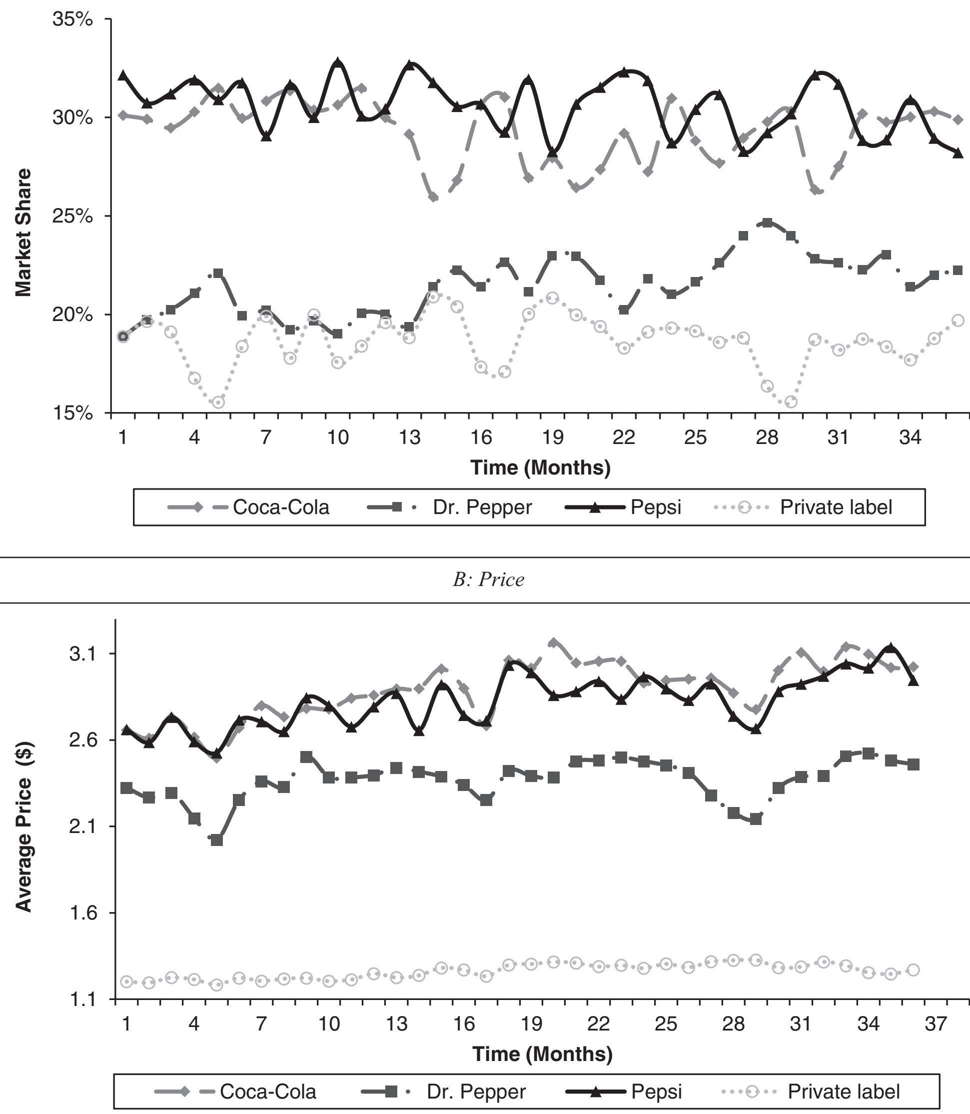
While aggregate metrics give managers an indication of the health of the brand and serve as an aggregate proxy for performance, they do not provide in-depth information about which customers grew and which ones did not. Further, they do not address the inherent heterogeneity among customer responses to marketing. To illustrate this point, we compute household-level market share (the percentage of purchases of

[5]An exception is Satomura, Kim, and Allenby (2011), who estimate budget constraint in a controlled conjoint setting. We apply our methodology creatively on real-world transaction data to assess the future profit stream at the customer level (CLV).

[6]The proposed CLV model could accommodate more granular (brand-size) data, provided there was enough variation in consumption patterns.

Figure 1
TIME TRENDS IN KEY VARIABLES

A: Market Share



the focal brand relative to total number of purchases). Figure 2 describes the distribution of household-level market share across the four major brands we consider.

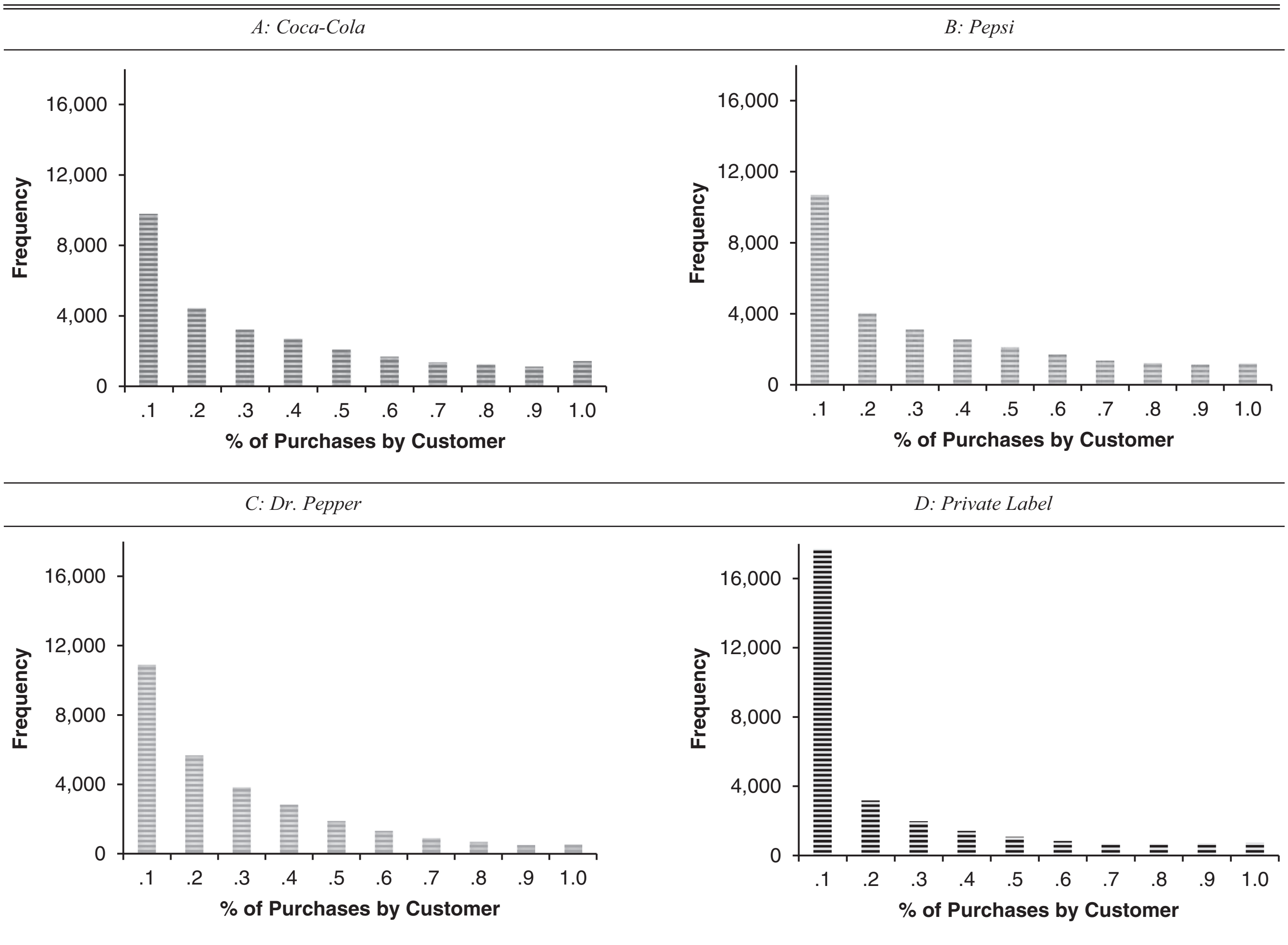
There are two key points to note in Figure 2. First, there is a wide variation in customer purchases, suggesting that heterogeneity is indeed important and needs to be considered. Second, the distribution of purchases across brands also varies. A key question is that of how customers react when one brand decides to modify its price. That is, given an increase in the price of Coca-Cola, customers could (1) increase their consumption of another brand and reduce their share of Coca-Cola purchases while still maintaining their overall consumption level, or (2) continue to purchase Coca-Cola but reduce the quantity consumed to remain within the budget constraint. Overall, the variations in the data help motivate the need to use a sophisticated modeling approach to accurately address these issues. In the following subsection, we further motivate the need for applying a multiple discreteness framework to the current context by providing evidence from the data and literature.

Multiple Discreteness Check

To check the extent of multiple discreteness within the data (40,098 customers), we computed the number of interior and corner solutions among the consumers. Table 1 displays the results.

From Table 1, we can see that about 45% of the transactions are in fact interior solutions, leading to the multiple discreteness issue, which exists at the weekly level as well. We found that at the weekly level, almost 30% of the

Figure 2
CUSTOMER-LEVEL PURCHASE DISTRIBUTION



transactions are multiple discrete. We cross-checked data from other categories (canned pasta and yogurt) for evidence of the multiple discreteness and found that results that were consistent with our findings in the carbonated beverages category. It is important to note that although we implement the proposed model for the carbonated beverages category, the proposed framework is easily adaptable to other CPG categories[7] as well. The conventional methods of CLV modeling (which rely on classic choice, frequency/timing, and quantity modeling) would end up combining a large percentage of the transactions into a single brand purchase, which could in turn significantly bias the estimation and lead to inaccurate CLV calculations. The proposed model not only accounts for the aforementioned multiple discreteness issue but also integrates the three main decisions (choice, frequency/timing, and quantity) within the same model.

METHODOLOGY

Formally, CLV is defined as the sum of the cumulated cash flows—discounted using the weighted average cost of capital—of a customer over his/her entire lifetime (Venkatesan, Kumar, and Bohling 2007). The lifetime value of the customer has two components, predicted contribution margin and predicted marketing cost:

$$(1)\qquad \mathrm{CLV_i = NPV\ of\ GC_i - NPV\ of\ MC_i} = \sum_{t=t_1}^{T}\sum_{j=1}^{J}\frac{\widehat{q}_{ijt}(m_j P_{jt})}{(1+d)^{t-t_1}} - \frac{\overline{MC}_{jt}}{(1+d)^{t-t_1}},$$

where GC_i = gross contribution made by customer i; $\widehat{q}_{ijt}$ = predicted quantity of brand j purchased by customer i at time t (in units); m_j = profit margin of brand j (as a percentage); P_{jt} = price of brand j at time t (in dollars); $\overline{MC}_{jt}$ = average marketing cost per customer incurred by brand j at time t; d = discount rate (12% annually); and j = brand indicator (ranging from 1 to J).

The first term in Equation 1[8] depicts the profit stream of each customer in the database and discounts this value to the present

[7]Even for categories that do not exhibit very high multiple discreteness, the proposed model will simplify to a discretized modeling framework, thus simplifying estimation.

