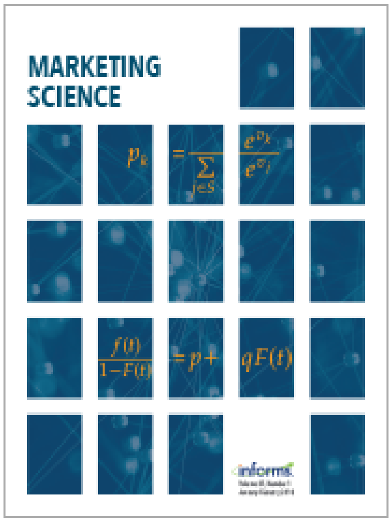


Marketing Science

Market Structure Across Retail Formats

Karsten Hansen, Vishal Singh,

MARKETING SCIENCE
Vol. 28, No. 4, July–August 2009, pp. 656–673
ISSN 0732-2399 | EISSN 1526-548X | 09 | 2804 | 0656

DOI 10.1287/mksc.1080.0432

Market Structure Across Retail Formats

Karsten Hansen
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
karsten-hansen@kellogg.northwestern.edu

Vishal Singh
Stern School of Business, New York University, New York, New York 10012,
vsingh@stern.nyu.edu

We study how market structure within a product category varies across retail formats. Building on the literature on internal market structure, we estimate a joint store and brand choice model where the loading matrix of brand attributes are allowed to be retail format specific. The approach allows us to recover brand maps for different retail formats while controlling for the short-term marketing mix activities at these stores and the self-selection of households that frequent a particular format. The model is applied to consumer panel data from two product categories, where households are observed to make purchases across three store types: high-end grocery store, traditional supermarket, and large everyday low pricing (EDLP) formats. Our results show strong correlations between the marketing mix sensitivities, store format preference, and unobserved brand attributes. These correlations translate into significant differences in market structure across retail formats and in the direction and size of preference vectors for unobservable brand attributes. We find a tight clustering of brands at the EDLP format, whereas brands are found to compete in distinct subgroups at other stores. Results show that failure to account for retail format effects can substantially bias the understanding of underlying market structure and could lead to incorrect implications in applications such as new product entry.

Key words: market structure; brand maps; retail formats; heterogeneity
History: Received: March 14, 2007; accepted: June 29, 2007; processed by Greg Allenby. Published online in *Articles in Advance* January 7, 2009.

1. Introduction

The analysis of market structure (Day et al. 1979) constitutes an important area of research in marketing and has received considerable attention in the literature (see Elrod et al. 2002 for a recent overview). The basic endeavor in this line of research is to identify (using market shares or consumer preference data) the extent to which different brands compete with each other in the marketplace. A popular approach to understanding such interbrand similarities or rivalries is to construct market maps that pictorially depict the locations of brands in a multiattribute space (Elrod 1988). Such spatial representation of brands is appealing as it relates closely with the notion of perceptual maps, regarded as the fundamental blocks for positioning strategies.

In this paper, we study how market structure in a product category varies across retail formats. Our focus is on frequently purchased consumer packaged goods sold through supermarkets. The grocery industry in the United States has undergone significant changes over the past two decades. Traditionally, the industry consisted of small family-owned businesses that operated on the principle of high margins and low turnover. The Atlantic and Pacific Tea Company (A&P) introduced the concept of an organized multiple-store network that marked the transition to a new phase in grocery retailing and the birth of the chain store. With multiple stores, A&P generated the high volumes necessary to obtain quantity discounts from the manufacturers while reaping economies-of-scale benefits on self-produced private label items. The supermarket format that first emerged in the 1930s took the concept of high volume and low prices to a new level. Supermarkets were much larger than the existing grocery stores and were primarily located in low-rent areas. They offered limited store services and relied on nationally advertised brands as opposed to private labels. Over time, supermarkets grew in importance and by 1970 had become primary food distributors, replacing smaller grocery stores.

Over the past couple of decades, food retailing has undergone another transition with the rapid growth of alternative retail formats that have entered the grocery business on a large scale. These formats range from value-oriented retailers (e.g., supercenters and price clubs) that are typically significantly larger in size compared with supermarkets to smaller high-end specialty stores (e.g., Trader Joe's, Whole Foods) that provide consumers with upscale product offerings. For example, Wal-Mart, which was hardly a player in the food sector until the mid-1990s is currently the largest supermarket chain in the country. Similarly, Whole Foods has grown at a rate of 20% over the past five years in an otherwise stagnant industry, and the current market value of Whole Foods is nearly the same as Safeway Inc., the third-largest traditional supermarket chain (Whole Foods Market 2006). As several industry analysts point out, there is a growing shopper trend of moving toward either of the two sides of the price spectrum: super-cheap one-stop-shopping format versus high-end organic format targeting specialized niches (Gogoi 2006). Changing consumer habits have prompted even traditional supermarkets to operate stores of different formats. For example, Kroger, Inc. operates high-end stores (e.g., Fresh Fare), traditional supermarkets (e.g., Kroger), and price-impact warehouse stores (e.g., Food4Less).

The changing competitive structure of the industry has also generated academic interest in issues related to different store formats. Messinger and Narasimhan (1997) and Lal and Rao (1997) have presented theoretical models to understand the competition between everyday low pricing (EDLP) and Hi-Lo operators and theoretical explanations for the general movement toward EDLP as a positioning strategy. On the empirical side, researchers have explored issues such as the relationship between household shopping behavior and store preference. For example, Bell and Lattin (1998) and Bell et al. (1998) study the effects of consumer- and shopping-related characteristics such as basket size on store choice between EDLP and Hi-Lo stores. Fox et al. (2004) study consumer expenditures at different retail formats as a function of shopper characteristics and marketing mix variables.[1] Unlike these papers that look at consumer store choice, our primary interest in this paper is on understanding the market structure *within* a product category and how the interbrand rivalries differ across store formats. In particular, our focus is on understanding how consumer perceptions about a particular brand, say, Folger's coffee, differs across retail formats. Similarly, we are interested in analyzing competitive structure in a product category across retail formats and the degree to which two brands—say, Folger's and Maxwell House—are perceived to be close substitutes in a large warehouse-type store versus a high-end grocery store.

