
Management Science

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

# An Interproduct Competition Model Incorporating Branding Hierarchy and Product Similarities Using Store-Level Data

Sudhir Voleti, Praveen K. Kopalle, Pulak Ghosh

MANAGEMENT SCIENCE
Vol. 61, No. 11, November 2015, pp. 2720–2738
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

informs®
http://dx.doi.org/10.1287/mnsc.2014.2039


# An Interproduct Competition Model Incorporating Branding Hierarchy and Product Similarities Using Store-Level Data

Sudhir Voleti
Indian School of Business, Gachibowli, Hyderabad 50032, India, sudhir.voleti@gmail.com

Praveen K. Kopalle
Tuck School of Business at Dartmouth, Dartmouth College, Hanover, New Hampshire 03755, kopalle@dartmouth.edu

Pulak Ghosh
Indian Institute of Management, Bangalore, Bengaluru 560076, India, pulakghoshc@gmail.com

We develop and implement a Bayesian semiparametric model of demand under interproduct competition that enables us to assess the respective contributions of brand-SKU (stock keeping unit) hierarchy and interproduct similarity to explaining and predicting demand. To incorporate brand-SKU hierarchy effects, we use Bayesian hierarchical clustering inherent in a nested Dirichlet process to simultaneously partition brands, and SKUs conditional on brands, into groups of "similarity clusters." We examine cluster memberships and postprocess the Markov chain Monte Carlo output to infer cluster properties by accounting for parameter uncertainty. Our proposed approach lends to a spatial competition interpretation in latent attribute space and helps uncover the extent to which competition across SKUs in the latent attribute space is local or global. In a related vein, we discuss the implications of well-defined groups of similar SKUs as subcategory or submarket boundaries in latent attribute space. We empirically test our model using aggregate beer category sales data from a midsize U.S. retail chain. We find that branding hierarchy effects dominate those from product similarity. We find that the model partitions the 15 brands in the data into 4 brand clusters and the 96 SKUs into 25 SKU clusters conditional on brand cluster membership. In estimating a set of models of spatial interproduct competition, we find that SKU competition is more local than global in that only subsets of products compete within groups of comparable products. Finally, we discuss the substantive implications of our results.

*Keywords*: competition; nested Dirchlet process; brand-SKU clustering; product similarity; spatial models
*History*: Received June 29, 2012; accepted March 6, 2014, by Pradeep Chintagunta, marketing. Published online in *Articles in Advance* February 19, 2015.



## 1. Introduction

Competitive effects are integral to market outcomes (e.g., unit sales), particularly in the consumer packaged goods (CPG) industry where there is abundant brand and stock keeping unit (SKU) proliferation. Although prescriptive models have significantly enhanced our conceptual understanding of how firms should respond to competitive actions, at a more fundamental level, it is equally important to more fully capture how competitive actions impact sales. Accordingly, there has been a lot of interest around modeling competition effects. Consider the following hypothetical example from the beer category. Let us assume from a market response standpoint that the competitive effects of SKUs between brands "Budweiser" and "Coors" may be more intense than, say, between Budweiser and "Heineken." This is because either customers view Budweiser and Coors to be more similar in observed attributes (such as price, brand, and SKU characteristics) or because the two may be more similar in latent attributes (such as quality, brand positioning, the response to marketing mix elements etc.). In other words, the net competitive impact of Coors' SKU on a Budweiser's SKU is based on the following three factors: (i) the extent to which Budweiser and Coors are (perceived) "similar" to each other and Heineken is not, which leads to "similarity clusters" among brands; (ii) the extent to which SKUs among brand clusters are similar or dissimilar to one another, which leads to SKU level clustering conditioned upon brand level clusters; (iii) the product attributes and marketing mix variables of the various SKUs both within and outside a brand cluster. This simple example illustrates that demand models of competition need to incorporate both (i) brand and SKU level hierarchies and (ii) the corresponding similarity measures across SKUs to more appropriately capture competitive effects.

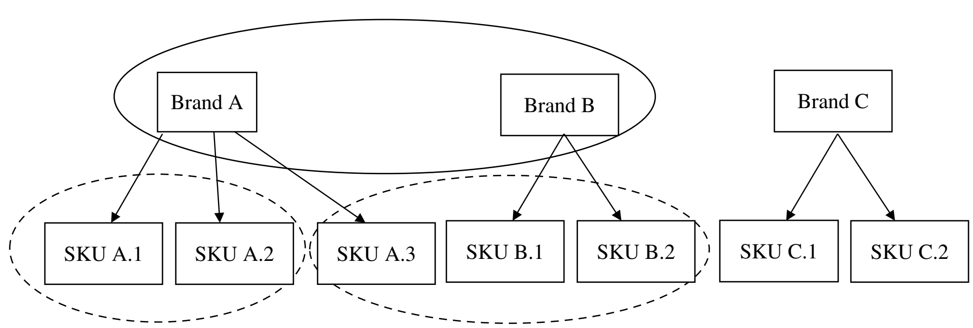
In general, CPG products typically compete as part of hierarchical branding structures involving a brand layer at the top and another SKU layer at the bottom. However, most models estimate interproduct competition effects at either the brand level (e.g., Ailawadi et al. 2005, Elrod and Keane 1995, Chintagunta et al. 1991) or the SKU level (e.g., Chintagunta 1994, Bronnenberg and Vanhonacker 1996). Thus, the literature has largely taken a "single-level" perspective on interproduct competition. Models of brand-level competition forgo the richness and explanatory power of the specific product attributes present at the SKU level (Fader and Hardie 1996). On the other hand, models of SKU level competition treat the brand as just another SKU attribute. Both empirical and experimental studies document that the brand is more than just another product attribute in that it impacts consumers' evaluation of all the other product attributes (e.g., Sullivan 1998). Further, the literature on market structure (e.g., Kannan and Wright 1991a, b), consideration set formation (e.g., Shocker et al. 1991), and sequential choice models (Nedungadi 1990), suggests that consumers process hierarchical alternative sets (in this case, brand information) and choose from the alternatives (SKUs) that belong to the brands in the set. This suggests that models of competition take a "multilevel" perspective (brand, then SKU) of interproduct competition.

Competition between two SKUs (or brands) arises as a consequence of product substitutability, which in turn depends on how similar consumers perceive the products to be. For instance, one stream of literature (Yang et al. 2002, Kim and Chintagunta 2012) finds that apart from intrinsic brand preference, consumption occasions, motivation, and contexts provide additional sources of brand preference heterogeneity across consumers. Thus, similarity across SKUs or brands may be along multiple dimensions, such as observable attributes (brand, price etc.) and latent attributes (such as perceived quality, price elasticity, response to promotions, consumption context etc.). To cater to a diverse set of consumer preferences over products, firms (manufacturers or retailers) deploy portfolios of brands and SKUs in most CPG categories and use marketing-mix-based tools to influence consumers' purchase decisions. Firms are likely to be interested in leveraging economies of scale and/or scope in their use of marketing-mix tools to improve profitability. This would require that firms be able to identify precisely which sets of products (brands and subsequently, SKUs) are "similar" to one another in marketing-mix response terms. Since CPG products in grocery stores have a discernible hierarchy in their branding structure (in that brands deploy SKUs), the problem of identifying sets of products similar in their marketing mix responses also assumes a hierarchical form. We illustrate this desired dependency structure among similarity clusters (and the complications therein) using Figure 1, where A, B, and C denote brands with SKUs under each brand.

Suppose brands A and B are in the same cluster (the solid oval). Now, if we apply any standard clustering procedure, we may find SKUs A.1 and C.2 in the same cluster, whereas we know that brands A and C are dissimilar. Ideally, we would want the clustering to be restricted only to SKUs of brands A and B (the dashed oval in the figure) since we know the concerned brands are more similar. We thus seek to achieve a multilevel dependency structure preferably within a nonparametric framework (so that we can avoid making parametric assumptions on outcome distributions for the products to be clustered). To this end, we propose the nested Dirichlet process model (Rodríguez et al. 2008) that accounts for the brand-SKU hierarchy by grouping together similar brands at an upper "parent" level and, *simultaneously*, at a "child" level, where similar SKUs are grouped together within an umbrella of brand clusters.

In typical demand models, cross effects (e.g., cross-price elasticities) are normally used to capture the effect of competition (e.g., Leeflang et al. 2000). Whereas such an approach works well with a small number of SKUs, parameter proliferation erodes degrees of freedom when a large number of SKUs are present in the data (e.g., Sudhir 2001). On the other hand, if perceived similarity between SKUs

**Figure 1** Illustration of the Hierarchical Product Aggregation Structure in CPG Categories



is known a priori, then within the popular logit formulation, one can use the generalized extreme value-based nested logit model (e.g., Kannan and Wright 1991a, b) to construct similarity groupings of SKUs that compete more with one another than with SKUs outside the grouping. In practice, however, perceived SKU similarity is rarely known ex ante for most product categories. Although specific hypotheses about similar-SKU groupings can be tested using the nested logit, even with a modest number of SKUs (say, $n$), empirically testing for the right set of similar SKU groupings becomes computationally challenging as there would be $2^n - 1$ possible SKU groupings. Hence, there is value in a method that empirically and endogenously identifies SKU groupings such that SKUs are more similar within than outside it; and that is consistent with the brand memberships of SKUs in any cluster, i.e., similar SKUs from a SKU-level cluster belong to similar brands from a brand-level cluster.

Finally, we note that data on perceived interproduct (SKU or brand) similarities can either be directly collected from consumers as primary data (e.g., Horsky and Nelson 1992) or derived from secondary data on sales and marketing-mix activities. However, issues such as cost, interpretability, the concordance between stated and revealed preference (e.g., Park and Srinivasan 1994) and variation in perceptions across respondents are liable to arise in the case of primary data.

We address the following three key research questions in this paper. (1) What is the role and relative value of incorporating interproduct similarities and brand-SKU hierarchy in demand models? (2) Which underlying similarity-based brand and SKU clusters (conditional on brand memberships) are inferred from our analysis? A related aspect is accounting for uncertainty in cluster-specific parameters to enable inference on the clusters obtained. (3) What is the extent to which competition in latent attribute space is local (only a subset of products compete with a focal product) or global (every product competes with every other product)? Correspondingly, the contribution of this paper is threefold: First, we develop and implement a model of demand with interproduct competition that simultaneously incorporates branding hierarchy effects and interproduct similarity across all the products in the sample. Our semiparametric model enables us to assess the respective value of branding hierarchy and interproduct similarity in explaining and predicting demand. Second, we leverage Bayesian hierarchical clustering inherent in our nested Dirichlet process model to partition brands and SKUs (conditional on brands) into groups of "similarity clusters," examine cluster memberships, and infer cluster properties after accounting for parameter uncertainty. Third, our proposed approach leads to a spatial interpretation of competition in latent attribute space and helps uncover the extent to which competition across SKUs in the latent attribute space is local or global. In a related vein, we discuss the implications of well-defined groups of similar SKUs as subcategory or submarket boundaries in latent attribute space. We empirically test our model using store-level, beer category sales data from a retail chain and derive corresponding managerial implications.

## 2. Model of Interproduct Competition

We propose a two-stage model of interproduct competition that simultaneously incorporates brand-SKU hierarchy and product similarities. In Stage 1, we obtain an initial set of estimates for a basic demand model and use the estimates to construct a base level of competition variable that incorporates both interproduct similarity in observed attribute space and the impact of marketing mix variables. In Stage 2, using an initialization sample we deploy the brand-SKU hierarchy, via a nested Dirichlet process (nDP), and develop an interproduct similarity measure in latent attribute space. The new similarity measure is used to revise the competition variable as a latent distance-based measure and the demand model specification is then reestimated in Stage 2 using a calibration sample and the nDP approach. Finally, we use the Stage 2 results to compare our model against alternative (benchmark) models, address our research questions, and bring out the substantive implications.