[8]Equation 1 can be easily modified to account for variations across retailers as well. For example, if we observe retailer-level data, we can incorporate this information in the CLV computation by including another summation term, thus allowing CLV to be computed at the brand, retailer, and category levels. We elaborate on this modification in the "Discussion and Implications" section.

Table 1
INCIDENCE OF MULTIPLE DISCRETENESS IN DATA

Number of Brands Purchased	*Number of Transactions*	*Percentage of Total Transactions*
1	426,096	54.61
2	251,249	31.20
3	88,741	11.37
4	14,229	1.82

value. The second term in Equation 1 describes the marketing expenses borne by the firm toward customer i. Specific to our case, CPG firms do not market individually to each customer. Instead, CPG customers are typically reached through mass-marketing channels such as television commercials, newspaper inserts, in-store displays, and so on. Thus, the marketing cost per customer in the CPG setting is likely to vary across brands but not much across customers. In the empirical application presented in our study, we assume a zero-base marketing spending, similar to Yoo, Hanssens, and Kim (2011). The model describing the customer's budget-constrained utility maximization problem is presented in the following sections, along with brief discussions of each component.

To model the stochastic component $(\widehat{q}_{ijt})$, we propose a paramorphic (as opposed to a process model of psychological processes) or structural approach of purchase behavior wherein the consumer maximizes his/her utility for each trip across a variety of brands. We adopt this approach because our main objective is to be able to predict behaviors or actions at the household level and eventually to compute CLV in the CPG setting.

The Budget Constraint in the CPG Context

In this subsection, we elaborate on the theoretical underpinnings of the budget constraint construct, its boundaries, and its definition. Literature on mental accounting (Cheema and Soman 2006; Thaler 1985) has shown that consumers impose restrictions on themselves to avoid overspending and overconsumption. These restrictions are usually in the form of mental dollars that consumers assign toward consumption, which have been shown to exist in the grocery setting (Milkman and Beshears 2009). Furthermore, research in categorization (Antonides, Manon de Groot, and Van Raaij 2011; Ratneshwar, Pechmann, and Shocker 1996) has shown that consumers represent products and substitutes differently. Thus, a model of consumer purchase behavior that imposes a budget constraint at predefined category level needs to account for substitutes both within and outside the category. Thus, we specify the budget constraint, or monetary ceiling, to be the maximum monies allocated by the consumer toward the focal category as well as substitutes that may be considered outside the focal category. For example, the budget constraint that we attempt to quantify in this study is the maximum amount of money that the consumer allocates toward the carbonated soft drinks category plus substitute product categories (water, juice, etc.).[9] In the following section, we develop the consumer's overall utility maximization problem and describe the salient features of the model.

Consumer's Utility Specification

In this section, we specify a direct utility model in which consumers are assumed to be utility maximizers subject to a budgetary constraint. The consumer's overall utility (U_{it}) can be expressed as a function of his/her utility from consumption and category-level savings. The savings utility, which tracks the overall spending within the focal category as well as the budget constraint, acts as a counterbalance to the consumption utility (Prelec and Loewenstein 1998). Utility from consumption is derived from purchase of specific brands from a subset of offerings. Typically, from a discrete modeling approach, this is the utility derived when a consumer purchases a brand. In this context, due to the multiple discreteness issue, the consumer is assumed to purchase a set of brands (as opposed to one brand). The consumption utility (U_{it}^{Cons}) is therefore a sum of utilities ($\sum_{j=1}^{J} U_{ijt}$) that the consumer gains from consuming/purchasing a set of brands. The second component of the consumer's overall utility is the utility from savings (U_{it}^{Sav}).

The consumer's category-level utility from savings is described as a function of his/her category-level monetary savings from a shopping trip. We specify the monetary savings as the difference between the consumer's budgetary ceiling, or mental account (y_{it}), and the amount of money spent toward the category at time t ($\sum_{j=1}^{J} P_{jt} q_{ijt}$). The budget constraint (y_{it}) is the maximum allocation to goods in a mental category (focal product category + substitutes outside the product category) and helps ensure that the overall utility is concave with positive, but diminishing, marginal returns:

$$(2) \qquad U_{it} = U_{it}^{Cons} + U_{it}^{Sav} = \sum_{j=1}^{J} U_{ijt} + f\left(y_{it} - \sum_{j=1}^{J} P_{jt} q_{ijt}\right),$$

where U_{it} = overall utility from consumption by consumer i at time t; U_{ijt} = brand-level utility for consumer i at time t for brand j; y_{it} = unobserved budget allocation within category by consumer i at time t; P_{jt} = price of brand j at time t; and q_{ijt} = quantity of brand j consumed by consumer i at time t.

The overall utility U_{ijt} from Equation 2 can be decomposed into subutilities for each brand (Equation 3). The "1 +" in "$1 + q_{ijt}$" in Equation 3 allows for the possibility of corner solutions in the model, where q_{ijt} can take zero values. This specification is important since there could be situations wherein a consumer who is extremely loyal to a specific brand will never purchase any other brand, thus leading to quantity demanded for other brands to be zero. Furthermore, this formulation works well for CLV modeling because it incorporates choice, quantity, and frequency (or timing) decisions within the same utility specification. Thus, the current modeling approach avoids problems of overspecification and maintains model parsimony, while still addressing multiple discreteness and the budget-constrained nature of consumer decision making. The utility function in Equation 3 is concave and allows for diminishing marginal returns. That is, according to Equation 3, if a household purchases more of brand j, the marginal utility derived will

[9]We impose no restrictions on the manner in which these dollars are allocated across substitutes. Furthermore, although we present a monetary budget constraint, the constraint could be related to space/inventory as well. For example, a household could have constraints on storage that would act as a quantity ceiling. We expect that this would be correlated with the monetary constraint modeled in this study.

decrease for that brand (diminishing marginal utility). While we do not explicitly model the mechanism for decreasing marginal utilities (e.g., for satiation or overconsumption), the utility function accounts for it. The overall consumer utility at time t is thus given by

$$(3)\qquad U_{it} = \sum_{j=1}^{J}\left[\psi_{ijt}\ln\left(1+q_{ijt}\right)\right] + \lambda_i \ln\left[y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)\right].$$

The savings side of U_{it} can be described log-linearly where λ_i (Equation 3) is introduced to convert the monetary savings into utility. In an economic sense, the Lagrangian multiplier λ_i can be interpreted as the marginal effect of the monetary constraint on the overall utility of the customer or the rate of change in U_{it} relative to the monetary constraint (Baxley and Moorhouse 1984). It helps us transform the constrained optimization problem into an unconstrained optimization problem by providing a link between the constraint (y_{it}) and the overall utility (U_{it}). This method of using Lagrangian multipliers to solve constrained optimization problems has been used by various researchers in economics, marketing, and engineering (Chintagunta and Nair 2011). The baseline utility (ψ_{ijt}) in Equation 3 can be written as a function of stochastic (ε_{ijt}) and deterministic (ψ^*_{ijt}) parts:

$$(4)\qquad \psi_{ijt} = \psi^*_{ijt} + \varepsilon_{ijt}, \text{ and}$$
$$\varepsilon_{ijt} \sim N\left(0, \sigma^2\right).$$

Equation 3 leads to the Karush–Kuhn–Tucker conditions of constrained utility maximization wherein interior ($q_{ijt} > 0$) or corner solutions ($q_{ijt} = 0$) are possible. We can derive the overall likelihood by connecting the error (ε_{ijt}) to the observed demand (q_{ijt}) in each of these conditions. When consumer i purchases brand j at time t, yielding observed demand (q_{ijt}) greater than 0 (interior solution), the first-order condition for Equation 3 leads to a normal density function:

$$(5a)\qquad \frac{\partial U_{it}}{\partial q_{ijt}} = \frac{\psi_{ijt}}{1+q_{ijt}} - \frac{\lambda_i P_{jt}}{y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)} = 0, \text{ if } q_{ijt} > 0;$$
$$\Rightarrow \varepsilon_{ijt} = \frac{\lambda_i P_{ijt}\left(1+q_{ijt}\right)}{y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)} - \psi^*_{ijt}, \text{ if } q_{ijt} > 0.$$

Alternatively, the consumer may not purchase brand j at time t, thus yielding observed demand (q_{ijt}) equal to 0. This leads to a probability mass function and denotes the corner solution:

$$(5b)\qquad \frac{\partial U_{it}}{\partial q_{ijt}} = \frac{\psi_{ijt}}{1+q_{ijt}} - \frac{\lambda_i P_{jt}}{y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)} < 0, \text{ if } q_{ijt} = 0;$$
$$\Rightarrow \varepsilon_{ijt} < \frac{\lambda_i P_{ijt}\left(1+q_{ijt}\right)}{y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)} - \psi^*_{ijt}, \text{ if } q_{ijt} = 0.$$

We now link the baseline utility to covariates by specifying the deterministic portion (ψ^*_{ijt}) as a linear function of covariates that describe the customer's purchase behavior. In Equation 6, we include full heterogeneity in the intercept and the state dependence parameters while including brand-specific parameters for the other variables:

$$(6)\qquad \psi^*_{ijt} = \alpha_{ij} + \delta_i SD_{ijt} + \beta_j X_{it},$$

where α_{ij} = brand-specific (j) and customer-specific (i) intercept term; δ_i = customer-specific (i) state dependence parameter; SD_{ijt} = state dependence variable (measured currently as a dummy variable that takes the value 1 if customer i bought brand j at time t − 1 and 0 otherwise); β_j = brand-specific (j) parameter; and X_{it} = customer-specific (i) variable at time t.

We can further decompose the budget constraint parameter (y_{it}) to vary with time as a function of factors that are both intrinsic as well as extraneous to the environment. In the current implementation, we decompose the budget constraint parameter to be a function of demographics (age) and seasonality effects (summer months).[10] We include Age^2_{it} in Equation 7 in order to test for any quadratic effects of age on the budgetary constraint for each customer. We also expect that the consumer's budget does not stay the same throughout the year. Especially for frequently purchased goods, the consumer's budgetary allocation changes depending on seasonal effects. To account for this, we also include a seasonality dummy variable to capture the effects of seasonality on the consumer's budget allocation. Thus,

$$(7)\qquad y_{it} = \zeta_{0i} + \zeta_1 Age_{it} + \zeta_2 Age^2_{it} + \zeta_3 Seas_t,$$

where ζ_{0i} = baseline budget constraint parameter (estimated) for consumer i; Age_{it} = age of consumer i at time t; and $Seas_t$ = dummy variable that takes the value 1 if month = May–August (summer months) and 0 otherwise.

We incorporate heterogeneity in the consumer's inherent brand preference parameter (α_{ij}), the state dependence coefficient (δ_i), and the baseline budget parameter (y_i). We assume that the aforementioned coefficients follow a normal distribution with location parameters specified as follows:

$$(8)\qquad \alpha_{ij} \sim N\left(\overline{\alpha_j},\ V_{\alpha j}\right);$$
$$\delta_i \sim N\left(\overline{\delta},\ V_\delta\right); \text{ and}$$
$$\zeta_{0i} \sim N\left(\overline{\zeta_0},\ V_{\zeta_0}\right),$$

where $(\overline{\alpha_j},\ V_{\alpha_j})$, $(\overline{\delta},\ V_\delta)$, and $(\overline{\zeta_0}, V_{\zeta_0})$ represent the population means and variances of the distributions of α_{ij}, δ_i, and ζ_{0i}, respectively.