Perceptions about a particular brand and the competitive interactions across brands could vary across retail formats because of consumer- and retailer-related factors. For instance, consumers that frequent a particular format may differ in terms of their demographics, shopping characteristics, preferences for different brand attributes, and sensitivity to price promotions. Similarly, different retail formats could vary in terms of marketing mix activities for a particular brand or category, product assortment, presence and positioning of store brands, and so forth. For example, the role of the category or a particular brand in the category could vary across formats on whether it is treated as traffic or as a profit generator. Similarly, a brand may be perceived by consumers to be an expensive or cheap alternative depending on the presence of other premium or value brands at the store. Given the evolution of the food industry as a three-tier system on price and quality, understanding the perceptions and relative positioning of the brand across retail formats could be important from a managerial perspective as it not only dictates the current marketing activities but could also have longer-term implications in terms of brand repositioning and product line extensions.

Our approach to understanding these issues is to build on the literature on internal market structure analysis that use revealed preference data (Elrod 1988, Chintagunta 1994, Elrod and Keane 1995, Erdem 1996, Erdem and Winer 1999). These methods infer both the brand attributes and consumer preferences for those attributes by exploiting how heterogeneous consumers' preferences for brands in the marketplace are correlated across those brands. In the estimation of these models, researchers impose a factor structure on the covariance matrix of preferences to provide brand maps. Our modeling approach extends this literature in several directions. First, we use a hierarchical structure to incorporate consumer observables (e.g., demographics) and allow for valuation of brand attributes to be correlated with marketing mix variables. More important, by allowing the loading matrix of brand attributes to be retail format specific, we can recover market maps specific to each format. Finally, to allow for systematic differences in consumers that prefer a particular retail format, we estimate a two-stage model of store choice followed by brand choice. In the empirical application, both store- and brand-choice models are estimated jointly using a Markov chain Monte Carlo (MCMC) method. Our modeling approach is appealing as we can obtain brand maps for different retail formats while controlling for systematic differences in the short-term marketing mix activities at these stores and the self-selection of households that frequent a particular retail format.

The data for this study come from a large Midwestern city. We observe detailed purchase histories for over 500 households for a period of two years. These households are observed to make purchases from five supermarkets: two high-end grocery

[1] Related literature looks at the impact of traditional supermarkets on store sales because of a change in price format to EDLP (Hoch et al. 1994) or because of entry by large discount stores (Singh et al. 2006, Zhu et al. 2005).

stores with an average store size of 17,000 square feet, one traditional supermarket which is 39,000 square feet, and two larger warehouse-type stores with an average store size of 74,000 square feet. The stores differ significantly in terms of the demographics of the household that patronize each format and in terms of pricing and promotional activities. The two large stores in the data are positioned explicitly as EDLP operators. For the empirical application we use household panel data from two product categories: ice cream and paper towels. In both categories, we find large differences in terms of product assortment, market shares of different brands, presence and market shares of private labels, pricing, and promotional activities. In general, high-end grocery stores offer the smallest assortment and higher prices, whereas EDLP operators offer a larger assortment with many low-priced value brands. Regular grocery stores (representing typical supermarkets) are observed to offer the highest level of promotions and perform significantly better in terms of store brand market shares.

Parameter estimates show systematic differences in households that prefer particular store formats, with observed household characteristics explaining a significantly higher proportion of heterogeneity variance in format choice than in marketing mix sensitivities. More important, we find strong correlations between marketing mix sensitivities, store format preference, and unobserved brand attributes. These correlations translate into significant differences in market structure across retail formats and in the direction and size of preference vectors for unobservable brand attributes. Interestingly, we find a tight clustering of brands at the EDLP format, whereas brands are found to compete in distinct subgroups such as premium and fat-free at other stores. Consumer perceptions about individual brands and the perceived similarities between competitive sets of brands are found to vary significantly across formats. For both product categories in our empirical analysis, we find more intense brand competition at the EDLP format, which could be driven by factors such as the presence of low-priced alternatives, retailers' marketing mix activities, and self-selection of more price-sensitive households to this format. We use the estimated model to analyze the impact of a new product entry and show that failure to account for retail format effects significantly biases the impact of entry on some incumbent brand shares.

Overall, our results show significant differences in market structure across store formats because of differences in sensitivities to the marketing mix as well as self-selection of households that frequent different stores. Whereas a typical approach in the literature to study market structure using scanner panel data is to pool household purchase behavior across all stores in the market, our results suggest that such a strategy would not be appropriate if households in the data are observed to make purchases from different types of stores. As pointed out above, the grocery industry is evolving toward a three-tier structure on price and quality that differs significantly in terms of product assortments and marketing mix activities. In the current paper, our application concerns three types of grocery stores. The differences in customer base as well as marketing mix activities are likely to be magnified when comparing supermarkets with discount formats such as supercenters and high-end organic stores such as Whole Foods. In these situations, pooling data across stores could not only create bias in the measures capturing the causal environment but also provide a murky picture of the underlying market structure. Because an understanding of the competitive patterns among brands constitutes the necessary first step in developing short- and long-term marketing strategies, the approach developed in this paper can be quite useful from a managerial perspective.

The rest of the paper is organized as follows: In the next section, we present the model. Section 3 discusses the data used in the study, and empirical results are presented in §4. Section 5 concludes with a discussion on the limitation of current work and directions for future research.

2. Model

The existing methods for market structure analysis can be broadly divided into external analyses (which presume that the attributes as well as values of brands on the attributes are known to the researcher) and internal analyses (where the number of attributes and locations of brands on these attributes are solely driven by preferences or choices). Researchers have proposed methods using hierarchical (Ramaswamy and DeSarbo 1990, Kannan and Wright 1991) and nonhierarchical (Elrod 1988, Chintagunta 1994) approaches, using aggregate market share (Allenby 1989, Cooper 1988, Shugan 1987, Chintagunta et al. 2002) and individual panel (Elrod 1988, Chintagunta 1994, Elrod and Keane 1995, Erdem and Winer 1999) data.[2] Our focus in this paper is on internal market structure analysis using revealed preference choice data.