### 2.1. Stage 1 Model

We begin the first stage with a typical and popular log-log model of demand (e.g., Hoch et al. 1995, Wedel and Zhang 2004) in which we regress the log of sales quantity over a set of demand determinants, and account for SKU-specific random effects. The logarithmic function offers a natural diminishing-returns pattern, accommodation of various response shapes and rates, and a ready interpretation of coefficients as elasticities. Let $j = 1, 2, \ldots, J$ index SKU, $t = 1, 2, \ldots, T$ time-period and $b = 1, 2, \ldots, B$ index brand. Note the subscript $j(b)$ denote SKU $j$ belonging to brand $b$. Accordingly, our Stage 1 demand specification is

$$\begin{aligned}&\ln(\mathrm{SALES}_{jt})\\&\quad=\boldsymbol{\alpha}_1\mathbf{Z}_t+\alpha_{2,j}\mathrm{COMPTN}_{jt}^{(\mathrm{OLS})}+\boldsymbol{\beta}_j\ln(\mathbf{MMIX}_{jt}^{(\mathrm{P})})\\&\qquad+r_j(b)+\varepsilon_{jt},\quad \varepsilon_{jt}\sim\mathrm{IIDN}(0,\sigma_\varepsilon^2),\quad j=1,2,\ldots,J;\\&\qquad\qquad t=1,2,\ldots,T.\qquad(1)\end{aligned}$$

Here, $\mathbf{Z}_t$ is a vector of control variables (e.g., seasonality; and log is taken for continuous variables) and $\mathrm{COMPTN}_{jt}$ is a scalar term summarizing the competitive effects on SKU $j$ due to all other SKUs in period $t$ and $\mathbf{MMIX}_{jt}^{(\mathrm{P})}$ is a vector of SKU-specific heterogeneous price, promotion (or advertising), and distribution variables corrected for endogeneity.[1] Correspondingly, $\alpha_1$ is a vector of coefficients, $\alpha_{2,j}$ the sales impact of competition on SKU $j$, and $\boldsymbol{\beta}_j$ a vector of heterogeneous marketing mix (henceforth, MMIX) elasticities specific to SKU $j$. Since the log-log model corresponds to a multiplicative model, to avoid a situation in which a product with zero promotion (say) ends up with zero sales, we set the minimum promotion level for any SKU in time $t$ to a small number. Measurement error $\varepsilon_{jt}$is assumed to be white noise and drawn independently and identically distributed (IID) from a mean zero Normal. There are no brand effects modeled separately in Equation (1). All time-invariant product attributes (both observed and latent) are captured directly by the SKU-specific random effects term $r_{j(b)}$. To construct a parsimonious interproduct competition metric, $\mathrm{COMPTN}_{jt}$, we weigh the MMIX by interproduct similarity for every product pair of interest. Evidence in the marketing literature suggests that the competitive effect of a rival product on a focal product's demand directly relates to (i) how similar the products are (or are perceived to be) by consumers (e.g., see Sethuraman 2003 on positioning private labels against national brands on attributes), and (ii) the MMIX activity of the rival product (e.g., Eliashberg and Chatterjee 1985) and, especially, its price (Russell and Kamakura 1993). In this vein, we posit that the competitive effect of rival item $i$ on focal item $j$ in period $t$, termed $\mathrm{COMPTN}(jt, it)$, arises when the degree of "similarity" between $i$ and $j$ (denoted by $\mathrm{SIMIL}_{ji}$), scales up or down the competitive impact of the marketing mix of rival $i$ in period $t$ (denoted by $\mathrm{MMIX.EFFECT}_{it}$); thus,

$$\mathrm{COMPTN}(jt, it) = f_1(\mathrm{SIMIL}_{ji}) \times f_2(\mathrm{MMIX.EFFECT}_{it}). \tag{2}$$

Since only the rival's MMIX finds place in the competition expression, unless two products have identical contemporaneous MMIX levels, competition between two products will be asymmetric in that $\mathrm{COMPTN}(jt, it) \neq \mathrm{COMPTN}(it, jt)$. This is consistent with the findings of Kamakura and Russell (1989).

Let $D$ be the dimension of the number of observable, time-invariant, and binary product attributes in the data. To operationalize $f_1(.)$ and $f_2(.)$ in Equation (2), we first run an ordinary least squares (OLS) log-log regression of sales, from an *external* sample (not part of the initialization, calibration, or holdout samples) over observed binary attributes and MMIX variables. Let $\boldsymbol{\beta}_{\mathrm{ATTRIB}}^{(\mathrm{OLS})}$ and $\boldsymbol{\beta}_{\mathrm{MMIX}}^{(\mathrm{OLS})}$ be the $D \times 1$ and $3 \times 1$ vectors of coefficients, respectively. We operationalize the degree of similarity between any two SKUs $i$ and $j$ as follows:

$$\mathrm{SIMIL}_{ji}^{(\mathrm{OLS})} = \sum_{\mathrm{d}=1}^{\mathrm{D}} \beta_{d,\,\mathrm{ATTRIB}}^{(\mathrm{OLS})} \times I(i \text{ and } j \text{ share attribute } d). \tag{3}$$

Here, $I(.)$ is the indicator function and takes the value 1 if both SKUs share a common attribute. The coefficients in $\boldsymbol{\beta}_{\mathrm{ATTRIB}}^{(\mathrm{OLS})}$ represent the attributes' impact on sales. Thus, SIMIL measures how similar $i$ and $j$ are weighted by the relative importance to sales of shared attributes. Using a similar rationale for MMIX variables, we rewrite Equation (2) as

$$\begin{aligned}\mathrm{COMPTN}(jt, it)^{(\mathrm{OLS})} &= \mathrm{SIMIL}_{ji}^{(\mathrm{OLS})} \times \left(\sum_{l=1}^{3} \beta_{l,\,\mathrm{MMIX}}^{(\mathrm{OLS})} \ln\left(\mathrm{MMIX}_l^{(P)}\right)_{it}\right)\\ &= \mathrm{SIMIL}_{ji}^{(\mathrm{OLS})} \times \ln(\mathbf{MMIX}_{it}^{(\mathbf{P})})\boldsymbol{\beta}_{\mathrm{MMIX}}^{(\mathrm{OLS})}.\end{aligned} \tag{4}$$

The OLS superscript in Equations (3) and (4) distinguish the base level SIMIL and COMPTN variables from those that are derived using an nDP specification later in Stage 2. Finally, we aggregate the competitive effect of all products $i \neq j$ on focal product $j$ in $t$ as

$$\begin{aligned}\mathrm{COMPTN}_{jt} &= \mathrm{COMPTN}(jt, \cdot)\\ &= \sum_{i \neq j} \mathrm{COMPTN}(jt, it).\end{aligned} \tag{5}$$

Let the vector of SKU specific response elasticities in Equation (1) be denoted by $\theta_{j(b)}$ and be drawn from some general multivariate density $F_{\mathrm{MV}}(.)$; thus,

$$\boldsymbol{\theta}_j(b) = [\boldsymbol{\alpha}_{2,j} \boldsymbol{\beta}_j^T r_j(b)]^T \sim F_{\mathrm{MV}}(.). \tag{6}$$

### 2.2. Stage 2 Model

Different specifications for $F_{\mathrm{MV}}(.)$ in Equation (6) yield different demand models. For instance, one common specification would be $F_{\mathrm{MV}}(.) \equiv \mathrm{MVN}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$. This would imply that $\theta_{j(b)}$ are drawn from a multivariate normal density. However, a strictly parametric specification for $F_{\mathrm{MV}}$ may be restrictive because $\theta_{j(b)}$'s true distribution might be skewed, multimodal, or otherwise nonstandard in shape and behavior. To retain flexibility and for robustness against particular distributional assumptions, Bayesian semiparametric and nonparametric models such as those based on

[1] Some variables such as the time-varying MMIX elements may be strategically set by manufacturers and retailers. We treat these potentially endogenous variables using a set of instruments. We use only their two-stage least squares (2SLS) projection onto a corresponding set $\mathbf{M}_l$ of exogenous or predetermined variables (detailed in the data section) to obtain the corresponding endogeneity corrected values $\mathrm{MMIX}_l^{(\mathrm{P})}$; thus, $\mathbf{MMIX}_l^{(\mathrm{P})} = \mathbf{M}_l(\mathbf{M}_l^T\mathbf{M}_l)\mathbf{M}_l^T\mathbf{MMIX}_l$, $l = 1, 2, 3$.

a Dirichlet process prior have been employed in the past (e.g., Ansari and Mela 2003, Braun and Bonfrer 2011, and Wedel and Zhang 2004). We adapt a hierarchical generalization of the standard Dirichlet process (DP) prior model called the nested Dirichlet process model to accommodate constraints implied by the branding hierarchy and to implement the required dependence structure as shown in Figure 1. Although we briefly describe the standard DP model (and later the nDP), for a thorough review of DP, see Hjort et al. (2010).

**2.2.1. Standard Dirichlet Process (DP) Model.** In the standard DP model, $\boldsymbol{\theta}_{j(b)}$are drawn from a general distribution (say, $\mathbf{G}$) whose functional form is ex ante unknown. Hence, $\mathbf{G}$ is treated as a set of parameters with a prior denoted by DP($\alpha$,$\mathbf{G}_0$). Conditional on $\mathbf{G}$, $\boldsymbol{\theta}_{j(b)}$ is drawn IID from $\mathbf{G}$:

$$\boldsymbol{\theta}_j \mid \mathbf{G} \overset{\text{IID}}{\sim} \mathbf{G}, \quad \mathbf{G} \sim \mathrm{DP}(\alpha, \mathbf{G}_0). \tag{7}$$

Since the distribution $\mathbf{G}$ is drawn from DP($\alpha$,$\mathbf{G}_0$), the DP prior is essentially a distribution on the space of distributions. It is parameterized by a base distribution $\mathbf{G}_0$ and by a positive concentration parameter $\alpha$ that measures variability around $\mathbf{G}_0$. Of particular interest is the fact any realization of $\mathbf{G}$ is almost surely discrete (Ferguson 1973). The discreteness property of the DP ensures that the $\theta_{j(b)}$ draws are discrete such that groups of similar products would share an identical $\theta$ value. While estimating DP, suppose $L$ groups of similar products emerge, each with parameter value $\theta_l$ for $l = 1, 2, \ldots, L$. If products $i$ and $j$ belong to the same group $l$, then the model implies that $\boldsymbol{\theta}_i = \boldsymbol{\theta}_j = \boldsymbol{\theta}_l$ for that realization of $\mathbf{G}$. In reality, in the MCMC (Markov chain Monte Carlo) algorithm, different realizations of $\mathbf{G}$ occur. Each draw of $\mathbf{G}$ and the number of groups that emerge can be viewed as a model selection problem (Hjort et al. 2010, Chap. 7). Correspondingly, $\theta_i$ values for any product $i$ are the model averaged values across different draws of $\mathbf{G}$. Thus, when $F_{\mathrm{MV}}(.) \equiv \mathrm{DP}(\alpha,\mathbf{G}_0)$, the model output yields not just estimates of the random coefficients vector but also a substantively interesting grouping of SKUs into "similar-SKU" clusters. An important advantage of the DP prior is that the appropriate number of mixture components is determined from within the model by the prior and the data.