Likelihood

Using the assumption of normal errors, Equations 5a and 5b can be combined to form the overall likelihood, which is a combination of density (for interior solution) and mass (for corner solutions). We represent the parameter space as an array Θ_i for expositional purposes, such that $\Theta_i = \{\alpha_{ij}, \delta_i, \beta, \zeta_{0i}, \zeta_{1-3}\}$, and write the likelihood for household i as

[10]The budget constraint parameter varies across time deterministically (as a function of exogenous variables). The specification and estimation of a dynamic stochastic budget constraint model is beyond the scope of this research.

$$(9)\quad L_i(\Theta) = \int_{-\infty}^{\infty} L_{0i}(\Theta_i)^{I(q_{ijt}>0)} \times L_{1i}(\Theta_i)^{[1-I(q_{ijt}>0)]} f(\Theta_i)d\Theta_i$$
$$= \int_{-\infty}^{\infty} \prod_{t=1}^{T}\prod_{j=0}^{J}\left[\phi(\varepsilon_{ijt}) \times |J|_{\varepsilon_{ijt}\to q_{ijt}}\right]^{I(q_{ijt}>0)}$$
$$\times\, \Phi(\varepsilon_{ijt})^{[1-I(q_{ijt}>0)]} f(\Theta_i)d\Theta_i$$

where

$$I(q_{ijt}>0) = \begin{cases}1; & \text{when } q_{ijt}>0\\ 0; & \text{else}\end{cases};$$

$\phi(.)$ = probability density function of the normal distribution; $\Phi(.)$ = truncated normal distribution; $|J|_{\varepsilon_{ijt}\to q_{ijt}}$ = Jacobian of the transformation from the random utility error (ε_{ijt}) to the likelihood of observed data (q_{ijt}); and $f(\Theta_i)$ = heterogeneity distribution of parameter space Θ_i with location parameters $\overline{\Theta}$, V_Θ.

The Jacobian for our model is given by the first order derivative of the error term with respect to q_{ijt}, as follows:

$$(10)\quad |J|_{\varepsilon_{ijt}\to q_{ijt}} = \frac{\partial \varepsilon_{ijt}}{\partial q_{ijt}} = \frac{\lambda_i P_{jt}}{y_{it} - \sum_{j=1}^{J} P_{jt}q_{ijt}} + \frac{\lambda_i P_{jt}^2\left(1+q_{ijt}\right)}{\left(y_{it} - \sum_{j=1}^{J} P_{jt}q_{ijt}\right)^2}.$$

Let N be a collection of all i households in the data. Then, the overall likelihood for the data can be given by

$$(11)\quad L(\Theta)_{overall} = \prod_{i=1}^{N} L_i(\Theta).$$

In the following section, we comment on the theoretical and empirical identification issues faced when estimating the proposed model.

Model Identification

Given the structure of our model, it is important to provide some intuition regarding the identification of the model parameters. The overall utility model (Equation 3) consists of two main components that need to be estimated in order to achieve our stated objectives, namely, (1) the baseline utility, ψ_{ijt}, through its associated hierarchical parameters (α_{ij}, δ_i, and β_j); and (2) the budget constraint, y_{it}, through its associated hierarchical parameters (ζ_{0i}, ζ_1, ζ_2, and ζ_3). Recall that according to Equation 7, y_{it} is allowed to vary deterministically as a function of a baseline budget constraint (ζ_{0i}) along with exogenous covariates. An identification problem arises when we attempt to simultaneously evaluate the intrinsic preference at the brand level α_{ij}, the baseline budget constraint ζ_{0i}, and the Lagrangian λ_i. That is, it is possible that one could generate the same observed data (P_{jt} and q_{ijt}) using more than one unique combination of the parameters α_{ij}, ζ_{0i}, and λ_i. Thus, given the data (which include price and quantity information at the customer–brand level), it is not possible to empirically identify all three parameters listed. Therefore, we need to fix at least one of these parameters in order to identify the others jointly. As stated before, our main parameters of interest are the baseline utilities and the budget constraint parameter. To uniquely identify α_{ij} and ζ_{0i}, we first fix $\lambda_i = 1$ and $\sigma^2 = 1$. We provide more details on the specific elements in the data that allow us to reliably recover the parameters, as well as theoretical arguments on identification, in Web Appendix A.

The budget constraint (y_{it}) is modeled in exponential form in order to constrain it to positive values (since it is impossible to have negative budgets). We also impose logical ceilings on the budget parameter such that the estimated value for customer i does not exceed the observed maximum purchase value (in dollars) within the data, such that $y_{it} \geq \max(\sum_{j\in J} P_{jt}q_{ijt})$.

Estimation

We estimated the proposed model using a hybrid Bayesian Markov chain Monte Carlo algorithm. The use of Bayesian methods is needed because one of our objectives is to infer the budget constraint (y_{it}). The Bayesian approach allows us to create latent variables, use data augmentation methods, and estimate the parameters sequentially. The assumption of normal errors allows us to break down the estimation process into more efficient Gibbs sampling (from full conditionals) and Metropolis–Hastings (M-H) sampling methods.

We outline our estimation process next. We first begin by drawing ψ_{ijt} according to whether q_{ijt} is equal to or greater than 0. In the case when $q_{ijt} > 0$ (interior solution), we use the normal distribution to infer ψ_{ijt}, and when $q_{ijt} = 0$ (corner solution), we use the truncated normal distribution to infer ψ_{ijt}. Given ψ_{ijt}, we now treat the underlying estimation of α_{ij}, δ_i, and β_j similar to a multivariate regression with heterogeneous parameters, which can be estimated using Gibbs sampling. The remaining parameters (ζ_{0i} and ζ_1–ζ_4) are drawn using the M-H algorithm because we cannot derive the full conditional distributions for them. We specify the prior distribution on the hyperparameters ($\overline{\alpha}_j, V_{\alpha_j}$), ($\overline{\delta}, V_\delta$), and ($\overline{\zeta}_0, V_{\zeta_0}$) to be noninformative and flat. The prior means were normally distributed and the prior variances were inverse Wishart distributed. Our overall estimation algorithm is described in more detail in the Appendix.

Variable Operationalization

As elucidated in Equation 6, we introduce brand- and customer-level covariates to explain variance in the baseline utility equation. We elaborate on the variables used in this study next.

State dependence. Following prior literature on state dependence in consumer choice, we include a state dependence term (SD_{ijt}) to track the inertia in the consumer's purchase pattern. In the current implementation, we specify state dependence as a dummy variable similar to that of past research investigating state dependence in choice modeling (Dubé, Hitsch, and Rossi 2010; Seetharaman, Ainslie, and Chintagunta 1999). Specifically, if the consumer buys brand j during the previous shopping occasion (t − 1), then the state dependence term for that brand is equal to 1:

$$(12)\quad SD_{ijt} = I\left\{q_{ijt-1} > 0\right\}.$$

The specification in Equation 12 induces a first-order Markov process on choices. It is also important to note that SD_{ijt} is brand specific and can take multiple nonzero values for each purchase occasion, due to the multiple discreteness issue (whereby the consumer could have purchased more than one brand at t − 1). We refer to δ_i as the state dependence coefficient that captures the effect of the state dependence term (SD_{ijt}). If $\delta_i > 0$, the model implies that the purchase of a brand reinforces the household's latent utility for that brand.

Table 2
VARIABLE OPERATIONALIZATION

Variable	*Operationalization*
State dependence (SD_{ijt})	SD_{ijt} = 1 if $q_{ijt-1} > 0$, and 0 otherwise (adapted from Dubé, Hitsch, and Rossi 2010).
Category consumption intensity ($LASTQty_{it}$ and $Recency_{it}$)	$LASTQty_{it}$ is measured as the total quantity purchased by consumer i at time t − 1. (adapted from Chintagunta and Haldar 1998). $Recency_{it}$ is measured as the time (in months) since the last purchase for consumer i (adapted from Kumar and Shah 2009).
Deal usage intensity ($DEAL\_intensity_{it}$)	$DEAL\_intensity_{it}$ is measured as the cumulative number of times that consumer i has purchased the brand when it was offered as a deal (expressed as percentage of total number of purchases made). It must be noted that the measure is updated as t increases and is also normalized by the denominator, restricting values between 0 and 1 (adapted from Shah, Kumar, and Kim 2014).
Coupon usage intensity ($COUP\_intensity_{it}$)	$COUP\_intensity_{it}$ is measured as the cumulative number of times that consumer i has purchased the brand when using a coupon (expressed as percentage of total number of purchases made) (adapted from Shah, Kumar, and Kim 2014).
Store usage intensity ($STORE\_intensity_{it}$)	$STORE\_intensity_{it}$ is measured as the cumulative number of purchases made in a specific store format (in this study, food stores) as a percentage of total number of purchases made. Similar to other intensity measures (Shah, Kumar, and Kim 2014), this measure is updated as t increases.
Diet soda purchase intensity ($DIETSODA\_intensity_{it}$)	$DIETSODA\_intensity_{it}$ is measured as the cumulative number of diet soda purchases as a percentage of total purchases made by the consumer (adapted from Shah, Kumar, and Kim 2014).

By accounting for brand- and customer-specific intercepts (α_{ij}), we capture the household's underlying preferences for brands and also explicitly separate them from the household's tendency to be state dependent ($\delta_i > 0$).

Past purchase behavior. In Equation 6, we also specify X_{it} as a matrix of customer-level variables that drive consumer purchase behavior. Table 2 shows the variables used in this study, as well as their operationalizations and expected effects.

To capture the consumer's consumption intensity within the category, we use total quantity purchased at the previous purchase occasion ($LASTQty_{it}$) and the recency of last purchase ($Recency_{it}$). These variables are expected to explain the consumer's category-level consumption patterns by accounting for the incidence of a past purchase as well as the depth of the previous purchase. Prior research has shown that there exists a negative effect of recency of purchase on CLV (Kumar and Shah 2009). In the CPG context, recency will have a negative effect on quantity purchased. That is, the longer the time since the last purchase, the less likely the customer is to purchase within the category. For example, consumers who have not made a category purchase recently (high values of $Recency_{it}$) are likely to have churned and thus derive much lower utility from consuming the brand. To capture the effect of the depth of the previous purchase, we include the lagged values of quantity purchased as a covariate (Chintagunta and Haldar 1998; Jain and Vilcassim 1991). This variable will also account for observable differences in consumption among households (e.g., heavy vs. light users) as well as control for category consumption levels per household (Jain and Vilcassim 1991).[11]

The general behavioral tendency of a customer to selectively purchase brands that are offered as "deals" is defined in this study as $DEAL\_intensity_{it}$. This variable indexes the consumer's deal usage intensity, or the extent to which the consumer purchases brands that are featured as deals or prominently displayed within the store. The role of deals in the CPG setting is not only to provide monetary savings to the customer but also to signal quality. Past research has shown that deal usage with regard to national brands (which command higher loyalty) is associated with higher perceived savings (Ailawadi, Neslin, and Gedenk 2001) and results in higher derived utility. Thus, deal usage is expected to have a positive effect on utility for national brands. However, these latent savings are not perceived for store brands (which do not command loyalty or high perceived quality). Thus, highly deal-intensive consumers would, in fact, derive a lower utility for private labels, leading to lower purchase quantities.