Elrod (1988) represents the first attempt in the literature to produce spatial brand maps using probabilistic models of purchase frequencies. He infers a spatial structure by assuming that consumer preferences are multinormally distributed and by decomposing the preferences into importance weights and

[2] Although the majority of the work in this literature focuses on consumer packaged goods, it can be applied to other settings (e.g., understanding brand equity in online auctions, as in Bradlow and Park 2007).

brand locations. Chintagunta (1994) uses a similar structure as Elrod (1988) but uses consumer panel data rather than purchase frequencies. In addition, he accounts for heterogeneity in importance weights using a latent class structure (Kamakura and Russell 1989). Elrod and Keane (1995) provide a probit-based extension to these papers. They compare their proposed approach against several existing models and also provide a comprehensive review of the literature on market structure. Erdem (1996) and Erdem and Winer (1999) further improve on the methodology by incorporating choice dynamics and allowing for intercategory relationships in attribute perceptions, respectively. Our modeling approach builds on these papers and allows for brand perceptions and utility weights to be retail format specific. In addition, we obtain brand maps for different retail formats while controlling for the short-term marketing mix activities specific to each store and the self-selection of households that frequent a particular retail format.

2.1. Brand Choice Model for Inferring Market Structure

Consider a brand choice model of the form

$$V_{ht} = \mu_h + X_{ht}\beta_h + \varepsilon_{ht}, \quad t = 1, \ldots, T_h;\ h = 1, \ldots, H, \tag{1}$$

where V_{ht} is a vector of utilities for J brands at purchase occasion t for household h, μ_h is a vector of brand intercepts or brand preferences, X_{ht} is a matrix of marketing mix variables, β_h are sensitivities to the marketing mix variables, and ε_{ht} is a vector of iid error terms.

We obtain a market structure model by factoring μ_h into unobservable brand attributes and preferences for these attributes:

$$\mu_h = \alpha\psi_h, \quad h = 1, \ldots, H, \tag{2}$$

where α is a $J \times M$ matrix of unobservable brand attributes (or "brand perceptions") and ψ_h is an M vector of household h's preference weights. Note that besides the market structure papers mentioned above, a number of researchers have used a similar factor structure to capture heterogeneity in a parsimonious fashion in logit (e.g., Erdem and Keane 1996, Chintagunta et al. 2002) and probit (Keane 1997, Chintagunta 1998), as well as multicategory settings (Erdem 1998, Singh et al. 2005).

We make the following assumptions about heterogeneity in preference weights and marketing mix sensitivities:

$$\psi_h = \Pi_\psi z_h + \nu_{h\psi}, \tag{3}$$

$$\beta_h = \Pi_\beta z_h + \Gamma\nu_{h\psi} + \nu_{h\beta}, \tag{4}$$

$$\nu_{h\psi} \sim \mathrm{N}(0, I_M), \tag{5}$$

$$\nu_{h\beta} \sim \mathrm{N}(0, \Omega_\beta), \tag{6}$$

where z_h is a vector of household characteristics (e.g., demographics) impacting preferences and marketing mix sensitivities. This specification extends the current market structure literature in two ways. First, we use a hierarchical structure to incorporate household demographics. This allows a researcher to infer whether observable household characteristics can explain variation in brand preferences along the dimensions of the map. Second, we allow brand preferences to be correlated with marketing mix sensitivities through the Γ matrix.

Upon estimation, the elements of the loading matrix are usually plotted in a brand map displaying the locations of each brand in the market along with a vector displaying the market's mean preference for the (unobservable) brand attributes. Brands located perceptually close to each other compete more closely compared to brands located far from each other.

2.2. Controlling for Retail Format

To control for retail format, we generalize the model by allowing the loading matrix α to be retail format specific. In particular, we assume that utility conditional on a visit to retail format f is

$$V_{htf} = \alpha_f\psi_h + X_{htf}\beta_h + \varepsilon_{htf}. \tag{7}$$

The row dimension of α_f, J_f, varies across retail formats (because the range of brands varies across retail formats). To keep the model somewhat parsimonious, we restrict the marketing mix sensitivities *within* households to be constant across formats. In other words, a given household will not be more or less price or deal sensitive in one format versus another.[3] However, the model will allow for the *average* price sensitivity across households to vary across formats (see below). We assume that the ε_{htf}s are Type I extreme value distributed, so that the resulting brand choice model is of the logit form.

The model described above is quite flexible in that it allows the underlying market structure to vary across store formats because of factors such as differences in product assortment and other marketing mix activities. It is also conceivable that households with particular observed or unobserved characteristics may prefer to shop at particular formats. For instance, larger families with limited budgets may prefer cheaper warehouse-type stores, whereas single-member consumers may prefer a more convenient high-end store. To account for such selection effects, we add a simple model of retail format choice. Note that our objective here is not to build and analyze a full-fledged model of store choice. Rather, we want

[3] It should be noted that households' price *elasticity* may vary across formats because brand perceptions vary across formats.

to control for the possibility that households that frequent a specific format are—on average—different in their (ψ_h, β_h) configurations than the overall population. In the empirical application, we analyze three types of store formats and assume that latent utility for retail format f is

$$U_{ht}(f) = \eta_{fh} + W_{fh}\gamma + \omega_{fht}, \quad f = 1, 2, 3, \tag{8}$$

where

$$\eta_h \equiv (\eta_{1h}, \eta_{2h}) = \Pi_\eta z_h + \Lambda_\psi \psi_h + \Lambda_\beta \beta_h + \nu_{h\eta}, \tag{9}$$

$$\nu_{h\eta} \sim \mathrm{N}(0, \Omega_\eta). \tag{10}$$

In the empirical section, we let $f = 1$ denote the "EDLP" format, $f = 2$ the "Regular" format, and $f = 3$ the "High-End" format. This specification allows format preferences η_h to be correlated with household demographics z_h and, more important, attribute preferences ψ_h and marketing mix sensitivities β_h. We set $\eta_{3h} = 0$ for identification purposes. The array W_h contains format-specific time-invariant factors determining store choice. In our application, this will be a dummy indicating close proximity to the particular format. We assume that the ω_{htf}s are Type I extreme value distributed, so that the format choice probabilities are of the logit form.