**2.2.2. Nested (Multilevel) Dirichlet Process Model.** The nDP is obtained by replacing the $\mathbf{G}_0$ in Equation (7) with another DP, say DP($\rho$, $\mathbf{H}$). Thus, we *nest* a new "child" DP within an existing "parent" DP. Following Ho et al. (2013), the resulting specification is written as nDP($\alpha$, $\rho$, $\mathbf{H}$). In our application, let $\mathbf{F}_{b(j)}(.)$ denote the unit sales distribution of SKUs $j = 1, 2, \ldots, J_b$ under brand $b$. Let $\boldsymbol{\theta}_{j(b)}$ denote the random parameter vector for SKU $j$ of brand $b$. Let a set of distributions to model $\mathbf{F}_b$ and $\theta_{j(b)}$ be distributed according to an nDP; thus,

$$\mathbf{F}_{b(j)}(.) \mid \{\pi_k, \mathbf{G}_k\} \overset{\text{IID}}{\sim} \mathbf{Q}, \quad \mathbf{Q} = \sum_{k=1}^{\infty} \pi_k \delta_{G_k}(.),$$
$$b = 1, 2, \ldots, B;\ j = 1, \ldots, J_b;\ k \in \{1, \ldots, \infty\}, \tag{8}$$

and

$$\mathbf{G}_k(.) \mid \{\omega_l^{(k)}, \mathbf{a}_l^{(k)}\} \equiv \sum_{l=1}^{\infty} \omega_l^{(k)} \delta_{a_l}^{(k)}(.), \quad \mathbf{a}_l^{(k)} \overset{\text{IID}}{\sim} \mathbf{H},$$
$$l \in \{1, \ldots, \infty\}. \tag{9}$$

Here, $\pi_k$ and $\omega_l^{(k)}$ represent Dirichlet distributed mixture probabilities at the brand and SKU levels, respectively. $\delta_{G_k}(.)$ and $\delta_{a_l}^{(k)}(.)$ represent point masses at $\mathbf{G}_k$ and $\mathbf{a}_l^{(k)}$, at the brand and SKU levels, respectively. Section 3.3. details the estimation of the quantities of interest in Equations (8) and (9).

As a motivating example for the nDP, consider assessing (perceived) similarity across brands among a store's clientele. The marketing outcomes (e.g., unit sales) of SKUs in each brand define a brand-specific distribution, which can be nonnormal, presenting skewness, multimodality, and/or heavy tails. There may be subsets of brands that share similar brand-specific distributions of SKU outcomes (e.g., unit sales). We argue that brands with similar distributions of SKU outcomes are, with a calculable probability, part of the same similarity set, aggregated across time. In this setting, it is of interest to cluster brands according to the full distribution of SKU outcomes, and to identify outlying brands. Thus, if for brands $b$ and $b'$, the distribution of unit sales outcomes is denoted by $\mathbf{F}_b$ and $\mathbf{F}_{b'}$, then under the nDP model, the probability that these two distributions follow the same latent discrete distribution $\mathbf{G}_k$ is strictly positive and is given by $1/(1+\alpha)$. On the other hand, it is also interesting to simultaneously cluster SKUs within the brands, and do so via borrowing information across brands that present SKUs with similar characteristics. Thus, we cluster both brands and SKUs simultaneously, consistent with the branding hierarchy, and thereby solving the problem presented in Figure 1.

As output, for each realization of draws from the nDP prior, we obtain a classification vector $C_k$ for $k = 1, 2, \ldots, K$ clusters, at the parent level for similar brands. That is, if brands $b$ and $b'$ follow $\mathbf{G}_k$, then $\{b, b'\} \in C_k$. Within each of the $K$ brand clusters, we obtain classification vectors at the child level for SKUs that are similar, $C_{k,l}^{(\mathrm{SKU})}$, for $l = 1, 2, \ldots, l_k$. Thus, for SKU $j$ of brand $b$, $\theta_{j(b)} = \mathbf{a}_l^{(k)}$ if $b \in C_k$ and $j \in C_{k,l}^{(\mathrm{SKU})}$.

We empirically test the assumption—that dependencies across the branding structure in our application (between the brand and the SKU levels) impact interproduct competition—underpinning the proposed nDP model. If a benchmark model, which treats brand merely as any other latent product attribute, displays fit and prediction as well as or better than the proposed nDP model, then our conjecture would be falsified. We split the data set into initialization, calibration, and holdout samples. The holdout sample is about half the size of the calibration sample (following Steckel and Vanhonacker 1993). We assess calibration-sample model fit and holdout-sample predictive validity using the RMSE criterion. In addition, we examine the face validity of the fixed, main effects parameters for consistency with theory and past studies, the proportion of parameters significant at the 95% credible interval range, and the average interval length (to assess the precision of estimates).

The nDP procedure has found use in areas such as machine learning and biostatistics (Ho et al. 2013, in the context of multicenter cancer trials), but not as yet, to the best of our knowledge, in the business research literature. In biostatistics, Ghosh et al. (2010) obtain predictions of cancer occurrence across states in the United States using the standard DP and the nDP model and compare them. They find that the latter model improves prediction considerably over the standard DP because the nDP is able to borrow information across units of analysis. Further, they report that the clustering of similar states provided by the nDP is useful and interpretable.

**2.2.3. Analyzing Clustering Output.** The rich MCMC output from the nDP model requires careful analysis and processing for usable interpretation and inference. At least three challenges emerge in analyzing the output from our proposed model implementation. First, the model is a Bayesian mixture model and is susceptible to the label switching problem (Diebolt and Robert 1994). A related complexity is that the model allows the number of clusters to be different across the Gibbs sampler iterations. Second, characterizing the clustering structure (or "partition") that is best supported by the data may not be a straightforward exercise because of the probabilistic nature of the cluster hierarchies and the lack of a "natural" estimator for partitioning (Lau and Green 2007). Finally, there is a need to infer the significance of the cluster allocations of the units, which requires that we assess the uncertainty associated with the process of assigning units to clusters in the "best" clustering solution.

To address these issues, we combine the use of posterior pairwise similarity matrices (PSMs) and a clustering inference procedure proposed by Molitor et al. (2010) into a four-step algorithm (detailed in Web Appendix B, available as supplemental material at http://dx.doi.org/10.1287/mnsc.2014.2039).

**2.2.4. A Latent Distance Based Measure of Competition.** Both the standard DP and the nDP outputs comprise a set of SKU-specific five-dimensional random parameter vectors $\boldsymbol{\theta}$, which, for any given SKU, represent the impact on the SKU's sales of the SKU's MMIX response, its product attribute effects, and the heterogeneous impact of competition from other SKUs. The $\boldsymbol{\theta}$ also lend readily to a spatial interpretation—as location addresses of SKUs in the latent space of MMIX response, competition impact, and product attributes. Following the literature on latent space models (e.g., Hoff et al. 2002, Handcock et al. 2007, Braun and Bonfrer 2011), we assume that each SKU has a position in five-dimensional latent product space and $\boldsymbol{\theta}_j$ denotes the five-dimensional latent coordinates of item $j$ in that space. Thus, using our initialization sample, we obtain a "latent distance metric" in the $p$-norm between any two SKUs $i$ and $j$, i.e.,

$$\begin{aligned}\text{dist}(i,j)^{(p)} &= d_{ij}^{(p)}\\ &= \left[|\alpha_{2,i}-\alpha_{2,j}|^p+\sum_{q=1}^{3}|\beta_{i,q}-\beta_{j,q}|^p+|r_i-r_j|^p\right]^{1/p}. \quad (10)\end{aligned}$$

We use the commonly used Euclidian norm ($p=2$). As a robustness check, in §4.5, we examine the sensitivity of results to variation in $p$. SIMIL in Equation (3), representing the degree of interproduct similarity in product attributes, relates inversely to this intuitively appealing notion of interproduct "distance" in latent attribute space. The idea is not new in the literature. For instance, in their study of consumer preference formation, Carpenter and Nakamoto (1989) relate preference similarity to the "addresses" of products in attribute space. Given the above interproduct distance $d_{ij}$, we now redefine $\text{SIMIL}_{ji}^{(\text{OLS})}$ in Equation (3) as some function $g(.)$ of $d_{ij}$, i.e., $\text{SIMIL}_{ji}^{(\text{nDP})} = g(d_{ji})$. In §2.4, we discuss the specification of the $g(.)$ function. In Equation (4), we rework the competitive impact of rival's MMIX on the focal item using the parameter vectors from the nDP output as

$$\begin{aligned}&\text{COMPTN}(jt,it)^{(\text{nDP})}\\ &\quad = g(d_{ji})\times\left[\ln(\mathbf{MMIX}_{it}^{(\text{P})})\boldsymbol{\beta}^{(\text{nDP})}+\mathbf{r}_i^{(\text{nDP})}\right], \quad (11)\\ &\text{COMPTN}_{jt}^{(\text{nDP})} = \sum_{i\neq j}\text{COMPTN}(jt,it)^{(\text{nDP})}.\end{aligned}$$

The new $\text{SIMIL}^{(\text{nDP})}$ is much more comprehensive than the $\text{SIMIL}^{(\text{OLS})}$ we used earlier because it takes into account not just latent product attributes, but also SKU response to the other three MMIX variables and to competitive effects from rival items. Consequently, our $\text{COMPTN}^{(\text{nDP})}$ variable, obtained using the results of the initialization sample, is also a more comprehensive competition measure than $\text{COMPTN}^{(\text{OLS})}$

because it (i) combines observed, time-varying MMIX elements with latent MMIX response (summarized in the $\beta^{(\text{nDP})}$ and in the random effect term $\mathbf{r}$) at a disaggregate SKU level; and (ii) scales the competitive impact of these observed and latent effects on focal item sales by degree of product similarity (captured in $\text{SIMIL}^{(\text{nDP})}$ via our latent distance framework). Further, our competition metric continues to be asymmetric across product pairs and yields intuitive substitution patterns.

**2.2.5. Revised Demand Specification.** Using the latent distance based $\text{COMPTN}^{(\text{nDP})}$ measure, we now define a revised demand model and estimate it via nDP using the calibration sample. Accordingly, the parameters in this model are superscripted with "(2)" to distinguish them from those in the sales function in Equation (1):

$$\ln(\text{SALES}_{jt}) = \boldsymbol{\alpha}_1^{(2)}\mathbf{Z}_t + \boldsymbol{\beta}_j^{(2)}\ln(\mathbf{MMIX}_{jt}^{(\mathrm{P})}) + \mathbf{r}_j^{(2)} + \alpha_{2,j}^{(2)}\,\text{COMPTN}_{jt}^{(\text{nDP})} + \boldsymbol{\alpha}_3^{(2)}\mathbf{X}_j + \boldsymbol{\varepsilon}_{jt}^{(2)}. \quad (12)$$

Apart from the COMPTN term, another way in which Equation (12) differs from Equation (1) is that time-invariant product attributes are included in the model as main, fixed effects ($\mathbf{X}_j$). The demand model in Equation (12) yields easily interpretable marginal cross effects. For instance, a change in rival item $i$'s price impacts focal item $j$ as

$$\frac{\partial \ln(\text{Sales}_{jt})}{\partial \ln(\text{PRICE}_{it})} = \alpha_{2,j}^{(2)} \times \text{SIMIL}_{ij}^{(\text{nDP})} \times \beta_{\text{PRICE},\,i}^{(\text{nDP})}. \quad (13)$$

## 2.3. Benchmark Models

Conceptually, the revised demand model in Equation (12) incorporates the following two key factors: (i) a competition measure that is based on latent, interproduct similarity as seen in Equations (10) and (11); and (ii) the brand-SKU hierarchy intrinsic in the nDP specification. These two factors can be viewed as generating a set of $2\times 2$ comparisons, as shown in Table 1.[2] One axis represents whether the demand model takes into consideration the brand-SKU hierarchy and the other considers whether the demand model incorporates competitive effects (COMPTN term) based on interproduct similarity in latent attribute space. The top right cell includes our proposed nDP specification. The top left cell does not consider the brand-SKU hierarchy, i.e., the random parameters are distributed according to a standard DP model, but incorporates the competition measure that is based on interproduct similarity in latent attribute space. The bottom right cell incorporates the brand-SKU hierarchy by having the random parameters distributed according to an nDP but

**Table 1 Framework for Benchmark Models**

| | | | |
|---|---|---|---|
| Whether model accounts for SIMIL based COMPTN? | Yes | Standard DP model with COMPTN term | Proposed nDP model with COMPTN term |
| | No | Standard log-log demand model | nDP model without the COMPTN term |
| | | No | Yes |
| | | Whether model accounts for brand-SKU hierarchy? | |

does not consider the interproduct similarity based competition term. The bottom-left cell does not consider either the branding hierarchy or the competition term, resulting in a standard log-log demand model, based on a multivariate Gaussian density. It is easy to see that the difference in results between the proposed nDP and the standard DP model would highlight the value of incorporating brand-SKU hierarchy constraints in a demand system. Likewise, a comparison between the proposed nDP and the nDP model without COMPTN would demonstrate the value of incorporating interproduct similarity based competition measure in a demand system.