Similar to the deal intensity variable, we operationalize $COUP\_intensity_{it}$ in order to capture the coupon usage behavior of consumers. Consumers who are serial coupon users are likely to purchase only the value of the coupon offered rather than indulge in cross-buying or up-buying within the category (Shah, Kumar, and Kim 2014). Drawing parallels from this research, we expect that consumers who consistently use coupons are likely to purchase lesser quantities.

Because the data are from consumers who made purchases from either food stores or nonfood stores (e.g., drugstores), we can study whether consumers who are especially loyal to a specific kind of store are more or less likely to purchase within a category. Especially important is the fact that high purchase intensity at food stores might lead to different effects for different brands (Ailawadi, Pauwels, and Steenkamp 2008). For example, consumers who are frequent drugstore purchasers may not purchase private labels (possibly due to an availability issue). In this study, we use the variable $STORE\_intensity_{it}$ to measure whether store format loyalty influences the overall quantity purchased. Finally, with increasing attention being cast on the health impact of foods (especially carbonated sodas), consumers are moving toward diet sodas as an alternative due to their lower sugar and calorie content. In fact, recent research by Ma, Ailawadi, and Grewal (2013) shows that consumers diagnosed with diabetes change their consumption patterns to accommodate a diet lower in sugar and carbohydrates, which in our case translates to a shift from regular to diet soda. Diet products, thus, are likely to be

[11]The inclusion of $LASTQty_{it}$, along with the use of month-level data for aggregation, also helps control for inventory effects that might influence the budget constraint parameter.

perceived as having a higher utility due to their health-related advantages. Therefore, higher diet soda consumption in the past (measured as DIETSODA_intensity$_{it}$) is likely to lead to a higher consumption in the future. The summary statistics of the data are provided in Table 3.

RESULTS

Simulation Study

To check the robustness of our model specification and estimation methodology, we first conducted a simulation study to calibrate the performance of our model. Data were generated according to the utility specified in Equation 4, assuming a three-brand market. We generated consumption data for 500 consumers, each having an observation length of 20 time periods. All the parameters were well recovered, having true values within 95% credible intervals, thus confirming that our estimation method can recover the true parameters and can be implemented on real transaction data. Details on the simulation exercise are provided in Web Appendix B.

Model Evaluation and Performance

We estimate the proposed model on a randomly selected sample of 500 customers (total number of transactions = 12,837) from the previously described consumer scanner panel data for the carbonated beverages category.[12] We used 20,000 iterations of the Markov chain to generate parameter estimates, with the first 10,000 discarded as burn-in. In order to assess the performance of the model, we use the mean absolute deviation (MAD) and mean absolute percentage error (MAPE) to assess the predictive accuracy of our model. We rely on MAPE as a preferred metric to gauge model fit because it is unit-free and easier to interpret. We gauge model performance for in-sample as well as out-of-sample fit.

In this section, we compare our modeling approach with a more conventional choice and quantity modeling approach that is typical for extant CLV models (Gupta et al. 2006). First, we compare our proposed model that accounts for multiple discreteness with a conventional one that does not. Theoretically, a simpler discrete choice–quantity model would ignore the variety in the consumer's purchases (which prior literature has documented as being prevalent in several categories the CPG setting), thus giving managers a unidimensional view of the customer. The proposed model has several advantages over the conventional approach because it (1) takes variety of purchases (within the same period) into account, (2) explicitly addresses the role of competition in CLV, and (3) models the monetary ceiling (budget constraint) so that managers can more effectively assess the potential of the customer. From an empirical standpoint, the proposed model also presents advantages over a discrete choice–quantity model. Dubé (2004) shows that the multiple discrete model tends to perform better in predicting overall sales in the CPG setting and also finds differences in parameter estimates that could lead to suboptimal decision making.

We compared our model with a multivariate choice-then-quantity approach. Specifically, we estimated a multivariate probit choice model using the simulated maximum likelihood approach to predict customer choice across various brands and subsequently used a regression model to predict quantity (for model and estimation details, see Web Appendix C). To assess out-of-sample fit, we estimated the model using the first 30 months of data and used the remaining 6 months as holdout. In Table 4, we report in-sample and out-of-sample fit statistics (MAD and MAPE) for each brand, as well as overall category level quantity. As we can see, the proposed model predicts brand-level quantity purchased ($\widehat{q_{ijt}}$) quite well, yielding average MAPEs across brands of 20.74% (in-sample) and 23.09% (out-of-sample). When we consider the total category quantity purchased, the model performance dips slightly, to MAPEs of 27.75% (in-sample) and 29.87% (out-of-sample). This result is markedly better than that of the benchmark model, which has average MAPEs of 48.90% (in-sample) and 50.64% (out-of-sample) when predicting brand-level quantities. At the category level, the MAPEs are 41.61% (in-sample) and 43.89% (out-of-sample), which are both worse than those of the proposed model. The choice-then-quantity model performs much worse in this case because it involves specifying multiple equations (each associated with a random utility error) with several parameters. The proposed model is superior to the conventional CLV modeling approaches because it exploits quantity information within the choice framework and prevents parameter proliferation (Chintagunta and Nair 2011).

We also comment on the cost of relying purely on conventional CLV models to assess the drivers of purchase behavior in the CPG setting. To further demonstrate this idea, we estimated a conventional CLV model for Coca-Cola and compared the results with those of the proposed model. We found that a simpler discrete choice–quantity model greatly overestimates the effects of the drivers of consumer purchase behavior. Specifically, in Coca-Cola's context, ignoring multiple discreteness leads to a much higher parameter estimate for LASTQty$_{it}$ and COUP_intensity$_{it}$. Furthermore, ignoring multiple discreteness leads to a nonsignificant effect for DEAL_intensity$_{it}$, which could result in misguided decision making from the brand's perspective. These differences in parameter estimates highlight how conventional CLV approaches could lead to potentially inaccurate understanding of consumer purchase behavior in the CPG setting. Moreover, as we elaborate in our counterfactual analyses, the proposed model is able to elicit nonlinearities in price elasticities; that is, a drop in prices leads to a consistently larger impact on quantity purchased than an increase in price. This result, too, is largely unaccounted for in conventional CLV models that estimate a single linear price elasticity parameter. We believe that these differences are key issues that managers need to consider when making marketing mix decisions in the CPG setting.

Findings from Model Estimation[13]

Consumer's budget constraint. One of the main modeling issues that we deal with in this study is the explicit estimation of the consumer's budget parameter using Bayesian methods. To our knowledge, this is the first study to estimate the consumer's budget using transaction data and to

[12]We repeated the analysis for three different samples of 500 customers and arrived at similar estimation results.

[13]To establish external validity, we presented our findings to executives from one of the largest firms in this industry, who provided valuable qualitative insights corroborating the results.

Table 3
SUMMARY STATISTICS OF RELEVANT VARIABLES

Variable	*M*	*SD*	Correlations									
			$Price^{Coca\text{-}Cola}$	$Price^{Dr.\ Pepper}$	$Price^{Pepsi}$	$Price^{Private\ Label}$	$LASTQty_{it}$	$Recency_{it}$	$DEAL\_intensity_{it}$	$COUP\_intensity_{it}$	$STORE\_intensity_{it}$	$DIETSODA\_intensity_{it}$
$Price^{Coca\text{-}Cola}$	2.9	.16	1									
$Price^{Dr.\ Pepper}$	2.37	.12	.732***	1								
$Price^{Pepsi}$	2.83	.14	.836***	.735***	1							
$Price^{Private\ Label}$	1.26	.04	.725***	.243***	.558***	1						
$LASTQty_{it}$	6.92	6.21	−.008	.010	−.009***	−.017**	1					
$Recency_{it}$	1.3	.86	.053***	.022**	.035***	.049***	−.104***	1				
$DEAL\_intensity_{it}$	.25	.25	−.013	−.010	−.016*	−.012	−.110***	.011	1			
$COUP\_intensity_{it}$	.04	.09	−.003	−.002	−.001	.001	.035***	.001	.287***	1		
$STORE\_intensity_{it}$	.43	.27	−.020**	−.011	−.021**	−.022**	−.207***	.023***	.400***	.006	1	
$DIETSODA\_intensity_{it}$	.31	.28	.007	.006	.006	.005	−.084***	−.022**	.156***	.033***	.130***	1

*$p < .1$.
**$p < .05$.
***$p < .01$.

Table 4

MODEL PERFORMANCE

	MAD		*MAPE*	
	In-Sample	*Out-of-Sample*	*In-Sample*	*Out-of-Sample*
Proposed Model				
Brand-level quantity (q_{ijt})				
Coca-Cola	.54	.61	21.01	22.94
Dr. Pepper	.54	.65	23.51	26.70
Pepsi	.53	.55	19.57	21.12
Private label	.44	.53	18.88	21.60
Category-level quantity ($\sum_{j \in J} q_{ijt}$)	1.39	1.56	27.75	29.87
Benchmark Model				
Brand-level quantity (q_{ijt})				
Coca-Cola	1.09	1.21	50.91	51.64
Dr. Pepper	.78	.86	44.85	46.93
Pepsi	1.24	1.27	51.02	52.96
Private label	.74	.95	48.84	51.01
Category-level quantity ($\sum_{j \in J} q_{ijt}$)	3.54	3.81	41.61	43.89

use this estimation to calculate CLV. In Table 5, we report the parameter estimates for Equation 10. We find that the average consumer baseline budget allocation for the carbonated beverages category is exp(3.371) = \$29.40 for a month. Consistent with Du and Kamakura (2008), we find that there is significant heterogeneity in the budget parameter. This heterogeneity in the consumer's budget is important to consider, especially in the CPG industry, wherein each consumer/household can have different thresholds and priorities when allocating a budget toward a particular category.

In Figure 3, we can see the distribution of the estimated budget parameter across households in the estimation sample. As elaborated previously, the baseline budget, which is estimated at the household level, represents a "monetary ceiling" for the household. As we elaborate in the "Discussion and Implications" section, CPG companies could potentially build customer profiles for high-budget customers and try to achieve a larger portion of their SOW. This form of hybrid segmentation (based on CLV and the budget constraint) presents an important opportunity for marketers to reallocate resources toward the right customer segments. Further, we find that the age of the head of the household has a positive effect on the budget. Specifically, as the consumer ages, the budget allocation toward carbonated beverages increases. This result could be because as the consumer ages, the number of members in the household (family size) and the household income also are likely to increase. This increase in family size and income could lead to increases in budget constraint. Because the squared term is not significant, we conclude that the effect is only linear and not quadratic. The nonsignificance of the quadratic term could be due to the range of age that we observe in the data. Future research could further explore the long-term effect of age on the consumer's budget constraint.