The complete model consists of (7)–(10) plus (3)–(6). An implication of the model is that both attribute preferences, marketing mix sensitivities, and retail format preferences are correlated. As such, the model captures the notion that households that have a preference for a specific retail format may systematically differ in their ψ and β vector compared to the population mean. To see this, note that by Bayes rule the distribution of attribute preference and marketing mix sensitivity for households homogenous with respect to $Z = z$ and $W = w$ and choosing format f is

$$p(\psi_h, \beta_h \mid z, w, D_f^F = 1) = \frac{\Pr(D_f^F = 1 \mid z, w, \psi_h, \beta_h)\, p(\psi_h, \beta_h \mid Z = z)}{p(z, w, D_f^F = 1)}, \tag{11}$$

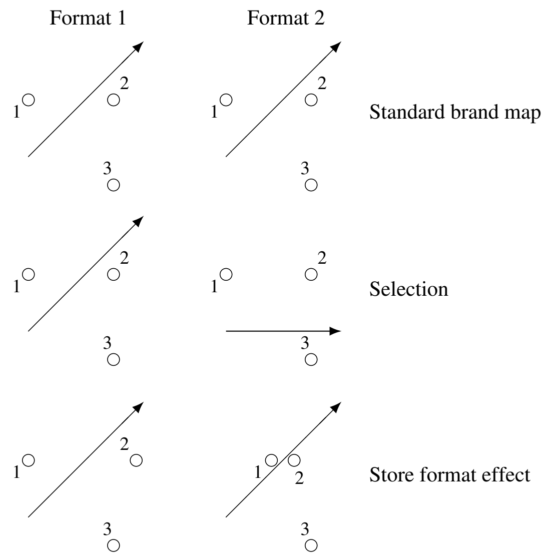
where D_f^F is a dummy equal to 1 if format f is chosen. The standard assumption in the existing literature is that the distribution of attribute preferences and marketing mix sensitivities is invariant across format choice; i.e.,

$$p(\psi_h, \beta_h \mid z, w, D_f^F = 1) = p(\psi_h, \beta_h \mid z, w), \quad \text{for } f = 1, 2, 3. \tag{12}$$

From Equation (9), we can see that our model specializes to this case when

$$\Lambda_\psi = \Lambda_\beta = 0; \tag{13}$$

Figure 1 Illustration of Brand Map Assumptions



i.e., choice of retail format is uncorrelated with preferences for the unobservable brand attributes and marketing mix sensitivities. These derivations demonstrate that unless this is the case, the average ψ level as well as the average marketing mix sensitivity will differ across retail formats. If a researcher wrongly assumes Equation (13), the resulting brand map and derived implications for competition will be inaccurate.

The extensions in the model described above over the existing literature is illustrated in Figure 1. The standard approach is based on estimating a marketwide map. This map will implicitly assume that brand positions do not vary across retail formats (top row in Figure 1), and the average preference vector for the two underlying attributes is assumed to be the same across formats. Our first extension is to allow the average preference (and marketing mix) vector to change across formats. This selection effect is illustrated in the maps in the middle row of Figure 1, and this effect will be present if $\Lambda_\psi \neq 0$. The selection effect alone can imply that even though brand perceptions are the same across formats, market structure may be very different across formats. For example, for the case shown in the middle row in Figure 1, brand 3 competes mainly with brand 1 in format 1 but competes with brand 2 in format 2. This is a pure selection effect and does not stem from underlying different perceptions across formats. Our second extension is to allow brand perceptions to vary across formats (the bottom row in the figure). Even if there are no selection effects, this may cause a different market structure across formats. For example, for the case shown in the bottom row in Figure 1, brand 1 competes mainly with brand 3 in format 1 but competes with brand 2 in format 2. In this case, it is assumed that there are no selection effects. In reality, there may be both selection and perception effects across formats (i.e., a mix of the cases in the middle and bottom row) and the model described above allows for this.

2.3. Estimation

In the empirical section, we only consider two-dimensional maps.[4] To identify the format-specific maps, we set one brand at the origin at the map; i.e.,

$$\alpha_{fj',1} = \alpha_{fj',2} = 0, \quad \text{for } f = 1, 2, 3. \tag{14}$$

In addition, we place one brand at the origin of the second dimension; i.e.,

$$\alpha_{fj^*,2} = 0, \quad \text{for } f = 1, 2, 3. \tag{15}$$

To aid in the interpretation of the resulting maps, we choose to put the same brand at the origin for all three formats and the same brand at the origin of the second dimension for all three formats. A final issue in the identification of the maps and implied distributions of preferences deals with the problem of *reflection* in factor analytic models. This refers to the simple fact that all loadings and factor scores can be multiplied by minus one without changing the likelihood. We deal with this by restricting "one factor one" loading to be positive for one brand, and one "factor two" loading to be positive for another brand.

For estimation, we take a Bayesian approach and rely on an MCMC algorithm to simulate posteriors of parameters of interest. As discussed in Allenby and Rossi (1999), Bayesian procedures are well suited for these models, especially when one is interested in making inference at the individual level. These methods have become standard in the brand choice literature in recent years. Interested readers are referred to Gelfand and Smith (1990) for a general overview of these methods, McCulloch and Rossi (1994) and Rossi et al. (1996) for multinomial probit models, and Allenby and Lenk (1994) for models where the error term follows an extreme value distribution. However, because our model is an extension of the previous models, we have provided a detailed description of the sampling algorithm in the appendix. Because of slow convergence of the factor loadings, we were forced to run the MCMC algorithm for abnormally many iterations. We ran the algorithm for a total of 1.1 million iterations, dropping the first 100,000 draws and using every 100th of the remaining draws for computing posterior summaries. This was necessary because of the high autocorrelations for the factor-loading parameters.

[4] Our estimation algorithm can handle higher-dimensional maps, but these generally do not yield interpretable results.