We consider the above four models—one proposed and three benchmark specifications and compare the results along a set of metrics, namely, (i) within-sample fit and holdout-sample predictive accuracy using the RMSE criterion, (ii) average 95% credible interval length (for estimate precision), (iii) the proportion of parameters significant at the 95% credible level (Conley et al. 2008), and (vi) the correlation or overlap between the MMIX response coefficients and in particular, price coefficient, across the four specifications.

## 2.4. Subcategory Boundaries in Latent Parameter Space

Pinkse et al. (2002) semiparametrically model price competition in geographic space (between U.S. gasoline stations) and define a variety of spatial configurations to determine the neighborhood structure—including both continuous (Euclidean distance metric) and dichotomous measures (whether or not $i$ and $j$ are neighbors). They find that competition is highly localized. Since $\text{COMTPN}^{(\text{nDP})}$ is based on a distance metric in latent attribute space, the proposed model also lends to a spatial competition interpretation. Further, as an analog to "local" competition in geographic space, different spatial configurations for local competitive effects can be examined. To illustrate, since $\text{COMTPN}^{(\text{nDP})}$ is a function of $g(d_{ij})$, different specifications for $g(\cdot)$ lead to different models of competitive effects. For instance, $g(d_{ij}) = (d_{ij})^{-1} \times I(d_{ij} > 0)$ implies global competition between the focal item $j$ and all products $i$ that are not colocated with $j$. For

[2] We thank an anonymous reviewer for this suggestion.

local competitive effects, based on some definition of a "neighborhood" around $j$, we get

$$g(d_{ij}) = (d_{ij})^{-1} \times I(d_{ij} > 0) \times I(i \neq \text{j is in the neighborhood of } j). \quad (14)$$

We define "neighborhood" in three different ways to obtain different spatial configurations for local competitive effects: (i) a radius measure of neighborhood, i.e., if $i$ is within a fixed distance of $j$ (we use the median distance among all distance pairs), it is a neighbor of $j$; (ii) a brand-cluster measure, i.e., if $i$ and $j$ belong to the same substantive brand cluster (i.e., the parent level cluster in the nDP), then they are neighbors; and (iii) a SKU-cluster definition of neighborhood along the lines of the brand-cluster neighborhood. Following standard practice in the spatial statistics literature, we row standardize the similarity matrices $g(\mathrm{d}_{ij})$ to create proportional spatial weights for products having an unequal number of neighbors. To assess which of the four spatial competition configurations (one global and three local) best fits the data, we use the RMSE metric on both the calibration and the holdout sample. If we do find differential fit among different spatial configurations, then the data suggest subcategory boundaries along the local clusters where outcome differences are sharpest. We discuss this aspect further in the context of the empirical application in the discussion section.

At this point, we note that our approach relates to that of Braun and Bonfrer (2011) in at least four ways: One is in our use of the Dirichlet process prior to determine clustering among units of analysis on the basis of latent attributes. We use the nested DP extension of the standard DP model to incorporate brand-SKU hierarchy dependencies in building the probability model. Second is the similarity in our DP model's endogenous clustering property to reduce problem dimensionality (dealing in clusters rather than individual units) and make tractable the computational challenge of building a probability model over the possible outcomes or interactions between the units of analysis. Third is in leveraging latent space modeling by assigning coordinates to the clusters in latent attribute space (however, the dimensionality of the space differs in the two cases). And fourth is in inferring and extracting insights from the latent space itself (we infer submarket boundaries at the SKU level, whereas Braun and Bonfrer (2011) infer social circle similarity by focusing on other individuals in the same cluster).

# 3. Empirical Application

## 3.1. Data Description

We use beer category data from 56 stores of a midsized grocery chain in the U.S. Northeast. Our rationale for using a single retailer's data is that interproduct competition that results in choice at the store shelf depends on the assortment of products in the category and the MMIX of rival products. Hence, data from a single retailer are likely to maintain consistency in both product assortment as well as the stores' clientele that are exposed to the products. Our data set contains 23 weeks of sales and MMIX data for 96 SKUs from 15 brands, yielding a total of 2,171 usable observations. The SKU attributes we consider are BRAND, CONTAINER type (bottle/can), beer TYPE (ale, light, craft, regular), beer COLOR (light, amber, dark, golden), and PACKAGING types (6P 12 oz., 12P 12 oz., 18 pack, and 24 pack). Control variables were MONTH dummies (*January* to *May*) for seasonality. The MMIX variables available were DISTRIBUTION (% stores in which the item is available), PROMOTION (% sales made on any promotion), national annual ADSPEND for that year, PRICE per fluid ounce and unit COST per fluid ounce (as price instrument).

Table 2 summarizes the major characteristics of the data and the variables used in the analysis. Considerable variation is visible in both the dependent and the MMIX variables. Table 3 profiles the 15 brands used in the analysis, and again we see a lot of heterogeneity among brands in their use of MMIX elements.

**Table 2 Summary of Analysis Variables—Beer Category Data from One Retailer**

| Group | Variable | Description | Mean (SD) |
|---|---|---|---|
| $Y$ | *Sales* | Beer volume in fluid ounze | 39,204 (61531) |
| Price | *Priceoz* | Retail price per fluid ounce | 0.07 (0.02) |
| | *Costoz* | Wholesale price per fluid ounce | 0.06 (0.01) |
| Distribution | *Distbn* | % stores that sold the SKU | 0.70 (0.29) |
| Promotion | *Unitsp* | % units sold on promotion | 0.22 (0.36) |
| | *Adspend* | National ad-spend ($mn p.a.) | 13.97 (20.38) |
| Control variables | *January* | If week is in January | 0.170 |
| | *February* | If week is in February | 0.173 |
| | *March* | If week is in March | 0.173 |
| | *April* | If week is in April | 0.175 |
| | *May* | If week is in May | 0.175 |
| | *June* | If week is in June | 0.134 |

| Group | Variable | Proportion |
|---|---|---|
| Beer type | Ale | 0.011 |
| | Lite | 0.386 |
| | Craft | 0.095 |
| | Regular | 0.519 |
| Beer color | Light | 0.679 |
| | Amber | 0.094 |
| | Dark | 0.042 |
| | Golden | 0.185 |
| Packaging | Bottle | 0.711 |
| | 6P 12 oz. | 0.299 |
| | 12P 12 oz. | 0.454 |
| | 18 pack | 0.013 |
| | 24 pack | 0.021 |

**Table 3 Summary Profiles of Brands**

| Brand | Revenue share (%) | Volume share (%) | Mean price per fluid ounce | Mean cost per fluid ounce | Mean distribution (%) | Mean promotion (%) | Mean relative adspend |
|---|---|---|---|---|---|---|---|
| Amstel | 2.50 | 1.81 | 0.093 | 0.079 | 73.5 | 33.9 | 14,757.0 |
| Bass | 1.83 | 1.29 | 0.096 | 0.075 | 67.3 | 16.6 | 0.0 |
| Budweiser | 9.15 | 10.06 | 0.065 | 0.052 | 60.9 | 22.9 | 115,656.6 |
| Busch | 1.04 | 1.53 | 0.045 | 0.037 | 81.5 | 0.0 | 4,653.5 |
| Coors | 7.17 | 7.89 | 0.066 | 0.051 | 69.7 | 18.9 | 85,032.6 |
| Corona | 16.57 | 11.74 | 0.100 | 0.082 | 97.7 | 35.4 | 21,766.0 |
| Dos Equis | 0.92 | 0.68 | 0.093 | 0.070 | 71.2 | 30.7 | 1,959.4 |
| Harp | 1.05 | 0.72 | 0.098 | 0.080 | 61.0 | 19.9 | 0.0 |
| Heineken | 8.15 | 5.86 | 0.095 | 0.079 | 71.9 | 29.0 | 45,333.9 |
| Labatt | 25.42 | 29.61 | 0.062 | 0.051 | 76.3 | 36.4 | 1,183.4 |
| Michelob | 8.66 | 9.48 | 0.064 | 0.052 | 71.8 | 19.9 | 13,670.1 |
| Miller | 4.86 | 6.22 | 0.051 | 0.040 | 49.1 | 15.5 | 72,477.7 |
| Rolling Rock | 2.04 | 2.30 | 0.063 | 0.050 | 78.0 | 11.7 | 791.0 |
| Stella Artois | 1.89 | 1.20 | 0.106 | 0.083 | 80.8 | 22.2 | 0.0 |
| Yuengling | 8.74 | 9.63 | 0.062 | 0.051 | 78.6 | 6.9 | 200.0 |
| Average | 6.67 | 6.67 | 0.077 | 0.062 | 0.726 | 0.213 | 25,165.4 |
| SD | 6.82 | 7.47 | 0.020 | 0.016 | 0.111 | 0.105 | 37,150.0 |

The **H** matrices of instruments (in Footnote 1) for the three time-varying MMIX elements consist of exogenous variables that include product attributes, time dummies, price instruments (i.e., purchasers' price indices for material inputs to beer such as malt and barley, and aluminum for making cans), regional food inflation indices from the Bureau of Labor Statistics, and one-period lagged MMIX variables. The $R$-square of Stage 1 (in the 2SLS) regressions are well above 85% and the respective correlations between the actual and predicted MMIX variables are also high (above 0.75). More details about the 2SLS regressions are available upon request from the authors. Below, we discuss the specific details regarding the implementation of our model using the data.

### 3.2. Model Specifics

First, we specify the hyperpriors for the parameters in the DP priors. A reasonable choice for the base distribution is "what one might use in a purely parametric model" (Braun and Bonfrer 2011, p. 517). We specify $\mathbf{G}_0$ in Equation (7) for the standard DP model and $\mathbf{H}$ in $\text{nDP}(\alpha, \rho, \mathbf{H})$ for the nDP model as a five-variate Gaussian $\text{MVN}(\mu_0, \mathbf{\Sigma}_0)$ and in each case, place a normal-inverse Gamma hyperprior on $(\boldsymbol{\mu}_0, \mathbf{\Sigma}_0)$ to allow this base distribution to be flexible and determined by data. This choice of base distribution is one of the most widely used in applied work and has the appealing property that is conditionally conjugate and hence the full conditional distributions needed for Gibbs sampling have standard conjugate forms (Hjort et al. 2010, Chap. 7). For the fixed effects parameters and for the precision parameters, we used commonly used diffuse priors from the Bayesian literature.

For the prior on the concentration parameter $\alpha$, at both the brand and the SKU levels, we use $\alpha \sim \text{Gamma}(a, b)$, which has a scale parameter convention such that $E(\text{alpha}) = a/b$, and has diffuse Gamma hyperpriors for $a$ and $b$, i.e., $a \sim \text{Gamma}(0.001, 0.001)$ and $b \sim \text{Gamma}(0.001, 0.001)$. Since the expected number of clusters depends only on the concentration parameters and the number of units to be clustered, $n$ (Antoniak 1974), the choice of a prior on $\alpha$ and $\rho$ assumes particular importance. We assess how sensitive the resulting clustering is because of the prior on the concentration parameters, in two ways. First, by examining how different functional forms for the prior on $\alpha$ and $\rho$ (such as Gamma, uniform, and fixed values for each) affect the number and composition of clusters. A variety of priors such as the Gamma (Escobar and West 1995), inverse Gamma (Medvedovic et al. 2004), and the uniform (Ohlssen et al. 2007) have been proposed in the statistics literature. We find that allowing $\alpha$ and $\rho$ to vary under either a Gamma or a uniform prior produces very similar results—the 95% credible intervals for $\alpha$ and $\rho$, and the corresponding number of estimated clusters overlap. The cluster memberships in the two cases are also broadly similar. Following Dunson in Hjort et al. (2010, Chap. 7), we believe that allowing the concentration parameters to vary and the data to inform their values is better than pegging them to a fixed value (e.g., Kleinman and Ibrahim 1998). We present the details in §4.5.