Inertia effects. Consistent with Seetharaman, Ainslie, and Chintagunta (1999), we find that there exists inertia in the marketplace wherein consumers prefer to stick to their past experiences. This result is consistent with theoretical explanations of routinized response behavior (Assael 1974), especially in heavily advertised, convenience goods associated with limited informational search and stronger brand attitudes. Our results are congruent with prior literature on modeling structural state dependence. In particular, our results echo the findings of Seetharaman and Chintagunta (1998), who find strong inertial behavior among consumers, using similar Nielsen scanner data. Furthermore, similar to Seetharaman (2004) and Dubé, Hitsch, and Rossi (2010), we highlight the importance of unobserved heterogeneity in the state dependence term. Specifically, we find that the inertia effect is heterogeneous in that consumers vary in their levels of inertia (some consumers may be more variety seeking than others). By profiling customers who are more and less variety seeking, firms can identify consumer segments that may have a higher tendency to indulge in brand switching.

Brand-specific effects. Table 6 describes the brand-specific parameter estimates for baseline utility (ψ_{ijt}). Looking at the heterogeneous intercept term (α_{ij}), we find that consumers are heterogeneous in their intrinsic preference level for brands in the carbonated beverage category.

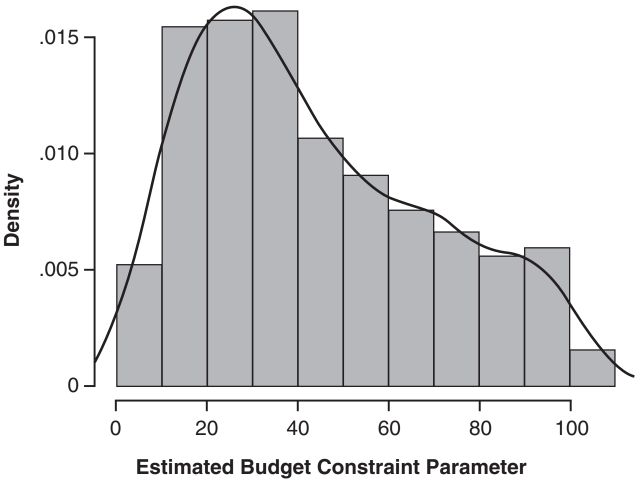
Table 5

BUDGET AND STATE DEPENDENCE PARAMETER ESTIMATES

Effect	*Parameter*	*M*	*SD*
Budget Constraint (y_{it})			
Intercept			
Mean	$\overline{\zeta_{0i}}$	3.371***	.026
Heterogeneity	V_{ζ_0}	.280***	.019
Extraneous factors			
Age_{it}	ζ_1	.017**	.009
Age_{it}^2	ζ_2	.001	.006
$Seas_t$	ζ_3	.013**	.006
State Dependence (SD_{ijt})			
Mean	$\overline{\delta_i}$	.148***	.017
Heterogeneity	V_{δ_i}	.030***	.005

$^{**}p < .05$.
$^{***}p < .01$.
Notes: The budget constraint (y_it) is modeled in log form.

Figure 3
DISTRIBUTION OF ESTIMATED BUDGET CONSTRAINT



Looking at the means of the α_{ij} distributions ($\overline{\alpha_{ij}}$), we find that the highest preference level is for Coca-Cola and the lowest is for private labels. This ordering follows the market share order in the category, where Coca-Cola has the largest market share and private labels have the smallest. As described previously, heterogeneity ($V_{\alpha_{ij}}$) is significantly large for this category, a result that is also demonstrated by Dubé (2004) in the same category. Notably, the heterogeneity term is large for Pepsi, indicating a high variance in intrinsic preferences among Pepsi's customer base.

Turning to the effect of the covariates, we find that $LASTQty_{it}$ positively affects the consumer's purchase behavior across all brands. That is, consumers who purchased large quantities in the past are likely to do the same in the current period. This result suggests that consumers do not necessarily take inventory into account when making frequent purchases in the carbonated beverages category. Specific to frequently purchased goods, heavy users could be developing habitual (or routinized) behavior of purchasing that leads to creation of behavioral loyalty. This finding is consistent with the general trend reported in CPG markets (pasta sauce and laundry detergent) by Chintagunta and Haldar (1998), who show that lag quantity decreases the hazard of purchase (and therefore increases probability of purchase). We also find that the effect of $Recency_{it}$ is significant and negative for Pepsi and Dr. Pepper but insignificant for the other brands (even though the sign of the coefficient is consistent). This result indicates that consumers who have not made a purchase in the category in a long time (high recency) have likely churned. Our results are congruent with Lewis (2004), who finds a negative effect of recency on future purchases such that the customer's propensity to continue purchasing decreases as the time since the previous order increases. The estimation results suggest that this variable is especially important for Pepsi and Dr. Pepper.

The results suggest that $DEAL\_intensity_{it}$ is positively associated with Coca-Cola and Dr. Pepper but negatively associated with private labels. Specifically, Ailawadi, Neslin, and Gedenk (2001) show that consumers who do not focus on the "deal" aspect of the purchase and therefore make fewer purchases of deals tend to gravitate toward store brands. Furthermore, consumers who tend to be quality conscious and deal prone tend to avoid private label brands and gravitate toward national brands. Turning to the effect of coupon usage behavior, consumers who are serial coupon users are found to be selective in their purchases and, thus, unlikely to exhibit high purchase behavior. This could be because these consumers purchase only the quantity/value indicated in the available coupon. Finally, we find that consumers who purchase frequently at food stores are likely to purchase private label brands. This could be a factor of the distribution intensity of private labels in these stores, which increases product availability. Although this result does not capture the retailer-specific loyalty, it gives managers an insight into which brands are more likely to be purchased in which store formats.

CLV IN THE CARBONATED BEVERAGES CATEGORY

CLV Measurement

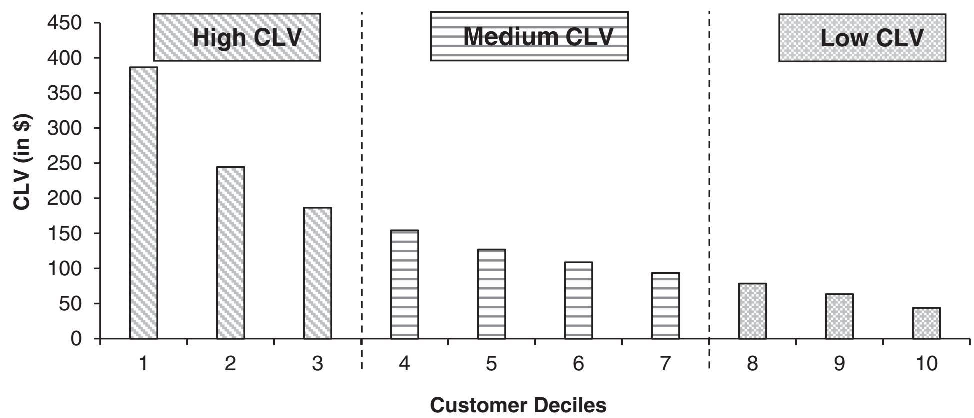
The main objective of this research is to compute the CLV of a customer in the CPG setting. Using the proposed

Table 6
BRAND-SPECIFIC PARAMETER ESTIMATES FOR BASELINE UTILITY

	Coca-Cola		*Dr. Pepper*		*Pepsi*		*Private Label*	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Intercept (α_{ij})								
Mean	−.501***	.049	−.662***	.060	−.607***	.069	−1.219***	.059
Heterogeneity	.399***	.034	.420***	.034	.830***	.070	.709***	.066
Covariates (β_j)								
$LASTQty_{it}$	.005**	.002	.005**	.002	.006**	.002	.007***	.003
$Recency_{it}$	−.001	.013	−.021**	.010	−.048***	.016	−.013	.016
$DEAL\_intensity_{it}$	.175**	.086	.339**	.114	−.009	.144	−.309**	.149
$COUP\_intensity_{it}$	−.036	.255	−.831**	.262	.355	.288	−.664**	.336
$STORE\_intensity_{it}$	−.107	.086	−.001	.101	−.106	.122	.248**	.115
$DIETSODA\_intensity_{it}$	.196**	.090	−.001	.094	.011	.117	−.061	.114

**$p < .05$.
***$p < .01$.

Figure 4
DISTRIBUTION OF CATEGORY-LEVEL CLV



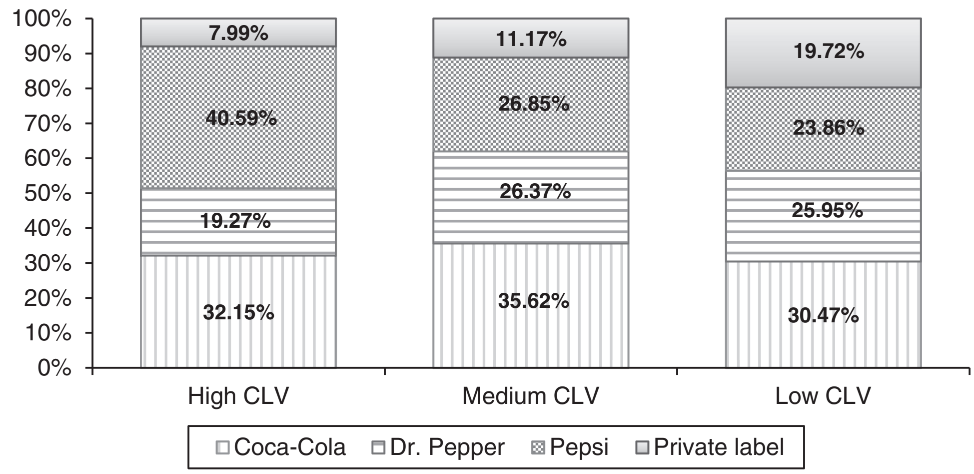
model, we can now predict the quantity purchased for each brand in the market ($\widehat{q_{ijt}}$), using the parameter estimates into the future and substituting the predicted values in Equation 1 to arrive at the CLV of a customer. First, we hold brand price (P_{jt}) at the mean and the brand-specific covariates (except $LASTQty_{it}$ and $Recency_{it}$) at the last recorded value for the CLV prediction, thus making the assumption that the consumer does not change his/her habits during the prediction window. Second, for each future period in the prediction window, we update the $LASTQty_{it}$, $Recency_{it}$, and SD_{ijt} variables according to the previous (predicted) values. Next, using the previously generated covariates, along with the parameter estimates, we generate the overall utility function (Equation 3) and subsequently maximize this expression to obtain purchase quantities for each brand. This process is repeated for the future time periods (36 months in our context). We choose a CLV prediction time window of 36 months for the following reasons. First, given the dynamic environment in which CPG firms typically operate, a prediction window of three years offers a good trade-off between predictive accuracy and horizon when computing CLV. Second, in general, the concept of discounting cash flows results in a majority of the customers' lifetime values being captured within the three-year window (Gupta and Lehmann 2005; Kumar and Shah 2009). For the context of the study, drawing on the guidance provided by Nielsen, we use a constant margin value of .28 for all the brands. Furthermore, following Yoo, Hanssens, and Kim (2011), we assume a marketing cost of zero without loss of generality because marketing investments in this category are made at the aggregate level and rarely vary across customers.

This analysis yields a mean CLV for a customer in this category of \$148.69, with a standard deviation of \$101.57. In order to investigate this distribution further, we summarize the CLV scores of the customers in ten deciles, where each decile represents the mean of 10% of the customers, organized in descending order of CLV scores (Figure 4). Similar to prior CLV work, we find that the bulk of dollars (in the form of CLV) are concentrated in the top few deciles. In fact, the first three deciles constitute almost 55% of the entire profits. This result, although familiar in a relationship marketing setting, is new to the CPG industry and presents further evidence that CPG brands need to move toward customer centricity rather than relying on aggregate measures of brand performance.