Table 1 **Format Summary Statistics (548 Households)**

	High end	Regular	EDLP
Size (sq. feet)	17,000	39,000	74,000
Loyal[1] (%)	14	21	47
Family size	1.46	1.76	2.99
Income ($)	49,400	29,500	36,600
White (%)	95	57	75
Age	61	62	50
Basket size ($)	17	20	35

[1] A household is labeled loyal if at least 75% of purchases occur at the format.

3. Data

The data for this study are drawn from a metropolitan city in the Midwest and contain detailed purchase histories for over 500 households for a two-year period in the early 1990s.[5] There are five supermarkets in the data set: two high-end grocery stores with an average store size of 17,000 square feet, one traditional supermarket which is 39,000 square feet, and two larger warehouse-type stores with average store size of 74,000 square feet. Besides the size, the stores also differ in terms of their pricing strategy with the two large stores explicitly positioned as EDLP operators. For the purpose of this paper, we combine the observations from the two high-end stores as they belong to the same chain. We also combine observations from the two larger EDLP stores (even though they belong to separate chains) as they are found to be very similar in terms of product assortment and marketing mix activities. For the remainder of the paper, we shall refer to these three formats as high end, regular, and EDLP.

In Table 1, we report the average demographics for the households that patronize each store type, where a household is labeled as loyal if at least 75% of its purchases occur at the format. Looking at the numbers presented in Table 1, we find that a large proportion of the households seem to be loyal to a particular store format, with only 18% of the total households classified as switchers based on the criterion described above. A few other patterns are also apparent. First, as reported in Bell and Lattin (1998), the basket size for consumers frequenting the EDLP format is significantly larger (more than double that at a high-end store). We also see significant differences in the demographic profile of the households that patronize different stores. For instance, the households loyal to the high-end store have higher income levels, smaller families, and are primarily white.

To study the differences in market structure across these three formats, we use data from two product

[5] These data have been used before in Bell and Lattin (1998), among others. We thank David Bell for providing the data.

Table 2 Summary Statistics, Ice Cream Category (496 Households)

Format	Brand	Size	Fat-free	Price	Share (%)	Feature (%)	Display (%)
EDLP	Sealtest	64	0	2.61	2	3	5
EDLP	Breyers	64	0	3.26	12	4	2
EDLP	Breyers	64	1	2.72	1	0	0
EDLP	Healthy Choice	64	1	3.81	2	5	2
EDLP	Simple Pleasures	16	1	9.18	4	3	1
EDLP	Häagen-Dazs	16	0	8.91	11	8	2
EDLP	Ben & Jerry's	16	0	9.60	8	4	1
EDLP	Dreyer's/Edy's	32	0	4.76	2	0	0
EDLP	Dreyer's/Edy's	64	0	3.63	6	9	4
EDLP	Dreyer's/Edy's	32	1	5.01	2	0	0
EDLP	Dreyer's/Edy's	64	1	3.63	4	5	3
EDLP	Country Charm	64	0	3.11	5	13	6
EDLP	Kemps	64	0	2.28	9	12	7
EDLP	Kemps	160	0	1.39	7	3	0
EDLP	Kemps	64	1	2.49	4	3	1
EDLP	Value Pak	160	0	1.26	23	7	5
Regular	Sealtest	64	0	3.02	4	21	2
Regular	Breyers	64	0	3.52	10	28	1
Regular	Breyers	64	1	3.23	6	22	1
Regular	Healthy Choice	64	1	4.23	2	8	0
Regular	Simple Pleasures	16	1	10.13	5	5	0
Regular	Häagen-Dazs	16	0	10.36	20	14	2
Regular	Ben & Jerry's	16	0	11.08	7	7	1
Regular	Dreyer's/Edy's	32	0	5.52	3	1	0
Regular	Dreyer's/Edy's	64	0	4.05	3	13	5
Regular	Dreyer's/Edy's	32	1	6.51	0	0	0
Regular	Dreyer's/Edy's	64	1	4.07	3	6	3
Regular	Country Charm	64	0	4.08	8	18	1
Regular	Dean's Light	64	1	3.26	4	13	0
Regular	Fieldcrest	64	0	1.84	13	11	1
Regular	Private label	32	0	4.82	6	12	0
Regular	Private label	64	0	3.02	6	12	1
High end	Sealtest	64	0	3.06	7	18	0
High end	Breyers	64	0	3.92	10	17	0
High end	Breyers	64	1	3.62	6	7	0
High end	Healthy Choice	64	1	4.24	2	4	0
High end	Simple Pleasures	16	1	9.79	14	4	0
High end	Häagen-Dazs	16	0	10.64	28	6	0
High end	Ben & Jerry's	16	0	10.63	20	3	0
High end	Private label	64	0	1.84	8	14	0
High end	Sealtest	64	1	3.28	4	3	0

categories: ice cream and paper towels. Summary statistics for the ice cream category are reported in Table 2. There are a total of 20 unique products in the category, which can be defined as a combination of brand, size, and fat-content attributes. Table 2 shows the market shares, price, and promotional activities at each format. In general, there are significant differences across store formats in terms of product assortments. The three premium brands (Simple Pleasures, Häagen-Dazs, and Ben & Jerry's) account for 62% of the market share at high-end grocery stores compared with only 23% at the EDLP format. On the other hand, the value brands (Kemps and Value Pak) capture 43% of the total category sales at the EDLP store; these brands are not available at the two other store formats. Private label is most prominent at regular grocery stores, followed by high-end stores. There are also significant differences in terms of prices with brands at the high-end store priced significantly higher compared with the EDLP format. Note that all prices have been normalized to 64 ounces. The last three columns show the promotional activities, which tend to be lower at the EDLP format.