Second, we examine how the shape and scale hyperparameter values for the Gamma-distributed priors for $\alpha$ and $\rho$ affect the number and composition of clusters. We tested $\text{Gamma}(1, 1)$, $\text{Gamma}(1, 2)$, $\text{Gamma}(2, 1)$, and $\text{Gamma}(2, 2)$ in addition to a non-informative $\text{Gamma}(a, b)$ prior. We find that on average, 91.7% of the SKU pairs match across different prior specifications. Thus, it appears that the different prior values used are in reasonable agreement regarding cluster membership recovery.

In Equation (11), heterogeneous MMIX response $\beta^{(\mathrm{nDP})}$ forms three of the five latent coordinates for an SKU's location. To ensure that differences in the scale of the MMIX variables do not unduly influence the magnitude of $\beta^{(\mathrm{nDP})}$, we standardize the MMIX variables so that the $\beta^{(\mathrm{nDP})}$ vector contains parameters that are comparable in scale. Note that it is only the relative distances between latent coordinates and not the absolute positioning of products in latent space that matter.

Kim et al. (2004) show that the number of estimated clusters in a DP routine is typically greater than the number of substantively meaningful and interpretable clusters. The solution to the overlapping clusters problem is intuitive. In a multivariate setting, if two clusters show overlap in their credibility intervals across all the dimensions of the multivariate density, then the two clusters can be considered to have significant overlap and hence part of the same substantive cluster. We use the above algorithm in our application. Although the algorithm is straightforward to apply in the standard DP case, in the nDP case however, we have to use it at two levels—first at the brand level and then at the brand-SKU level. Following Kim et al. (2004), we use the 90% credibility level to decide on overlaps among clusters. In our empirical application, we do find overlaps between clusters and subsequently, a smaller number of substantive than estimated clusters.

A related issue is that of products colocated in latent space. The DP's discrete output means that two SKUs, $i$ and $j$, may share the same set of latent coordinates implying that consumers perceive them as perfectly substitutable. Under COMPTN$^{(\mathrm{nDP})}$ however, since $d_{ij} = 0$, $g(d_{ij})$ would become infinity. To avoid such a situation and because $i \neq j$, we modify $g(.)$ as $g(d_{ij}) = 1/(d_{ij} + c)$, where $c$ is a small positive constant. We empirically set the value of $c$ by first looking at inter-SKU cluster distances and setting $c$ as half the minimum intercluster distance found in the data.

### 3.3. Estimation

Of the two broad types of Gibbs samplers available to draw from the DP prior—the Polya urn and blocked Gibbs samplers based on generalized stick-breaking priors—Ishwaran and James (2001) advocate using the blocked Gibbs sampler owing to its advantages.[3] We adopt a Gibbs sampler based on a stick-breaking prior for our model. However, we note that not all model parameters are block updated. Sethuraman (1994) provides an explicit characterization of $\mathbf{G}$ in terms of a stick-breaking construction, wherein $\mathbf{G}$ (in Equation (7)) is represented as an infinite mixture of discrete vectors ("atoms") $\boldsymbol{\theta}_k^*$ drawn from $\mathbf{G}_0$ with probabilities $w_k$ that sum to one drawn from a Dirichlet distribution:

$$\begin{aligned}\mathbf{G} = \sum_{k=1}^{\infty} w_k \delta_{\theta_k^*} \quad \text{with } w_k = \pi_k \prod_{i=1}^{k-1}(1-\pi_i),&\\ \text{for } k = 1, 2, \ldots, \infty, \text{ where}: \pi_k \mid \alpha \sim \mathrm{Beta}(1, \alpha)&\\ \text{and } \boldsymbol{\theta}_k^* \overset{\mathrm{IID}}{\sim} \mathbf{G}_0 \text{ for } k = 1, 2, \ldots, \infty,&\end{aligned} \tag{15}$$

where $\delta_{\theta_k}^*$ represents a point mass at $\boldsymbol{\theta}_k^*$. Similarly, for the nDP model in Equations (8) and (9), under the stick breaking characterization, the Dirichlet distributed mixture probabilities $\pi_k$ at the brand level and$\omega_l^{(k)}$ at the SKU level can be expressed as

$$\begin{aligned}\omega_l^{(k)} &= u_l^{(k)} \prod_{t=1}^{l-1}(1-u_t^{(k)}), \quad u_l^{(k)} \overset{\mathrm{IID}}{\sim} \mathrm{beta}(1, \rho),\\ \pi_k &= v_k \prod_{t=1}^{k-1}(1-v_t), \quad v_k \overset{\mathrm{IID}}{\sim} \mathrm{beta}(1, \alpha).\end{aligned} \tag{16}$$

The use of a finite dimensional $\mathbf{G}$ is appropriate because the weights in the infinite stick-breaking representation decrease rapidly for typical choices of $\alpha$ (Hjort et al. 2010, Chap. 7) and hence we obtain a fairly accurate approximation to the DP for a conservative choice on the upper bound of the number of clusters allowed. Following Ohlssen et al. (2007), we choose an upper bound on the number of clusters obtained to be greater than the square root of the units being clustered. Thus, for the 15 brands we choose an upper bound of $L = 10$ clusters and for the 96 SKUs, conditional on a brand cluster, upper bound $L_1 = 25$ clusters. The adequacy of the upper bounds is seen in the fact that the 97.5th percentile of the total number of clusters formed and used across iterations does not touch the upper bounds placed.

We ran the standard DP and the nDP models on a WinBUGS 1.4.3 platform. Our WinBUGS code for the nDP model is given in Web Appendix A. Runtime for the nDP was 2.4 seconds per iteration, 0.2 seconds per iteration for the standard DP model, and 0.01 seconds per iteration for the Normal-parametric model on a 2.8 GHz i7 processor. We ran a burn-in of 50,000 iterations and then chose every 30th draw of the next 30,000 draws from the converged joint posterior for each of two chains with different initial values. To assess convergence, we (i) ran a Geweke test, (ii) a Gelman–Rubin diagnostic test, and (iii) visually

[3] Block updates and superior mixing properties, conjugacy with the multinomial distribution, and ability to sample directly from $\mathbf{G}$ instead of marginalizing out $\mathbf{G}$.

inspected trace-plots of posterior draws of parameters. The Geweke test was run on the first 10% and the last 50% of the postburnout, thinned, posterior chains of each of the random parameters $\boldsymbol{\theta}_{j(b)}$, for the second-stage model in Equation (12). We find that in most of the cases, the $Z$ statistic from the Geweke test is unable to reject the null that the parameter means in the two sets of chains are the same. The range of the potential scale reduction factor in the Gelman–Rubin test output was within 1.1 for most parameters, tentatively supporting the convergence conjecture. We also visually inspected trace plots of posterior chains for a number of (randomly chosen) $\boldsymbol{\theta}_{j(b)}$ parameters and found that the chains mix well. These results, not shown for lack of space, are available upon request from the authors. Postprocessing for computing the best partition of units in the data, cluster overlaps and convergence diagnostic tests were all done on the *R* platform. We next discuss the results of our analysis and the corresponding managerial insights.

# 4. Results

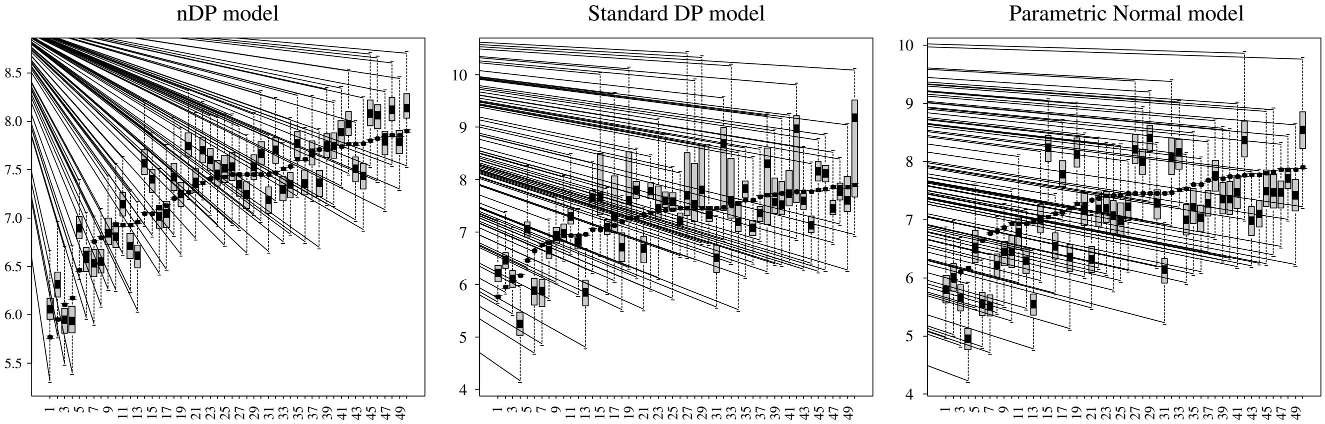
## 4.1. Proposed and Benchmark Model Comparisons

The results of Equation (12) from the proposed and the three benchmark specifications (from Table 1) are listed in Table 4. We first examine whether the additional complexity involved in specifying an nDP is worthwhile. Table 4, panel A compares the models on the basis of various measures of model fit and the quality of parameter estimation. Note that the observations for the predictive sample were randomly selected from the data. Our results show that the proposed nDP model outperforms the benchmarks on predictive accuracy in the holdout sample (RMSE-holdout). Following Gilbride and Lenk (2010) and Braun and Bonfrer (2011), we use posterior predictive checks (PPCs) to evaluate the contribution of our multilevel latent space model in replicating the data generation process in the holdout sample. As benchmark

**Table 4 Model Comparision (Stage 2 Results)**

| Metrics | Proposed nDP model | nDP model sans OMPTN | Standard DP model | Standard log-log model |
|---|---|---|---|---|
| | | Panel A | | |
| Holdout RMSE | 0.488 [0.435, 0.544] | 0.574 [0.522, 0.626] | 0.652 [0.608, 0.699] | 0.756 [0.665, 0.879] |
| % price coefficient estimates matching proposed models' | 100 | 86.40 | 40.62 | 37.50 |
| Average 95% interval length | 0.433 | 0.432 | 0.493 | 1.222 |
| % of parameters significant | 78.59 | 75.78 | 81.72 | 70.05 |
| | | Panel B | | |
| Latent attribute effects Mean [range] | 7.867 [7.75, 8.03] | 7.248 [7.13, 7.44] | 4.893 [4.74, 5.04] | −0.875 [−68.1, 21.8] |
| Price coefficient Mean [range] | −0.351 [−0.74, 0.00] | −0.284 [−0.72, 0.02] | 0.204 [−0.15, 0.64] | 0.375 [−3.26, 2.30] |
| Distribution coefficient Mean [range] | 0.801 [0.26, 1.14] | 0.843 [0.36, 1.30] | 0.834 [0.29, 1.06] | 0.659 [0.39, 0.84] |
| Promotion coefficient Mean [range] | 0.116 [0.02, 0.40] | 0.115 [−0.05, 1.30] | 0.116 [−0.07, 0.39] | 0.140 [−0.04, 0.39] |
| Competition response Mean [range] | −0.031 [−0.15, 0.07] | NA | −0.007 [−0.05, 0.03] | NA |
| | | Panel C | | |
| No. of estimated brand clusters | 6 | 7 | NA | NA |
| No. of substantive brand clusters | 4 | 3 | NA | NA |
| No. of estimated SKU clusters | 30 | 49 | 24 | NA |
| No. of substantive SKU clusters | 25 | 31 | 12 | NA |
| Concentration parameter Brand level ($\alpha$) | 3.45 [1.26, 6.71] | 2.60 [1.11, 5.37] | NA | NA |
| Concentration parameter SKU level ($\rho$) | 8.24 [5.38, 11.27] | 8.20 [4.93, 11.14] | 5.52 [3.13, 8.52] | NA |

**Figure 2** **Posterior Predictive Checks on the Holdout Sample**




models, we use the standard DP model and the Normal parametric model. In Figure 2, the solid dots show observed SKU outcomes (log sales) from the holdout sample and the box plots depict the dispersion of the predicted log sales across 1,000 draws from the Gibbs sampler. The box plots (representing the interquartile range) of the nDP model include noticeably more of the solid dots (actual observations) relative to the benchmark models. This indicates better fit, prediction, and data process replication under the nDP specification.