We also compared our proposed CLV segmentation approach with simpler heuristics that are commonly used by managers. The proposed CLV approach is an improvement over simpler naïve heuristics because it accounts for multiple discreteness, unobserved heterogeneity, competitive effect, variety seeking, and the consumer's budget constraint. Although past literature (Venkatesan and Kumar 2004) has shown that CLV outperforms conventional metrics and simpler heuristics in various business settings, we assess how well the traditional metrics match up against the proposed CLV. We focus primarily on purchase frequency, consumption level, and monetary value, which are commonly used in marketing practice due to their simple interpretation and implementation. Specifically, we segment customers according to these metrics and compute the mismatch between the deciles created using simpler heuristics and those created with the proposed CLV approach. That is, we compare how effectively the simpler heuristics segment customers relative to the approach presented in this study. If the mismatch were very low, then one could conclude that a simpler metric would suffice in evaluating the customer base. However, we find that across deciles, there is a significant mismatch between the metrics. The discordance between deciles was an average of 61.6% (79.6% for purchase frequency, 56% for total quantity consumed,[14] and 49.2% for total revenue) across metrics. In addition to this empirical evidence, we also highlight the strategic advantage that a model-based approach presents over a heuristic-based approach. A model-based approach (especially a structural model such as ours) not only outperforms heuristic-based methodologies in predicting customer behavior but also helps managers

[14]We also compared the proposed CLV classification with average consumption level (at the brand and category levels) and found the discordance to be high.

Figure 5
BRAND SHARE OF CATEGORY-LEVEL CLV



assess the drivers of customer purchase behavior and what strategies can influence it. This motivates the need to use model-based and predictive methods to assess customer value rather than relying on naïve heuristics that, while they may be easier to interpret and implement, may lead to suboptimal customer base evaluations.

Studying the Brand's Share of Total CLV

Our modeling approach allows us to study not only the CLV for the entire category but also the brand-level CLV for the category (Equation 1). Using the distribution of CLV scores (Figure 4) as a basis, we designate customers in deciles 1, 2, and 3 as "high CLV," deciles 4, 5, 6, and 7 as "medium CLV," and deciles 8, 9, and 10 as "low CLV." On the basis of this classification, we present the brand-level shares of CLV for high-, medium-, and low-CLV customers in the carbonated beverages category.

Figure 5 presents some notable results. Although Coca-Cola commands the largest market share in the carbonated beverages market, surprisingly, in the sample data set, Pepsi tends to attract a large percentage of the high-CLV segment (approximately 41%). This is further supported through the parameter estimates, in which we noted that the heterogeneity for inherent preferences was higher for Pepsi than for Coca-Cola, even though the mean preference level for Coca-Cola was greater. Furthermore, the majority of medium- and low-CLV customers are found to be Coca-Cola customers. We see that Coca-Cola seems to attract a majority of the low-CLV customers, presumably in an attempt to capture the "long tail." Although this strategy seems effective at the aggregate level, it is still important to capture the high-CLV customers because their spending power is higher. Finally, as expected, we see that private label customers tend to be few and predominantly lower-CLV customers. These customers tend to be value conscious and have little or no brand loyalty; the quality perceptions for private labels are lesser than those for national brands. Figure 5 represents an important status quo report of the state of brands in the carbonated beverages market with respect to CLV. Using the results, managers of each brand can gain a good understanding of the kind of customers that their respective brands have, rather than just using aggregate measures to assess brand performance.

POLICY SIMULATIONS

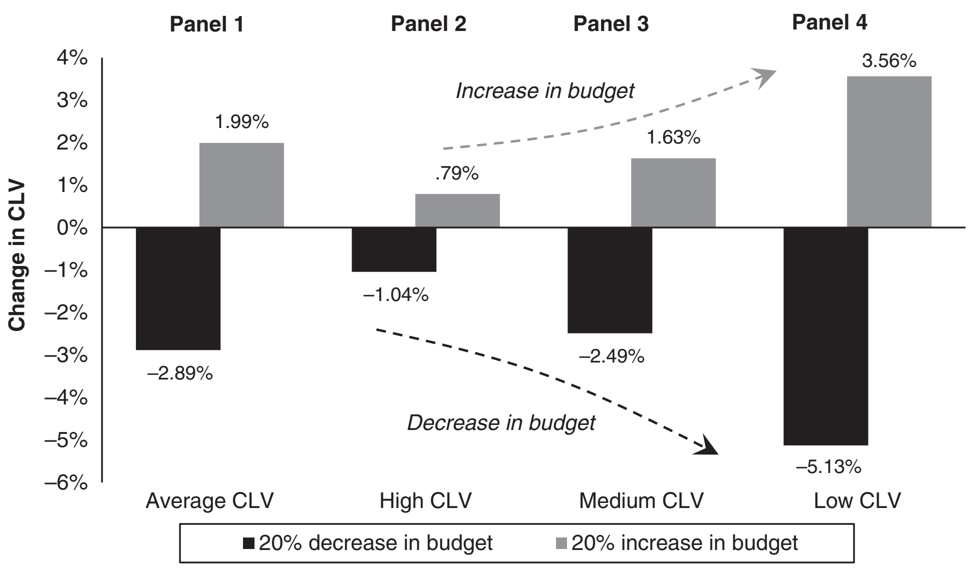
Simulation Exercise 1: Budget Constraints and CLV

In addition to segmenting customers into deciles, we are also interested in studying the relationship between the estimated consumer budget and CLV. High-CLV customers seem to have high budgetary allocations toward the category, and this trend holds for lower-CLV deciles as well. The correlation between CLV and budget is also significant and positive ($\rho = .78$, $p < .001$). That is, a lower CLV is associated with a lower budget constraint and vice versa. However, although the correlation is high, there are certain customers who have a high budget constraint but not a high enough CLV. This may indicate that these customers have a higher potential than their current spending level. These households with high potential could be identified and targeted strategies could be devised to improve their CLV (Venkatesan and Kumar 2004).

Given the relationship between budget constraints and CLV, it is useful to ask the questions, "How do consumers react to budget changes?" and "How does this affect CLV?" In fact, recent experimental research by Carlson et al. (2015) shows that consumers do, in fact, change consumption patterns in the presence of shrinking budgets. Because the proposed modeling framework is structural in nature, we are able to empirically investigate the budget effects on consumers. That is, we conduct theoretically grounded policy experiments in which we vary consumer budget constraints and assess the effects on CLV. Specifically, holding all other effects constant, we attempted to understand the effects of a 20% increase/decrease in budget constraint at the customer level on the customer's CLV. Figure 6 describes the results of the policy simulation.

In Panel 1 of Figure 6, we can see that the percentage change in average CLV for an increase in the budget constraint is less

Figure 6
COUNTERFACTUAL 1: IMPACT OF BUDGET CONSTRAINT ON CLV



than that of a decrease in the budget constraint. This nonlinear (concave) effect of budget constraint on CLV is important for managers to realize because it has implications for understanding consumers' mental accounts for certain categories. Furthermore, from Panels 2–4, we can see that the effects are different depending on the type of customer. Specifically, we can see that high-CLV customers' future profitability is least affected by changes in the budget constraint. However, low-CLV customers tend to be more sensitive to budget constraint changes. We find that low-CLV customers' average spend on carbonated beverages tends to be close to their budget constraints (which tend to be lower), thereby providing further support for this sensitivity to budget changes. Thus, brands that tend to attract low-CLV customers need to be aware of the conditions or situations (e.g., recessionary trends) that could influence the consumer's mental accounting process and, eventually, the budget constraint.

Simulation Exercise 2: Pricing and Consumption

One of the key firm action variables that managers use to improve brand performance in the CPG setting is price. Although we do not specifically estimate a price coefficient in the model, we can easily assess own- and cross-price effects through policy simulations. Furthermore, in our model, the budget constraint parameter acts as a ceiling and helps us identify competition between brands. That is, customers with large budgets are likely to be more price inelastic because when prices increase, they are more likely to absorb the extra cost of consumption than customers who have lower budget constraints. In such a case, the limiting nature of the budget constraint forces customers to reevaluate and adjust their consumption across brands in reaction to a price increase (Howell, Lee, and Allenby 2016). Given this, it is important for managers to assess *which* customers are more/less elastic and *where* the brand switching will occur. If Coca-Cola increases its price, which customers are more likely to purchase other brands and which brands will they consider as close substitutes in this market? Finally, do price increases and decreases lead to symmetric responses among consumers? We attempt to answer these key questions through a policy simulation exercise in which we simulate consumer responses to variations in price.

We generate two scenarios, wherein the focal brand's price either increases or decreases by 10%, with all other covariates and other brand prices held constant. Using the estimated parameters (Θ) along with the new price information, we simulate the consumer's quantity purchases ($\widehat{q_{ijt}}$). We report the findings from this policy simulation in Table 7.

In Table 7, we report the effects of a 10% increase (decrease) in the focal brand's price on the percentage change in average quantity demanded. First, we can see that the direction of the price elasticity is negative for price increases and positive for decreases. However, the magnitude of the effect across brands is not symmetric. The absolute value of the effect of a 10% decrease in price is greater than that of the corresponding increase in price. This nonlinearity in price elasticity is consistent across brands. Second, looking at the magnitude of the own effects (diagonal elements in Table 7), we see that private labels exhibit the highest price elasticity, with Coca-Cola, Pepsi, and Dr. Pepper following. Using this result, managers can assess how CPG customers react to changes in price. Furthermore, looking at the cross-price effects, we find that changes to Coca-Cola prices affect Pepsi and vice versa. This indicates that Pepsi and Coca-Cola are closely competing with each other and price is a key differentiator. This result is further substantiated in model-free analyses (Figure 1).

These results are congruent with prior research in the CPG setting. In the carbonated soda market, Chan (2006)

Table 7
OWN- AND CROSS-EFFECTS OF PRICE

	Price Elasticity: 10% Increase (Decrease) in Price			
Focal Brand ($P_j^{+10\%}$, $P_j^{-10\%}$)	*Coca-Cola*	*Dr. Pepper*	*Pepsi*	*Private Label*
Coca-Cola	−9.11 (10.02)	.68 (−.70)	1.54 (−1.67)	.96 (−.98)
Dr. Pepper	.67 (−.76)	−8.67 (9.42)	.78 (−.85)	1.74 (−1.78)
Pepsi	1.84 (−2.05)	.83 (−.88)	−9.00 (9.70)	1.02 (−1.40)
Private label	.74 (−.78)	.89 (−.96)	.70 (−.71)	−12.45 (12.86)

Notes: All reported values are percentages; percentage changes in quantity for decreases in price are in parentheses.

estimates a continuous choice model similar to the one used in this study and finds that consumers are very price elastic in this market. However, unlike Chan (2006), who finds that elasticities for Pepsi and Coca-Cola are highest, our results suggest that price elasticity is highest for private labels. We also compare these results with the price elasticities from a conventional CLV model and discuss their managerial impact. We find that the directionality of the price elasticities reported in Table 7 is consistent with the conventional model. However, the magnitudes vary depending on the brand. For example, the estimated conventional model suggests that quantity of Coca-Cola demanded would decrease by 4.13% when the price of Coca-Cola increased by 10%. This is in contrast with the 9.11% decrease in demand reported in Table 7. This difference could be important to a CPG manager making pricing decisions that could affect the firm's bottom line. Furthermore, we note that the conventional approach does not account for the nonlinearity in the price effects. That is, an increase in price yields a smaller effect on purchase quantity than a decrease in price. We believe that the nonlinearities in price effects are very important to CPG managers because they have direct implications for pricing decisions (increase vs. decrease) and may eventually affect the firm's bottom line. Ignoring this issue could lead to errant pricing decisions on the part of the CPG firm. Moreover, since the model is estimated at the individual level, managers can use this framework to assess the impact of price changes on various CLV segments (high/medium/low). Due to the presence of the heterogeneous budget constraint (from Equation 3), consumers with higher budget constraints are likely to be less price sensitive than consumers with lower budget constraints.