In Table 3, we report summary statistics for the paper towel category. For this category, because there are minute differences in per-roll prices (single and three rolls are the most dominant sizes in our data), we aggregate the universal price codes to brand level. Unlike the ice cream category, we find the product assortment for paper towels to be relatively similar across retail formats. However, there are significant differences in terms of market shares across brands as well as marketing activities. The market shares are fairly evenly distributed across brands at high-end

Table 3 Summary Statistics, Paper Towel Category (502 Households)

Format	Brand	Price	Share (%)	Feature (%)	Display (%)
EDLP	Bounty	0.93	14	5	3
EDLP	Brawny	0.81	9	9	13
EDLP	Scott	0.87	20	11	12
EDLP	Viva	0.87	10	4	4
EDLP	Sparkle	0.63	7	7	16
EDLP	Mardi Gras	0.62	20	11	45
EDLP	Private	0.47	15	2	4
EDLP	Hi-Dri	0.68	6	5	11
Regular	Bounty	1.16	23	17	19
Regular	Brawny	0.83	9	19	21
Regular	Scott	0.97	14	19	20
Regular	Viva	1.01	10	13	14
Regular	Sparkle	0.75	5	13	12
Regular	Mardi Gras	0.77	7	25	26
Regular	Private label	0.57	33	11	16
High end	Bounty	1.22	20	4	3
High end	Brawny	0.90	14	19	18
High end	Scott	1.07	17	6	5
High end	Viva	1.02	16	8	6
High end	Sparkle	0.72	17	26	27
High end	Mardi Gras	0.68	14	4	8
High end	Private label	0.77	2	1	1

stores with the exception of private labels which have a negligible share. On the other hand, private labels capture one-third of the market in regular grocery stores, followed by Bounty. At the EDLP format, the highest-selling brands are Scott and Mardi Gras, capturing 20% share each. In terms of paper towel prices, we find a similar pattern as the ice cream category, with the EDLP format offering the lowest prices followed by regular grocery stores. It is also interesting to see the promotional activities of brands across retail formats. Whereas regular stores seem to distribute promotional efforts evenly across brands, feature and display activities are concentrated to select brands at the other two store formats. The EDLP store promotes Mardi Gras extensively, but promotional activity is significantly higher for Brawny and Sparkle at high-end stores.

4. Results

We next present empirical results from the model and data described above to see how the differences in consumer self-selection, product assortment, and marketing activities translates to differences in consumer brand perceptions across store formats. We begin with a discussion on the parameter estimates from the hierarchical model and the correlations between marketing mix sensitivities, store format preference, and the unobserved brand attributes. This is followed by a discussion on brand maps specific to each retail format. We conclude this section with an application to new product entry.

4.1. Parameter Estimates

In Table 4, we report posterior estimates of the coefficients in the hierarchical model. Table 4 shows the posterior mean and standard deviation of the Π coefficients in (3), (4), and (9) and is organized as follows. All numbers reported in boldface are statistically "significant."[6] For each category, the first three rows show the marketing mix parameters, followed by the unobservable brand attributes. The last two rows represent the store format intercepts (high-end intercept is normalized to zero). The first column in Table 4 shows the intercept population estimates, followed by demographic interactions representing observed heterogeneity. We note that all the observable household variables have been normalized to have mean zero except for "family size" and the indicator variable "race." Finally, the last column, z_{frac}, shows the fraction of overall heterogeneity explained by the observed household characteristics.

We first comment on the results for the ice cream category. The mean marketing mix parameters have the expected sign, and the mean level of price sensitivity is seen to be higher for large and less affluent families and lower for "large basket" households (the basket size variable is bordering on significance). In general, observable household characteristics are found to explain a relatively small proportion (12% for feature, 18% for display, and 22% for price) of the overall variation in heterogeneity in the marketing mix sensitivities. Household characteristics explain even less of the variation in preferences for the two unobservable brand attributes.[7] This is not surprising as previous research has found demographic variables to be poor predictors of preferences estimated on scanner data (e.g., Gupta and Chintagunta 1994, Rossi et al. 1996), although Horsky et al. (2006) show the significant role of observed heterogeneity if one had information on consumers' brand preferences from external (e.g., survey) sources. For the format choice intercepts, we see a (slightly) bigger role for the observable characteristics. Larger, younger, and less affluent families are more likely to frequent EDLP stores. In addition, large basket households are much more likely to prefer EDLP stores to the other formats. This is consistent with the findings reported in Bell and Lattin (1998).

The results for the paper towel category more or less mirror those of ice cream. The mean marketing mix sensitivities have the expected sign and are precisely estimated. As for ice cream, larger and less affluent households are seen to be more price sensitive on average, whereas large basket households

[6] A 95% central posterior interval does not contain zero.

[7] We interpret the unobserved brand attributes in the next paragraphs.

Table 4 Hierarchical Model Estimates

	Const.	Famsize	Log(inc)	Race	Log(age)	Log(bsize)	z_{frac}[a]
Ice cream							
β_{price}	−**1.68** (0.16)	−**0.80** (0.13)	**0.97** (0.16)	0.52 (0.36)	−0.18 (0.17)	1.05 (0.57)	0.23
β_{feature}	**1.06** (0.10)	−0.10 (0.09)	**0.44** (0.11)	−0.09 (0.23)	0.06 (0.11)	−**0.78** (0.38)	0.12
β_{display}	**1.84** (0.21)	0.07 (0.16)	−**0.47** (0.20)	**0.89** (0.43)	−**0.44** (0.21)	−1.30 (0.73)	0.18
ψ_1	0.12 (0.06)	−0.09 (0.05)	0.11 (0.06)	−0.19 (0.13)	−0.03 (0.06)	0.21 (0.21)	0.02
ψ_2	−**0.13** (0.06)	−0.04 (0.05)	−0.09 (0.06)	−**0.46** (0.13)	−0.06 (0.06)	−**0.58** (0.21)	0.10
η_{EDLP}	0.40 (2.55)	**2.99** (0.61)	−**2.31** (1.05)	−0.43 (1.86)	−1.81 (0.89)	**9.44** (2.81)	0.27
η_{reg}	1.77 (2.12)	0.08 (0.43)	0.39 (0.78)	−**4.08** (1.40)	0.23 (0.66)	−2.24 (2.10)	0.14
Paper towels							
β_{price}	−**1.58** (0.27)	−**0.69** (0.21)	**0.75** (0.34)	0.43 (0.56)	0.001 (0.02)	**1.15** (0.42)	0.07
β_{feature}	**1.36** (0.12)	−0.09 (0.10)	0.14 (0.15)	0.32 (0.25)	−0.006 (0.01)	−0.08 (0.19)	0.02
β_{display}	**1.02** (0.10)	0.02 (0.09)	0.17 (0.14)	−0.17 (0.23)	−0.002 (0.01)	−0.11 (0.18)	0.01
ψ_1	−**0.30** (0.07)	0.01 (0.05)	−0.12 (0.08)	−0.16 (0.13)	0.0015 (0.004)	0.14 (0.10)	0.02
ψ_2	0.10 (0.07)	0.03 (0.05)	−0.10 (0.07)	−0.18 (0.13)	−0.0005 (0.004)	−0.07 (0.09)	0.02
η_{EDLP}	**4.35** (2.06)	**4.32** (0.72)	−**3.92** (1.01)	1.90 (1.63)	−**0.11** (0.05)	**5.42** (1.28)	0.26
η_{reg}	**4.50** (1.31)	0.18 (0.47)	0.21 (0.62)	−**2.50** (1.03)	0.03 (0.03)	−**2.20** (0.80)	0.13