In addition to traditional fit and prediction metrics, we also compare the parameter estimation results, particularly for the SKU-specific price coefficient. From a substantive standpoint, the price parameter is important for practitioners because it affects resource allocation decisions. Figure 3 shows a box plot comparing the dispersion in estimated price response across the four models. Unlike Boatwright et al. (1999) who recommend constraining the signs of parameters to that predicted by theory, we use unconstrained parameter estimation. Only the nDP models show the expected negative sign for unconstrained price parameters. Thus, Figure 3 and Table 4, panel A reemphasize the need to incorporate brand-SKU hierarchy in demand models.

We now address our first research question, namely, determining the relative contribution of incorporating brand hierarchy and interproduct similarity in a demand model. The % of price parameters from the benchmark models that overlap in their 95% credible intervals with corresponding parameters from the proposed nDP model is shown in Table 4, panel A. The actual posterior summaries (mean and range) of the 96 price parameters obtained are shown in the second row of Table 4, panel B. Whereas 86.4% of the price parameters match when interproduct similarity-based competition is dropped from the model and brand-SKU hierarchy effects retained, only 40.6% of the estimates match when the hierarchy is dropped and interproduct similarity retained. This indicates that incorporating brand-SKU hierarchy is indeed relatively more impactful and important than incorporating interproduct similarity-based competition.

Also, across the four models, the fixed main effects are broadly similar in magnitude, sign, and significance. This lends confidence and face validity to our demand modeling exercise. Table 4, panel B depicts posterior summaries of the random parameter vectors obtained under each of the four models. Consistent with evidence from past studies, we would expect the

**Figure 3** **Price Parameters from Proposed and Benchmark Models**

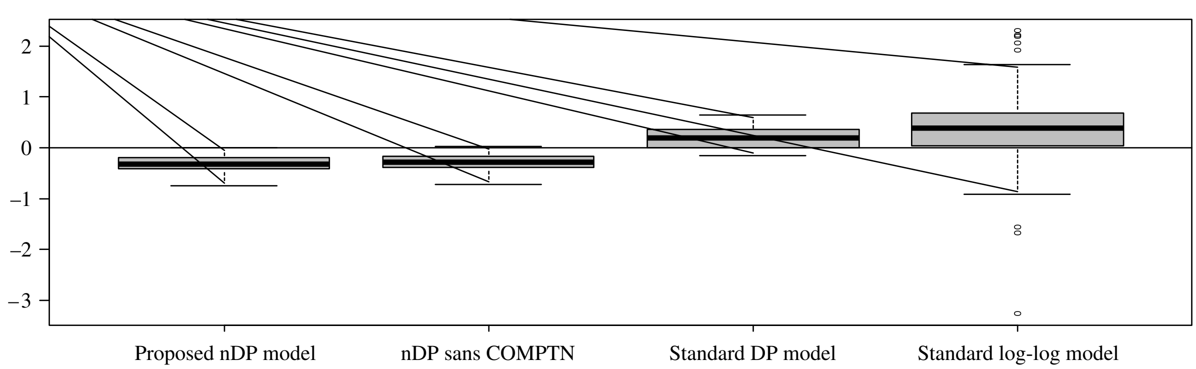


**Table 5 Summary of and Inference Over Brand-Level Cluster Characteristics**

| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|
| | | Cluster random effect | Cluster price coefficient | Cluster distribution coefficient | Cluster promotion coefficient | Cluster competitive impact |
| Brand cluster | Membership | $\beta_1$ posterior mean [95% interval] (Pr($\beta_L$ > Mean of $\beta$)) | | | | |
| 1 | Amstel, Busch, Dos Equis Heineken, Rolling Rock, Miller, Stella Artois, Yuengling | 7.844 [7.68, 7.99] (0.322) | **−0.367** [**−0.59**, **−0.18**] (**0.95**) | 0.839 [0.66, 1.05] (0.376) | 0.133 [0.03, 0.23] (0.544) | 0.01 [−0.15, 0.15] (0.684) |
| 2 | Bass, Harp | 7.842 [7.69, 8.00] (0.336) | −0.373 [−0.49, −0.23] (0.88) | 0.849 [0.68, 1.05] (0.432) | **0.277** [**0.13**, **0.44**] (**1**) | −0.001 [−0.16, 0.15] (0.62) |
| 3 | Budweiser, Coors, Corona, Michelob | 7.901 [7.75, 8.05] (0.796) | **−0.526** [**−0.73**, **−0.4**] (**0**) | 0.898 [0.68, 1.18] (0.78) | **0.103** [**0.05**, **0.2**] (**0.02**) | −0.067 [−0.21, 0.06] (0.21) |
| 4 | Labatt | 7.967 [7.76, 8.1] (0.87) | **−0.676** [**−0.76**, **−0.59**] (**0**) | **1.124** [**0.94**, **1.35**] (**0.998**) | 0.13 [0.1, 0.16] (0.48) | −0.124 [−0.27, 0.02] (0.102) |
| | Share weighted mean MMIX coefficients | 7.852 | −0.431 | 0.875 | 0.144 | −0.021 |

price and competition parameters to be negative, and the distribution and promotion parameters to be positive. The proposed model's estimates appear to be reasonable and most consistent with our expectations in this regard. Thus, overall, we find that the use of nDP to incorporate branding hierarchy does seem to matter in both fit and substantively interesting terms. §§4.2 and 4.3 focus on our second research question, i.e., similarity-based brand and SKU clusters.

### 4.2. An Analysis of Similarity

Table 4, panel C brings out the inferred latent structure of the data arising from an endogenous and simultaneous clustering at the brand and SKU levels. From a substantive standpoint, clusters showing large overlaps in all their clustering parameters can be merged to obtain a larger, more meaningful cluster (as described in §3.2). Table 4, panel C shows that the parent-level brand clusters in our proposed nDP model reduce from six to four and the corresponding collapse in the child level SKU clusters is from 30 to 25.[4] On the other hand, in the standard DP model, the number of SKU clusters reduce from 24 to 12. At the outset, although it may seem that the standard DP arrives at a more manageable number of SKU clusters, we note that it clusters SKUs without regard to branding hierarchy and cannot yield the latent brand clusters unlike the nested DP approach.

The bottom two rows in panel C display the posterior means and 95% credible intervals for the concentration parameters ($\alpha$ and $\rho$) at both the brand and SKU levels of clustering.

To illustrate inference on cluster-specific parameters, we take the example of brand-level clusters from the proposed model. Columns (1) and (2) in Table 5 show the substantive brand clusters and their membership, respectively. Note that this brand clustering in Table 5 is based on the three latent MMIX coefficients, the sales impact of competition, and the effect of aggregate time-invariant product attributes. Hence, it is quite possible that two brands that may be in the same price tier do not share membership in the same cluster because their latent price effects are significantly different. Also, these results may reflect regional idiosyncrasies as well (the data are from a northeastern retail chain). Columns (3)–(7) in Table 5 define a five-dimensional clustering parameter vector (say, $\boldsymbol{\theta}$) and correspond to the five latent parameters that form the basis for clustering. For these columns, each row corresponds to a cluster indexed by $l = 1, 2, \ldots, 4$ and describes $\boldsymbol{\theta}_l$, the clustering parameter vector for cluster $l$. Each cell in these columns gives three quantities—the posterior mean value of $\boldsymbol{\theta}_l$, the 95% credible interval (in square brackets) of $\boldsymbol{\theta}_l$, and the proportion of times $\boldsymbol{\theta}_l$ was greater than the average of $\boldsymbol{\theta}$ across all clusters and across 1,000 draws from the Gibbs sampler. Cells with a bold font indicate the presence of a "strong clustering signal" in that the $\boldsymbol{\theta}_1$ were either consistently above or below the overall mean for that column.[5] Thus, for instance,

[4] In any realization of the nDP, two SKUs sharing the same cluster label but belonging to different brand-level clusters are part of two different SKU level clusters. Hence, the number of estimated SKU clusters in the nDP model may well exceed those in the standard DP and the upper bound set for the SKU level in nDP, given a brand cluster.

[5] Web Appendix B describes the inference algorithm.

cluster 3 has a consistently low promotion elasticity parameter, whereas cluster 1 has a consistently high price elasticity parameter after taking into account parameter uncertainty.

If model-free data from the beer market could help explain at least some of these groupings, then it would lend a measure of external validity for the brand-level clustering. We test our conjecture that it is the unobserved "perceived quality" attributes that may help explain brand clustering through some model free evidence.[6] We used data from beeradvocate.com and ratebeer.com, the two largest beer-fan community websites (per Wikipedia), where we collected perceptual ratings on each brand (100 responses) on different one-dimensional constructs and also for "overall quality." Beer quality is a composite perceptual factor. It is typically broken down into a set of one-dimensional perceptual constructs such as taste, look, color, feel, aftertaste, smell etc. We ran a MANOVA procedure to assess whether and how strongly the set of perceived quality metrics relate to cluster groupings at the brand level. The results were the following: (i) an $F$-statistic of 2.408 at 5 and 1,494 degrees of freedom with a $p$-value of 0.03475 for beeradvocate.com data; and (ii) an $F$-statistic of 2.5808 with a $p$-value of 0.02471 for ratebeer.com data. Thus, we see that perceived quality is significantly different across clusters at the 95% confidence level in both cases. Also, individual ANOVA analyses revealed that the one-dimensional components of quality were significantly different across brand clusters. The results indicate that perceived quality, despite measurement error, seems to explain the cluster-wise differences in brands seen in our model results.

### 4.3. Substantive Implications

From a substantive standpoint, one interesting question is whether and how product similarities get reflected in marginal substitution patterns. Thus, for instance, if rival item A were to change its MMIX in some way, what effect would it have on focal item B's demand? How would the results change if items A and B were part of the same SKU-brand cluster, or were part of the same brand but not SKU level cluster? A more general question is, what are the practical implications of brand and SKU clusters?

To address these issues, we identify a few SKUs for illustrative purposes and work through their results. Table 6, panel A describes the identity of four SKUs, labeled 1 to 4. Items 1 and 2 belong to the same SKU cluster, item 3 belongs to the same brand cluster as one and two but a different SKU cluster and item 4 belongs to a different brand cluster altogether. We would expect SKUs within the same SKU cluster to be very similar, within the same brand cluster to be somewhat less similar, and those outside the brand cluster to not be similar (i.e., have low SIMIL scores). Table 6, panel B shows the SIMIL scores for the four items. The pattern of SIMIL scores appears consistent with expectations. Panel B also displays the values of the competition response parameter and price parameter, which are needed to compute the marginal effect of a price change in any item on the other items, according to Equation (13). Panel C depicts the marginal substitution effects (in standardized units) implied by our competition model because of a price change in a rival item (in the columns) on the demand for the focal item (in the row). The bold cells in the diagonal are the own price effects. Thus, for instance, the interpretation of the cell $(2, 1)$ in panel C is that a 1% change in the price of item 1 raises the demand for item 2 by 0.033%, whereas a 1% rise in item 2's price raises item 1's sales by 0.002% in cell $(1, 2)$. Thus, the asymmetry in marginal cross effects is easily seen.