DISCUSSION AND IMPLICATIONS

Significance of Study Findings

The CLV metric gives firms a long-term, forward-looking, profitability-oriented view of the customer base. However, academic work to date has been relatively silent in applying CLV in the CPG context. We believe that this research addresses some important issues in its attempt to bridge the gap between customer base evaluation (CLV metrics) and the CPG context. If they rely on short-term value metrics (market share, sales, etc.), CPG managers may find it difficult to establish a long-term profitability focus for marketing strategy. We attempt to resolve this issue in this study by proposing a structural approach to modeling the CLV of a CPG customer. We implement and develop insights for our modeling framework using transaction data in the carbonated beverages industry. We next discuss some findings and potential managerial implications of this research.

One of the unique aspects of this study is that in addition to measuring CLV, we also explicitly infer the consumer's budget constraints (through a Bayesian approach) and draw associations between budgetary allocations and CLV. Given our model specification, we are able to measure not only overall CLV for the category but also CLV at the brand level. Managers at CPG firms can make use of this information to understand (1) where their firm stands with regard to future customer profitability and (2) how to move up the profitability ladder (to attract high-CLV customers).

Our assessment of CLV in the carbonated soda market yields some noteworthy results. We find that approximately 40% of the share of the high-CLV customer segment within our sample belongs to Pepsi. Furthermore, from our analysis, it is clear that Coca-Cola customers are not necessarily the most behaviorally loyal. We identify specific past behavior variables that affect the future purchase pattern of the customer and show that these effects are different for different brands. Specifically, we note that depending on the focal brand, the drivers of CLV are different. For example, the effect of customer deal usage intensity is positive for Coca-Cola and Dr. Pepper but negative for private labels. Such outcomes provide very useful information for managers of CPG brands, who can now allocate marketing spend accordingly. Finally, because our model is structural in nature, we are able to conduct theoretically grounded policy simulations (what-if scenarios), a departure from reduced-form modeling approaches that are common in CLV literature. As an illustration, we conduct two managerially relevant counterfactuals to understand the relationship between budget and CLV as well the role of price in this market. In addition, we note that the proposed framework is flexible enough to be estimated at the subbrand (Diet Coke, regular Coke, etc.) and subcategory levels, depending on the managerial need. This flexibility adds to the practical applicability of the proposed framework.

The proposed CLV model can be implemented as a decision framework to assess and manage future value at the individual level or segment level. A natural application of the proposed CLV framework is in cases where manufacturers have access to individual customer-level data and have the logistic capability to market on an individual basis. For example, retailers such as Kroger and Costco that track individual households over time actively use coupons and other instruments to proactively market to individual households. On the other hand, in cases where brands do not have the capability to market to individual households, the proposed model could be applied on readily available scanner panel data (from Nielsen/IRI, etc.) and high-, medium-, and low-CLV segment profiles could be developed. These CLV segment profiles could be used in segment-level targeting strategies that are commonly applied in the CPG industry today. Thus, whether the firm uses segment-level or one-to-one targeting strategies, the proposed CLV framework could be easily adapted and implemented.

Although we do not observe retailer-specific data, we note that the CLV model presented in this study could, in theory, be extended to account for variation in CLV across retailers. To this end, Equation 1 can be easily modified to incorporate variations across retailers, as shown in Equation 13. Specifically, the model could be used to predict $\widehat{q}_{irjt}$ and then summed over R retailers to get the retailer-level CLV:

$$\text{(13)} \qquad CLV_{ir} = \sum_{t=t_1}^{T}\sum_{j=1}^{J}\frac{\widehat{q}_{irjt}\left(m_{rj}P_{rjt}\right)}{(1+d)^{t_0-t}} - \frac{\overline{MC}_{rjt}}{(1+d)^{t_0-t}};$$

$$CLV_i = \sum_{r=1}^{R} CLV_{ir}.$$

Using Equation 13, marketers could assess variations in CLV across retailers and gather consumer insights at the retailer as well as the manufacturer level.

Limitations and Opportunities for Future Research

We believe that this research opens several interesting avenues for further research and has the potential to help CPG firms build strategies to maximize customer-level profits. A possible limitation of the study is the potential heterogeneity in which categorization occurs in the consumer's mind. For example, some consumers might have broad categorizations ("drinks") as opposed to narrow categorizations ("cola"). The formal modeling of this categorization heterogeneity is beyond the scope of this study and could be a topic of future research. An useful avenue to explore could be to expand the analysis to consider a basket of goods such that we could study CLV from a retailer's perspective (provided the data were readily available). Although the framework could be extended to include the retailer's perspective, we are unable to implement this extension due to limited data. We note that the substantive findings for manufacturers are unlikely to change much due to the exclusion of a formal retailer decision model, for many reasons. First, the price paid (which is used in the model) captures the net effect of the retailer and manufacturer decisions, thus mitigating biases in the consumer demand model. Second, the manufacturer (especially in the carbonated soda market) does not heavily discriminate between retailers, such that there is no disproportional benefit to a specific retailer, thus reducing retailer-level differences in demand. This evidence was further supported through our interviews with executives from a large carbonated soda manufacturer. Finally, we attempt to control for some of the retail format–level heterogeneity within the model by using the $STORE\_intensity_{it}$ covariate. Although this does not capture retailer-level loyalty, it gives managers an idea of which formats tend to be preferred for their respective brands. Nevertheless, we acknowledge this as a potential data limitation and leave the investigation of a formal retailer-level model to future research. Due to this limitation, we note that the budget constraint estimated in this study applies mainly to the manufacturer and not the retailer (because consumers might switch retailers). An extension of the proposed model could be the specification of a nested retailer preference model within the demand model to elicit specific demand patterns across retailers. Such a model could be implemented using cooperative databases that track marketing information across retailers and manufacturers (Liu, Pancras, and Houtz 2015).

An worthwhile avenue for future research is the exploration of cross-category effects and the retailer's decision within the CLV framework. As Shankar and Kannan (2014) elaborate, retailers need to know which category needs to be stocked more and when bundling should be marketed by the retailer. A retailer-level CLV model that accounts for cross-category dependencies could be a logical next step in expanding the CLV concept to grocery purchases and could also help managers design profitable pricing strategies. While the proposed model is estimated at the brand level, it is conceivable that one could implement the model on more disaggregated choice sets (such as brand sizes) rather than just brands (Fader and Hardie 1996; Pancras 2011). Within our data, we do not observe enough variation in the consumption patterns across brand-size alternatives, so we are unable to estimate such a model without having to face increased complexities and identification issues. One such issue that could arise within the proposed framework is that there could be correlated unobservables (such as extraneous shocks) that might influence the covariates as well as the consumption patterns.

APPENDIX

The estimation of the proposed model is done efficiently using a hybrid Markov chain Monte Carlo algorithm where (1) the parameters α_{ij}, β_j, and δ_i and their respective hyperparameters are drawn using Gibbs sampling because we can write the full conditionals, and (2) the parameters ζ_{0i} and ζ_1–ζ_4 and their respective hyperparameters are drawn using the M-H algorithm. We cycle through Gibbs and M-H sampling until convergence is achieved. As per the model specification, we have the following parameters that need to be estimated:

$$\text{(A1)} \qquad \begin{aligned} &\alpha_{ij} \sim N\left(\overline{\alpha_j}, V_{\alpha j}\right);\\ &\delta_i \sim N\left(\overline{\delta}, V_{\delta}\right);\\ &\zeta_{0i} \sim N\left(\overline{\zeta_0}, V_{\zeta_0}\right);\\ &\beta_j;\\ &\zeta_1 - \zeta_4. \end{aligned}$$

Step 1: Data Augmentation for ψ_{ijt} and Gibbs Sampling for α_{ij}, $\overline{\alpha_j}$, $V_{\alpha j}$, δ_i, $\overline{\delta}$, V_δ, and β_j

Generate $\psi_{ijt}|\alpha_{ij}, \delta_i, y_i$, and β_j. There are two conditions that would govern the data augmentation of ψ_{ijt}. In the case of an interior solution ($q_{ijt} > 0$), the draw of ψ_{ijt} is done through a probability density function (see Equation 5a), such that

$$\text{(A2a)} \qquad \psi_{ijt}|\alpha_{ij}, \delta_i, y_i, \beta_j \sim N\left[\frac{\lambda_i P_{jt}\left(1+q_{ijt}\right)}{y_{it} - \sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)} - \psi^*_{ijt}, \sigma^2\right],$$

where $\psi^*_{ijt} = \alpha_{ij} + \delta_i SD_{ijt} + \beta_j X_{it}$ as described in Equation 6.

In the case of a corner solution ($q_{ijt} = 0$), the draw of ψ_{ijt} is done through a truncated normal distribution (see Equation 5b), such that

$$\text{(A2b)}\quad \psi_{ijt}|\alpha_{ij},\delta_i,y_i,\beta_j \sim TN\left[\frac{\lambda_i P_{jt}\left(1+q_{ijt}\right)}{y_{it}-\sum_{j=1}^{J}\left(P_{jt}q_{ijt}\right)}-\psi^{*}_{ijt},\sigma^2\right].$$

Generate α_{ij}, δ_i, $\beta_j|\psi_{ijt}$, $\overline{\alpha_j}$, V_α, $\overline{\delta}$, *and* V_δ. The draw of ψ_{ijt} converts Equation 6 into a standard multivariate regression model with heterogeneity. We can estimate the parameters listed in Equation A1 using Gibbs sampling because the full conditionals can be derived. Specifically, we draw the following densities:

$$\text{(A3)}\quad \alpha_{ij}|\overline{\alpha_j}, V_{\alpha j}, \psi_{ijt}, \delta_i, \beta_j;$$

$$\text{(A4)}\quad \delta_i|\overline{\delta}, V_\delta, \psi_{ijt}, \alpha_{ij}, \beta_j;$$

$$\text{(A5)}\quad \beta_j|\psi_{ijt}, \alpha_{ij}, \delta_i;$$

$$\text{(A6)}\quad \overline{\alpha_j}|\alpha_{ij}, V_{\alpha j};$$

$$\text{(A7)}\quad V_{\alpha j}|\alpha_{ij}, \overline{\alpha_j};$$

$$\text{(A8)}\quad \overline{\delta}|\delta_i, V_\delta;$$

$$\text{(A9)}\quad V_\delta|\delta_i, \overline{\delta}.$$

Step 2: M-H Algorithm for y_i*,* ζ_1*–*ζ_4 *Draws*

Because we do not have closed-form expressions for the posterior probability distributions of y_i and ζ_1–ζ_4, we need to use the M-H algorithm with random walk for estimation. From Equation 9,

$$\text{(A10)}\quad L_i(\Theta) = \int_{-\infty}^{\infty}\prod_{t=1}^{T}\prod_{j=0}^{J}\left[\phi\left(\varepsilon_{ijt}\right)\times |J|_{\varepsilon_{ijt}\rightarrow q_{ijt}}\right]^{I\left(q_{ijt}>0\right)} \times \Phi\left(\varepsilon_{ijt}\right)^{\left[1-I\left(q_{ijt}>0\right)\right]} f(\Theta_i)d\Theta_i.$$

Let $\zeta_{0i}^{(m)}$ denote the mth draw for ζ_{0i}. The next draw (m + 1) is given by

$$\text{(A11)}\quad \zeta_{0i}^{(m+1)} = \zeta_{0i}^{(m)} + \xi_{\zeta_0},$$

where ξ_y is a draw from the candidate generating density (normal distribution).