[a] z_{frac} is the fraction of overall variation in unobservable heterogeneity explained by the observables.

are less price sensitive. For this category, we observe an even lower role of demographics in explaining the overall variation in marketing mix sensitivities and the unobserved brand attributes. The signs and relative explanatory power of the observed household characteristics in explaining heterogeneity for store format choice intercepts are quite similar to use in the ice cream category. We see a similar pattern with larger, less affluent, younger, and large basket households preferring EDLP formats. Note that the format choice equation also included a dummy for close proximity to each format. Not surprisingly, the parameter estimates of this coefficient (γ) are found to be positive and highly significant (see Table 5).

We next turn our attention to the estimates of the covariance parameters Λ that are presented in Table 6. Note from Equation (9) that format preference η_h is a function of both Λ_ψ (brand attributes) and Λ_β (price, feature, display). Recall that if these parameters are zero, then the unobservable components of households' attribute preferences, marketing mix sensitivities, and format choices are uncorrelated. Table 6 shows that this is not the case for both categories. For both categories, there appears to be significant dependence between preferences for retail format and the second attribute. In addition, price parameter is negatively correlated with preference for the EDLP format in the ice cream category. More price-sensitive households have stronger preference for the EDLP format. Table 7 shows further evidence for significant correlations through posterior estimates of the household-level parameter correlations. For ice cream, we see significant correlations between format preferences and price sensitivity and preference for the second attribute. The negative correlation between ψ_2 and the format choice intercept implies that EDLP loyal households will on average have a smaller preference for the second attribute than regular and high-end format households. Similarly, EDLP and regular format households are more price sensitive than high-end loyal customers. For paper towels, we find both attributes to be correlated with retail format preferences. Both attribute preferences are positively correlated with EDLP and regular format preferences. This means that high-end loyal customers tend to prefer brands with more negative values on the two attribute axes. As for ice cream, we see that price sensitivities are negatively correlated with regular format preferences (the correlation with EDLP is zero). Overall, Table 7 shows strong support for the model allowing $\Lambda_\beta \neq 0$ and $\Lambda_\psi \neq 0$.

Additional evidence for $\Lambda_\beta \neq 0$, $\Lambda_\psi \neq 0$ is provided in Table 8. This table shows the log marginal density for each model computed using the standard

Table 5 γ Estimates

	γ
Ice cream	**7.39** (0.75)
Paper towels	**9.68** (0.95)

Table 6 Λ_ψ and Λ_β Estimates

	Λ_ψ		Λ_β		
Ice cream					
η_{EDLP}	−0.55 (0.65)	−**2.46** (0.70)	−**0.74** (0.36)	1.17 (1.01)	0.89 (1.33)
η_{reg}	−0.51 (0.39)	−0.48 (0.39)	−0.17 (0.24)	−0.15 (0.68)	0.06 (1.13)
Paper towels					
η_{EDLP}	1.35 (1.09)	**1.80** (0.86)	−0.18 (0.39)	1.98 (1.10)	0.94 (1.03)
η_{reg}	0.04 (0.62)	**1.54** (0.54)	−0.17 (0.23)	−0.24 (0.71)	0.06 (0.65)

Table 7 Household-Level Parameter Correlations

	ψ_1	ψ_2	β_{price}	β_{feature}	β_{display}	η_{EDLP}
Ice cream						
ψ_2	0.02 (0.05)					
β_{price}	**0.12** (0.04)	−0.02 (0.04)				
β_{feature}	0.08 (0.05)	0.03 (0.05)	**0.57** (0.06)			
β_{display}	−0.05 (0.05)	0.08 (0.05)	0.05 (0.15)	0.12 (0.15)		
η_{EDLP}	−0.10 (0.06)	−**0.26** (0.06)	−**0.29** (0.05)	−0.13 (0.09)	0.13 (0.17)	
η_{reg}	−0.11 (0.07)	0.01 (0.07)	−**0.22** (0.07)	−0.09 (0.10)	−0.02 (0.25)	0.16 (0.13)
Paper towels						
ψ_2	0.00 (0.06)					
β_{price}	0.02 (0.06)	−0.04 (0.05)				
β_{feature}	0.04 (0.05)	−0.04 (0.04)	**0.66** (0.06)			
β_{display}	−0.06 (0.25)	0.08 (0.07)	0.03 (0.11)	−0.02 (0.11)		
η_{EDLP}	**0.18** (0.08)	0.11 (0.11)	0.04 (0.08)	0.13 (0.09)	0.08 (0.10)	
η_{reg}	0.05 (0.11)	**0.28** (0.08)	−**0.23** (0.10)	−0.17 (0.12)	0.03 (0.13)	−0.10 (0.14)

Newton-Raftery estimator. We see that the full model (allowing $\Lambda_\beta \neq 0$ and $\Lambda_\psi \neq 0$) is favored over the no-selection mode. Taken together, Tables 7 and 8 provide strong evidence for the full model. Table 8 also gives an indication of how much we are straining the brand choice data by imposing the two-dimensional factor structure on the brand choice intercepts. We computed the log marginal density for the brand choice part of the full model and compared this to a brand choice model with unrestricted intercepts. Table 8 shows that we do lose some "fit" by imposing the factor structure—especially for the ice cream category. This is not surprising given the large number of products used in the empirical application. One possibility would be to go to a factor dimension higher than two. We did not pursue this because of the increased difficulty in estimation and—more important—the increased difficulty in interpreting the resulting higher-dimensional maps. In our new product application below, we will use only the paper towel category, and we view the two-dimensional maps as capturing the two most important factors (or principal components) of the free brand intercepts.