To see the practical implications of our model's hierarchy-consistent endogenous clustering, we take the perspectives of a retailer and a manufacturer. Retailers may want to know which national brands and products share similar profiles as their store brand products do. The SIMIL table in panel B aids in such an effort. Brand and SKU clusters help in understanding competitive groupings of products, which is necessary if the retailer is seeking to position the store brand against a particular national brand, for instance. This may have implications for rearranging products on store shelves based on product similarities. Going further, retailers may seek information on particular substitution patterns to better predict the effect of product promotions. Thus, if SKU 1 is on sale, the first column of Table 6, panel C shows the sales impact on the other SKUs. Equation (13), which governs the relation between asymmetric marginal cross effects, stresses the importance of the magnitude of the competitive response (or vulnerability) of the focal item $j$ $(\alpha_{2,j})$ and of the rival item $i$'s price effect, $\beta_i^{\text{PRICE}}$. To see this, consider that a 1% price discount on SKU 1 prompts a 0.03% drop in SKU 2's sales but larger drops of 0.06% and 0.07% in SKU 3's and SKU 4's sales, despite the fact that SKU 2 is closer to SKU 1 (same SKU cluster) than are SKUs 3 and 4. The higher absolute price effects of SKUs 3 and 4 more than compensate for the relatively lower SIMIL score in this case. It is easy to see how price promotions by different SKUs impact a focal SKU. For instance, from the third row in panel C, SKU 4's sales are most affected when SKU 3 promotes, followed by SKU 2 and SKU 1. The substitution table thus provides a means for retailers to better *simulate* the effects of changes in MMIX elements for sets of products on other products and

[6] We thank an anonymous referee for this suggestion.

**Table 6 Substitution Patterns from the Competition Model**

Panel A

| SKU ID | SKU description | Brand cluster | SKU cluster |
|---|---|---|---|
| 1 | Budweiser 6P 12 oz can | 3 | 18 |
| 2 | Michelob Lite 12P 12 oz can | 3 | 18 |
| 3 | Corona Extra 6P 12 oz NR bottle | 3 | 20 |
| 4 | Labatt 12P 12 oz can | 4 | 23 |

Panel B: SIMIL scores

| SKUs | 1 | 2 | 3 | COMPTN coefficient | Price coefficient |
|---|---|---|---|---|---|
| 1 | | | | −0.001 | −0.4139 |
| 2 | 5.41 | | | −0.015 | −0.419 |
| 3 | 2.49 | 4.58 | | −0.054 | −0.5605 |
| 4 | 1.80 | 2.68 | 2.51 | −0.095 | −0.7393 |

Panel C: Marginal substitution effects (in standardized units)

| | Rival SKUs | | | |
|---|---|---|---|---|
| Focal SKUs | 1 | 2 | 3 | 4 |
| 1 | **−0.414** | 0.002 | 0.001 | 0.001 |
| 2 | 0.033 | **−0.419** | 0.038 | 0.029 |
| 3 | 0.055 | 0.103 | **−0.561** | 0.100 |
| 4 | 0.071 | 0.107 | 0.134 | **−0.739** |

Panel D: Parameters for clusters 18, 20, and 23

| SKU IDs → | 1 and 2 | 3 | 4 |
|---|---|---|---|
| SKU clusters → | 18 | 20 | 23 |
| Latent attribute effects | 7.873 [7.69, 8.03] (0.698) | 7.938 [7.74, 8.11] (0.904) | 7.995 [7.76, 8.17] (0.916) |
| Price coefficient | −0.438 [−0.58, −0.31] (0.004) | −0.593 [−0.75, −0.45] (0) | −0.747 [−0.86, −0.64] (0) |
| Distribution coefficient | 0.779 [0.47, 1.07] (0.45) | 0.998 [0.77, 1.23] (0.996) | 1.189 [0.95, 1.46] (1) |
| Promotion coefficient | 0.086 [0.00, 0.18] (0.13) | 0.086 [−0.01, 0.21] (0.122) | 0.143 [0.1, 0.19] (0.922) |
| Competition response | −0.035 [−0.22, 0.16] (0.234) | −0.102 [−0.26, 0.07] (0.044) | −0.153 [−0.33, 0.01] (0.02) |

examine the corresponding sales differential. Assuming no supply side constraints, a follow-up analysis would include optimizing the MMIX variables to maximize category profitability. Note that although we estimate our model for a chain's data, one can easily apply it to individual stores. Interstore differences in competitive groupings and clientele preferences could then be assessed and acted upon (in the spirit of Hoch et al. 1995) and determine store-specific assortment and promotion decisions.

Much of the same benefits from our model (that endogenizes clustering under a branding hierarchy and a latent distance-based measure of competition) accrue to manufacturers as well. Manufacturers seeking to promote their products, would have a better understanding of (i) which store brand or rival national brand products compete most with their products and (ii) assess how well their brand line and product line is represented in different brand and SKU clusters that emerge. Table 6, panel D (marketing mix and competition parameters and the 95% intervals for the three SKU clusters) would aid such an analysis.

**4.3.1. Relating nDP to Preference-Based Competition Models.** In this section, we link our proposed model to other types of preference-based competition models in the marketing literature and their application contexts. First, our model relates to models of lexicographic consumer preference structures and to hierarchical elimination models (e.g., elimination by aspects (EBA) (Tversky 1972); elimination by tree (EBT), and the hierarchical elimination model (HEM) (Moore et al. 1986), most of which require data at the individual consumer level. Using data at an aggregate level and taking the perspective of firms (manufacturers and retailers), the nDP model can be viewed as "eliminating" sets of products that do not share "aspects" (in our case, brand membership within a similar brands cluster) from a downstream similarity set, based on latent product characteristics. Second, our framework relates to models of market structures (Shugan 2014) based on a brand-primary or form-primary sequential consumer choice process with different marketing implications in each case (e.g., Kalwani and Morrison's 1977 description of the Hendry model). Our model presents an aggregated data-based test of whether a product category follows a brand-primary or form-primary market structure. Since the benchmark, standard DP model is unconstrained in its clustering of SKUs while the proposed nDP model imposes branding structure constraints on SKU level clustering, superior in-sample fit and holdout sample predictions in favor of the nDP model relative to the standard DP model perhaps represents a brand-primary market structure. Further in this vein, Kim and Chintagunta (2012) use a brand-primary framework to find an additional and important source of preference heterogeneity in consumption occasion and context. They estimate segment-wise low-dimensional (factor analytic) brand maps to illustrate these effects, highlight insights on latent brand similarities, and thus on the potential for competition among brands. This connects well with our study's use of latent attribute space to identify and cluster brand locations accordingly. And third, our model relates to noncompensatory models of consideration sets. Gilbride and Allenby (2004) propose a first-stage, noncompensatory screening rule followed by compensatory second-stage choice to identify respondents' consideration sets. In a similar context, one interpretation of our nDP model is that it is premised on a noncompensatory brand-primary "screening" stage that clusters similar brands followed by a compensatory SKU clustering. Although we use the brand-SKU nested hierarchy to model a brand-primary structure in this application, we note that any nested hierarchical structure in which the parent and child levels are clearly defined can be similarly modeled. The implication is that in principle, the nDP model can accommodate any prespecified structure (such as form-primary) and test against any competing structure.

### 4.4. Configurations of Spatial Competition

Here, we examine our third research question. Following a rich stream of literature on spatial competition models in economics and marketing, we investigate interproduct competition in the beer category in different latent space configurations. In the industrial organization literature, most empirical models of differentiated product markets use a discrete choice framework with two commonly used models—a global random utility model (e.g., Berry et al. 1995, Sudhir 2001) in which every product competes with every other product and a local spatial model in which most price elasticities are either zero a priori or substantially smaller outside spatial clusters (e.g., Goldberg 1995, Verboven 1996 with the nested logit). The number of dimensions of product space has also varied. Whereas Bresnahan (1981, 1987) models one-dimensional competition where each product competes only with its two vertical neighbors in quality-tiers, Feenstra and Levinsohn (1995) allow for multiple dimensions in product space and compute endogenous market boundaries. In the marketing literature, Bronnenberg and Vanhonacker (1996) study brand choice in low-involvement categories and find that consumers compare prices within a limited set of brands (the "choice set"). They show that assuming global competition in the face of local price response may bias price elasticities and associated metrics.

In this study, we examine four spatial competition configurations—one global and three local, as described in §2.4. For these four configurations of spatial competition, Table 7 displays demand model comparison metrics (for the model in Equation (12)) for both the calibration sample and the holdout sample. The results point to three observations. First, the global competition assumption performs poorly relative to the local competition specifications, in both the calibration and the holdout samples. Thus, the notion that every product competes with every other does not find supporting evidence in our study. Second,

**Table 7 Model Comparison of Spatial Competition Configurations**

| | COMPTN specifications | | | |
|---|---|---|---|---|
| Model performance metrics | GLOBAL | Radius < Median interproduct distance | Within-brand cluster as neighbors | Within SKU cluster as neighbors |
| RMSE In-sample | 0.179 [0.177, 0.181] | 0.181 [0.178, 0.183] | 0.179 [0.176, 0.182] | 0.178 [0.176, 0.181] |
| RMSE Out-of-sample | 1.378 [1.168, 1.548] | 0.793 [0.637, 0.93] | 0.610 [0.532, 0.697] | 0.565 [0.514, 0.627] |

among the three local configurations, a purely spatial rule-based configuration that sets the neighborhood at the median inter-SKU distance fits poorly relative to the use of nDP generated clusters as neighborhood definitions. Thus, an exogenously imposed neighborhood rule appears to underperform relative to an endogenously determined neighborhood structure. Third, using either the parent-level brand clusters as closed neighborhoods or the finer SKU clusters as closed neighborhoods yields comparable fit to data (their RMSE credibility intervals overlap). However, the more refined SKU-level cluster significantly outperforms the radius-based neighborhood definition in holdout RMSE. Overall, the evidence suggests that the SKU clusters derived from the nDP model show coherence within the cluster and dissimilarity with other SKU clusters (even within the same parent-level brand cluster). Thus, they appear to play the role of latent submarket boundaries within which products compete directly and outside where the competition intensity diminishes. Thus, we note that store-level data are sufficient to discern these groups of similar SKUs conditioned on brand similarity.

### 4.5. Robustness Checks

We conduct robustness checks on alternative model specifications. First, we checked the sensitivity of model results to different priors on $\alpha$, as mentioned in §3.2. The 95% credible intervals for both the $\alpha$ parameters and the estimated numbers of clusters under a uniform and a Gamma prior overlap. The 95% credible intervals of estimated number of clusters at the brand level are $[6, 8]$ and $[6, 9]$, respectively. The corresponding intervals at the SKU level are $[19, 25]$ and $[17, 24]$. All the results reported in this study come from the model with the Gamma prior on $\alpha$. Second, we examined the potential impact of using particular hyperparameter values in the Gamma priors for the $\alpha$ parameters at the brand and SKU levels in the Stage 1 nDP model. We tested a variety of shape and scale hyperparameters and found that cluster formation (number of estimated clusters) and composition (cluster membership in terms of pairwise coincidences) in each case retained a large measure of similarity with the diffuse Gamma priors we use in the paper. Moreover, the models are comparable also on other important parameters. These results are provided in Web Appendix C. To assess the relative influence of prior and data on the posterior density of the number of clusters formed, we compared the expected variance of number of clusters (following the analytical expression in Kottas et al. 2005) with the corresponding observed variance. We find that in all cases, the variance of the observed number of clusters is a small fraction of that predicted by the DP model, suggesting that the data dominates the prior in exerting more influence over the posterior distribution of the clusters. Third, we examined the sensitivity of results to changing the norm $p$ in Equation (10). We found that the resulting distance matrices for $p = \{1, 2, 3\}$ were highly correlated (correlation $= 0.99$) and that using one distance matrix versus another did not significantly affect the results. The question may arise whether other, simpler measures of competition (e.g., market share assumed proportional to competitive effect) explain and predict better than our proposed and benchmark models. To assess this, we ran a demand model in which the competitive effect on a focal SKU is proportional to the MMIX of the rival SKU weighted by its market share. We find that such a model underperforms our proposed and benchmark models in terms of predictive fit. The results, not shown in the interest of space, are available upon request from the authors. In sum, we find that the results from the empirical application bear out the validity and the usefulness of our competition model (which incorporates interproduct similarity measure in latent space and brand-SKU hierarchy) in the context of addressing our research questions and shed light on the nature of interproduct competition in the beer category using store-level data alone.