The probability of accepting the new draw ($\zeta_{0i}^{(m+1)}$) is given by

$$\text{(A12)}\quad \min\left(\frac{\exp\left[-\frac{1}{2}\left(\zeta_{0i}^{(m+1)}-\overline{\zeta_0}\right)' V_y^{-1}\left(\zeta_{0i}^{(m+1)}-\overline{\zeta_0}\right)\right]\times L(\Theta_i)^{(m+1)}}{\exp\left[-\frac{1}{2}\left(\zeta_{0i}^{(m)}-\overline{\zeta_0}\right)' V_y^{-1}\left(\zeta_{0i}^{(m)}-\overline{\zeta_0}\right)\right]\times L(\Theta_i)^{(m)}},1\right).$$

If the new draw is rejected, then $\zeta_{0i}^{(m+1)} = \zeta_{0i}^{(m)}$. Using the drawn ζ_{0i} values, we can easily draw $\overline{\zeta_0}$ and V_{ζ_0} using Gibbs sampling, similar to the procedure described in Step 1. This procedure of generating the parameter using M-H algorithm is repeated for the ζ_1–ζ_4 parameters as well. Once this step is over, we iterate again over the densities drawn in Step 1 and then repeat this process until convergence is met.

REFERENCES

Ailawadi, Kusum L., Scott A. Neslin, and Karen Gedenk (2001), "Pursuing the Value-Conscious Consumer: Store Brands Versus National Brand Promotions," *Journal of Marketing*, 65 (January), 71–89.

———, Koen Pauwels, and Jan-Benedict E.M. Steenkamp (2008), "Private-Label Use and Store Loyalty," *Journal of Marketing*, 72 (November), 19–30.

Allender, William J., Timothy J. Richards, Sungho Park, and Stephen F. Hamilton (2013), "Demand for Variety Under Costly Consumer Search: A Multiple-Discrete/Continuous Approach," working paper, [available at http://dx.doi.org/10.2139/ssrn.2309136].

Antonides, Gerrit, I. Manon de Groot, and W. Fred van Raaij (2011), "Mental Budgeting and the Management of Household Finance," *Journal of Economic Psychology*, 32 (4), 546–55.

Assael, Henry (1974), "Product Classification and the Theory of Consumer Behavior," *Journal of the Academy of Marketing Science*, 2 (4), 539–52.

Baxley, John V. and John C. Moorhouse (1984), "Lagrange Multiplier Problems in Economics," *American Mathematical Monthly*, 91 (7), 404–12.

Bhat, Chandra R. (2008), "The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions," *Transportation Research Part B: Methodological*, 42 (3), 274–303.

Carlson, Kurt A., Jared Wolfe, Simon J. Blanchard, Joel C. Huber, and Dan Ariely (2015), "The Budget Contraction Effect: How Contracting Budgets Lead to Less Varied Choice," *Journal of Marketing Research*, 52 (June), 337–48.

Chan, Tat Y. (2006), "Estimating a Continuous Hedonic-Choice Model with an Application to Demand for Soft Drinks," *Rand Journal of Economics*, 37 (2), 466–82.

Cheema, Amar and Dilip Soman (2006), "Malleable Mental Accounting: The Effect of Flexibility on the Justification of Attractive Spending and Consumption Decisions," *Journal of Consumer Psychology*, 16 (1), 33–44.

Chintagunta, Pradeep K. and Sudeep Haldar (1998), "Investigating Purchase Timing Behavior in Two Related Product Categories," *Journal of Marketing Research*, 35 (February), 43–53.

——— and Harikesh S. Nair (2011), "Structural Workshop Paper: Discrete-Choice Models of Consumer Demand in Marketing," *Marketing Science*, 30 (6), 977–96.

Du, Rex Y. and Wagner A. Kamakura (2008), "Where Did All That Money Go? Understanding How Consumers Allocate Their Consumption Budget," *Journal of Marketing*, 72 (November), 109–31.

Dubé, Jean-Pierre (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 23 (1), 66–81.

———, Günter J. Hitsch, and Peter E. Rossi (2010), "State Dependence and Alternative Explanations for Consumer Inertia," *Rand Journal of Economics*, 41 (3), 417–45.

Fader, Peter S. and Bruce G.S. Hardie (1996), "Modeling Consumer Choice Among SKUs," *Journal of Marketing Research*, 33 (November), 442–52.

Gilboa, Itzhak, Andrew Postlewaite, and David Schmeidler (2010), "The Complexity of the Consumer Problem and Mental Accounting," working paper, Tel Aviv University.

Gordon, Brett R., Avi Goldfarb, and Yang Li (2013), "Does Price Elasticity Vary with Economic Growth? A Cross-Category Analysis," *Journal of Marketing Research*, 50 (February), 4–23.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V. Kumar, Nathaniel Lin, et al. (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139–55.

——— and Donald R. Lehmann (2005), *Managing Customers as Investments: The Strategic Value of Customers in the Long Run*. Upper Saddle River, NJ: Wharton School Publishing.

Heath, Chip and Jack B. Soll (1996), "Mental Budgeting and Consumer Decisions," *Journal of Consumer Research*, 23 (1), 40–52.

Hendel, Igal (1999), "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 66 (2), 423–46.

Howell, John R., Sanghak Lee, and Greg M. Allenby (2016), "Price Promotions in Choice Models," *Marketing Science*, forthcoming [DOI: 10.1287/mksc.2015.0948].

Jain, Dipak C. and Naufel J. Vilcassim (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, 10 (1), 1–23.

Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi (2002), "Modeling Consumer Demand for Variety," *Marketing Science*, 21 (3), 229–50.

———, ———, and ——— (2007), "Product Attributes and Models of Multiple Discreteness," *Journal of Econometrics*, 138 (1), 208–30.

Kumar, V. and Anita (Man) Luo (2008), "Integrating Purchase Timing, Choice, and Quantity Decisions Models: A Review of Model Specifications, Estimations, and Applications," in *Review of Marketing Research*, Vol. 4, Naresh K. Malhotra, ed. Bingley, UK: Emerald, 63–91.

——— and Denish Shah (2009), "Expanding the Role of Marketing: From Customer Equity to Market Capitalization," *Journal of Marketing*, 73 (November), 119–36.

Lewis, Michael (2004), "The Influence of Loyalty Programs and Short-Term Promotions on Customer Retention," *Journal of Marketing Research*, 41 (August), 281–92.

Liu, Hongju, Joseph Pancras, and Malcolm Houtz (2015), "Managing Customer Acquisition Risk Using Co-operative Databases," *Journal of Interactive Marketing*, 29 (February), 39–56.

Ma, Yu, Kusum L. Ailawadi, and Dhruv Grewal (2013), "Soda Versus Cereal and Sugar Versus Fat: Drivers of Healthful Food Intake and the Impact of Diabetes Diagnosis," *Journal of Marketing*, 77 (May), 101–20.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta (1999), "The 'Shopping Basket': A Model for Multicategory Purchase Incidence Decisions," *Marketing Science*, 18 (2), 95–114.

Milkman, Katherine L. and John Beshears (2009), "Mental Accounting and Small Windfalls: Evidence from an Online Grocer," *Journal of Economic Behavior & Organization*, 71 (2), 384–94.

Pancras, Joseph (2011), "The Nested Consideration Model: Investigating Dynamic Store Consideration Sets and Store Competition," *European Journal of Operational Research*, 214 (2), 340–47.

Prelec, Drazen and George Loewenstein (1998), "The Red and the Black: Mental Accounting of Savings and Debt," *Marketing Science*, 17 (1), 4–28.

Ratneshwar, S., Cornelia Pechmann, and Allan D. Shocker (1996), "Goal-Derived Categories and the Antecedents of Across-Category Consideration," *Journal of Consumer Research*, 23 (3), 240–50.

Richards, Timothy J., Miguel I. Gómez, and Geoffrey Pofahl (2012), "A Multiple-Discrete/Continuous Model of Price Promotion," *Journal of Retailing*, 88 (2), 206–25.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (January), 109–27.

Satomura, Takuya, Jaehwan Kim, and Greg M. Allenby (2011), "Multiple-Constraint Choice Models with Corner and Interior Solutions," *Marketing Science*, 30 (3), 481–90.

Seetharaman, P.B. (2004), "Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach," *Marketing Science*, 23 (2), 263–71.

———, Andrew Ainslie, and Pradeep K. Chintagunta (1999), "Investigating Household State Dependence Effects Across Categories," *Journal of Marketing Research*, 36 (November), 488–500.

——— and Pradeep Chintagunta (1998), "A Model of Inertia and Variety-Seeking with Marketing Variables," *International Journal of Research in Marketing*, 15 (1), 1–17.

Shah, Denish, V. Kumar, and Kihyun Hannah Kim (2014), "Managing Customer Profits: The Power of Habits," *Journal of Marketing Research*, 51 (December), 726–41.

Shankar, Venkatesh and P.K. Kannan (2014), "An Across-Store Analysis of Intrinsic and Extrinsic Cross-Category Effects," *Customer Needs and Solutions*, 1 (2), 143–53.

Stilley, Karen M., J. Jeffrey Inman, and Kirk L. Wakefield (2010), "Planning to Make Unplanned Purchases? The Role of In-Store Slack in Budget Deviation," *Journal of Consumer Research*, 37 (2), 264–78.

Thaler, Richard (1985), "Mental Accounting and Consumer Choice," *Marketing Science*, 4 (3), 199–214.

Van Oest, Rutger (2005), "Which Brands Gain Share from Which Brands? Inference from Store-Level Scanner Data," *Quantitative Marketing and Economics*, 3 (3), 281–304.

Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (October), 106–25.

———, ———, and Timothy Bohling (2007), "Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection," *Journal of Marketing Research*, 44 (November), 579–94.

Yoo, Shijin, Dominique M. Hanssens, and Ho Kim (2011), "Marketing and the Evolution of Customer Equity of Frequently Purchased Brands," working paper, Korea University Business School.