## 5. Conclusions

We develop and implement a multilevel model of demand under interproduct competition that incorporates both brand-SKU hierarchy and interproduct similarity. Our two-stage approach incorporates the branding hierarchy effects through a novel Bayesian semiparametric approach based on a nested Dirichlet process, and models interproduct similarity in latent attribute space. Focusing on our first research question, we compare the relative contribution of branding hierarchy and product similarity in modeling demand. Extending the branding literature to a competitive setting, we find that branding hierarchy effects dominate those from product similarity, justifying in part the model's additional complexity. Addressing our second research question, we (i) determine best partitions of brands and SKUs conditional on brand clusters in the data, (ii) account for the impact of parameter uncertainty, and (iii) demonstrate inference on cluster-specific parameters. A natural spatial interpretation arises if we view the estimated product-specific parameters as location coordinates of products in latent attribute space. Accordingly, to our third research question, we develop and estimate a set of models of spatial interproduct competition. Here, we find that SKU competition is more local than global in that only subsets of products compete within groups of comparable products.

Our paper contributes to the marketing literature in two broad ways. In methodological terms, we (i) demonstrate the adaptation and operationalization of the nested Dirichlet process model to analyze within-category, interproduct competition based solely on aggregate, store-level data, (ii) partition the hierarchical data into clusters of similar brands and clusters of similar SKUs conditional on the brand clusters, and (iii) extend spatial competition models to an inter-SKU latent attribute setting by defining neighborhood structures around the brand and SKU clusters obtained. From a substantive standpoint, we (i) demonstrate the relative importance of branding hierarchy versus product similarity in modeling demand under competition, (ii) endogenously cluster a set of SKUs in latent attribute space, (iii) demonstrate the implications of using different model specifications on substantively interesting outcomes—namely, marketing mix effects and substitution patterns, and (iv) find meaningful and interpretable brand and SKU clusters in the data based on cluster-specific parameters. Finally, we find differential competition intensity within and outside the brand-SKU clusters, which when taken into account yield superior model performance (fit and prediction) and the possibility of demarking substantively interesting subcategory boundaries.

Our research has limitations that future research could address. One avenue of future research is to extend the nDP model to incorporate panel data and validate the results with individual level data from panel members on their brand-level consideration or choice sets prior to purchase. Second, our model is implemented on one retailer's data in one category. Although this helps maintain consistency in the retailer's core clientele and the product assortment they are exposed to, for more generality, using data from more categories or from individual stores in different geographic regions would be another avenue of future research. Third, using data on SKU level promotion and advertising, our model can be extended to include interaction effects (e.g., price elasticity depends on advertising), and dynamic effects (e.g., impact of past period sales on current demand) in the demand specification. Fourth, the standard DP models (including the nDP extension) cluster units on all the dimensions considered. This may be too restrictive as two products could be very similar in four dimensions and not on the fifth. Rather than such "global" clustering, a "matrix stick breaking process" approach (e.g., Dunson et al. 2008) that allows for local dependent clustering appears to be a promising area for future research. Fifth, when faced with attributes that do not fall under a formal nesting structure (e.g., the "diet" attribute in the carbonated soda category), an extension to the class of Hierarchical DP models (Teh et al. 2006) can be considered. Thus, although work remains to be done, the concept of developing concise measures of competition that are stable seems viable and worthwhile.

## Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mnsc.2014.2039.

## Acknowledgments

The authors thank three anonymous referees, the associate editor, and Pradeep Chintagunta (the department editor) as well as the seminar participants at the Tuck School of Business at Dartmouth, Dartmouth College, and the attendees at the 2013 Northeast Marketing Consortium (hosted by Harvard Business School) for their helpful comments on earlier versions of this paper. The authors acknowledge the research assistant role of Ankit Anand at the Indian School of Business.

## References

Ailawadi K, Kopalle PK, Neslin S (2005) Predicting competitive response to a major policy change: Combining game-theoretic and empirical analyses. *Marketing Sci.* 24(1):12–24.

Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2): 131–145.

Antoniak CE (1974) Mixtures of Dirichlet processes with applications to nonparametric problems. *Ann. Statist.* 2(4):1152–1174.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Boatwright P, McCullogh R, Rossi P (1999) Account-level modeling for trade promotion: An application of a constrained parameter hierarchical model. *J. Amer. Statist. Assoc.* 94(448):1063–1073.

Braun M, Bonfrer A (2011) Scalable inference of customer similarities from interactions data using Dirichlet processes. *Marketing Sci.* 30(3):513–531.

Bresnahan TF (1981) Departures from marginal cost pricing in the American automobile industry: Estimates for 1977–1978. *J. Econometrics* 17(2):201–277.

Bresnahan TF (1987) Competition and collusion in the American automobile industry: The 1955 price war. *J. Indust. Econom.* 35(4):457–482.

Bronnenberg BJ, Vanhonacker WH (1996) Limited choice sets, local price response and implied measures of price competition. *J. Marketing Res.* 33(2):163–173.

Carpenter GS, Nakamoto K (1989) Consumer preference formation and pioneering advantage. *J. Marketing Res.* 26(3):285–298.

Chintagunta PK, Jain DC, Vilcassim NK (1991) Investigating heterogeneity in brand preferences in logit models for panel data. *J. Marketing Res.* 28(4):417–428.

Chintagunta PK (1994) Heterogeneous logit model implications for brand positioning. *J. Marketing Res.* 31(2):304–312.

Conley TG, Hansen CB, McCullogh RE, Rossi PE (2008) A semi-parametric Bayesian approach to the instrumental variable problem. *J. Econometrics* 144(1):276–305.

Diebolt J, Robert CP (1994) Estimation of finite mixture distributions through Bayesian sampling. *J. Royal Statist. Soc. Ser. B* 56(2):363–375.

Dunson DB, Xue Y, Carin L (2008) The matrix stick-breaking process. *J. Amer. Statist. Assoc.* 103(481):317–327.

Eliashberg J, Chatterjee R (1985) Analytical models of competition with implications for marketing: Issues, findings, and outlook. *J. Marketing Res.* 22(3):237–261.

Elrod T, Keane MP (1995) A factor-analytic probit model for representing market structure in panel data. *J. Marketing Res.* 32(1):1–16.

Escobar MD, West M (1995) Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* 90(430):577–588.

Fader PS, Hardie BGS (1996) Modeling consumer choice among SKUs. *J. Marketing Res.* 33(4):442–452.

Feenstra RC, Levinsohn JA (1995) Estimating markups and market conduct with multidimensional product attributes. *Rev. Econom. Stud.* 62(1):19–52.

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. *Ann. Statist.* 1(2):209–230.

Ghosh K, Ghosh P, Tiwari RC (2010) Invited discussion on nested Dirichlet process prior (by Rodriguez, Dunson and Gelfand). *J. Amer. Statist. Assoc.* 103:1147–1149.

Gilbride TJ, Allenby GM (2004) A choice model with conjunctive, disjunctive and compensatory screening rules. *Marketing Sci.* 23(3):391–406.

Gilbride TJ, Lenk PJ (2010) Posterior predictive model checking: An application to multivariate normal heterogeneity. *J. Marketing Res.* 47(4):896–909.

Goldberg PK (1995) Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry. *Econometrica* 63(4):891–951.

Handcock MS, Raftery AE, Tantrum JM (2007) Model-based clustering for social networks. *J. Royal Statist. Soc. Ser. A* 170(2): 301–354.

Hjort NL, Holmes C, Muller P, Walker SG (2010) *Bayesian Nonparametrics* (Cambridge University Press, New York).

Ho M-W, Tu W, Ghosh P, Tiwari CR (2013) A nested Dirichlet process analysis of cluster randomized trial data with application in geriatric care assessment. *J. Amer. Statist. Assoc.* 108(501): 48–68.

Hoch SJ, Kim BD, Montgomery AL, Rossi PE (1995) Determinants of store-level price elasticities. *J. Marketing Res.* 32(1):17–29.

Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.* 97(460): 1090–1098.

Horsky D, Nelson P (1992) New brand positioning and pricing in an oligopolistic market. *Marketing Sci.* 11(2):133–153.

Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* 96(453):161–173.

Kalwani MU, Morrison DG (1977) A parsimonious description of the Hendry System. *Management Sci.* 23(5):467–477.

Kamakura WA, Russell GT (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 25:379–390.

Kannan PK, Wright GP (1991a) Modeling and testing structured markets: A nested logit approach. *Marketing Sci.* 10(1):58–82.

Kannan PK, Wright GP (1991b) On testing competitive market structures. *Marketing Sci.* 10(4):338–347.

Kim JG, Menzefricke U, Feinberg FM (2004) Assessing heterogeneity in discrete choice models using a Dirichlet process prior. *Rev. Marketing Sci.* 2(1):1–39.

Kim M, Chintagunta PK (2012) Investigating brand preferences across social groups and consumption contexts. *Quant. Marketing Econom.* 10(3):305–333.

Kleinman KP, Ibrahim JG (1998) A semi-parametric Bayesian approach to generalized linear mixed models. *Statist. Medicine* 17(22):2579–2596.

Kottas A, Muller P, Quintana F (2005) Nonparametric Bayesian modeling for multivariate ordinal data. *J. Computational and Graphical Statist.* 14(3):610–625.

Lau JW, Green PJ (2007) Bayesian model-based clustering procedures. *J. Computational and Graphical Statist.* 16(3):526–558.

Leeflang PSH, Wittink DR, Wedel M, Naert PA (2000) *Building Models for Marketing Decisions* (Kluwer Academic Publishers, Dordrecht, The Netherlands).

Medvedovic M, Yeung KY, Bumgarner RE (2004) Bayesian mixture model based clustering of replicated microarray data. *Bioformatics* 20(8):1222–1232.

Molitor J, Papathomas M, Jerrett M, Richardson S (2010) Bayesian profile regression with an application to the national survey of children's health. *Biostatistics* 11(3):484–498.

Moore WL, Pessemier EA, Lehmann DL (1986) Hierarchical representations of market structures and choice processes through preference trees. *J. Bus. Res.* 14(5):371–386.

Nedungadi P (1990) Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *J. Consumer Res.* 17(3):263–276.

Ohlssen D, Sharples LD, Spiegelhalter DJ (2007) Flexible random-effects models using Bayesian semi-parametric models: Application to institutional comparisons. *Statist. Medicine* 26(9): 2088–2112.

Park C, Srinivasan V (1994) A survey based method for measuring and understanding brand equity and its extendability. *J. Marketing Res.* 31(2):271–288.

Pinkse J, Slade M, Brett C (2002) Spatial price competition: A semiparametric approach. *Econometrica* 70(3):1111–1153.

Rodríguez A, Dunson DB, Gelfand AE (2008) The nested Dirichlet process. *J. Amer. Statist. Assoc.* 103(483):1131–1154.

Russell GJ, Kamakura WA (1993) Understanding brand competition using micro and macro scanner data. *J. Marketing Res.* 31(2):289–303.

Sethuraman J (1994) A constructive definition of Dirichlet priors. *Statistica Sinica* 4:639–650.

Sethuraman R (2003) Measuring national brands' equity over store brands. *Rev. Marketing Sci.* 1(2):1–26.

Shocker AD,Ben-Akiva M, Boccaro B (1991) Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Lett.* 2(3):181–197.

Shugan SM (2014) Market structure research. Winer SR, Neslin SA, eds. *The History of Marketing Science*. (World Scientific Publishing, Singapore), 129–164.

Steckel JH, Vanhonacker WR (1993) Cross-validating regression models in marketing research. *Marketing Sci.* 12(4):415–427.

Sudhir K (2001) Competitive pricing behavior in the auto market: A structural analysis. *Marketing Sci.* 20(1):42–60.

Sullivan MW (1998) How brand names affect the demand for twin automobiles. *J. Marketing Res.* 35(2):154–165.

Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* 101(476):1566–1581.

Tversky A (1972) Elimination by aspects: A theory of choice. *Psych. Rev.* 79(4):281–299.

Verboven F (1996) International price discrimination in the European car market. *RAND J. Econom.* 27(2):240–268.

Wedel M, Zhang J (2004) Analyzing brand competition across subcategories. *J. Marketing Res.* 41(4):448–456.

Yang S, Allenby GM, Fennell G (2002) Modeling variation in brand preference: The roles of bjective environment and motivating conditions. *Marketing Sci.* 21(1):14–31.