4.2. Brand Maps Across Retail Formats

Our discussion on the correlations above indicates that brand perceptions are likely to be influenced by the selection of households that prefer a particular store format. In addition, as discussed in §3, we observe a large variation in product assortment and the marketing mix environment across store types. We next analyze how these factors translate to differences in market structure by developing brand maps specific to each format.

Table 8 Model Comparison, Log Marginal Density

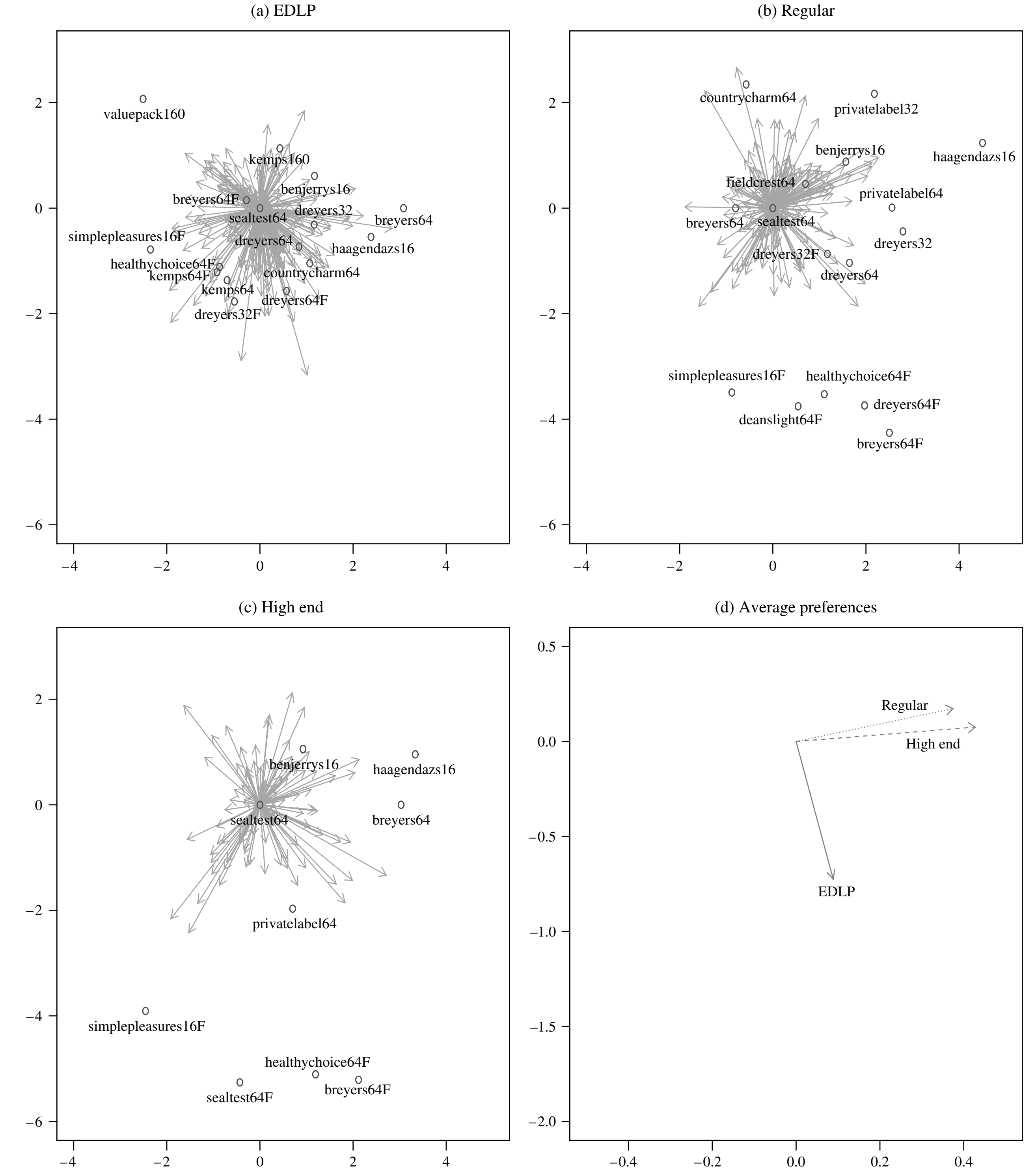
	Ice cream	Paper towels
Full model	−9,608.8	−9,749.5
No selection	−9,616.7	−9,813.5
Brand choice model only		
Factor structure	−8,399.5	−8,668.2
Free intercepts	−8,217.6	−8,614.2

The estimated brand maps for the ice cream category are shown in Figure 2. The arrows on each plot show the direction and size of preference vectors for the two unobservable brand attributes for households that are loyal to that format.[8] Note that these household-level estimates are a natural by-product of the MCMC algorithm used in estimation. The preference vectors indicate that within each format there are segments of households with strong preferences for brands in any region of the plot. For example, for the EDLP format, there are segments of households with preferences for brands in the north direction, segments who prefer brands in the east direction, and yet other segments who prefer brands in the west or south direction. A similar pattern is found for the other formats. The degree of selection is best seen from the average preference vectors for each format (shown in Figure 2(d)). On average, preferences increase in the east and south direction of the plot, with EDLP loyal customers on average having the strongest preference for brands located in the south/southeast region of the plot. Regular and high-end loyal customers on average favor brands in the east direction. As is evident in the format-specific maps, these averages are formed over very heterogeneous customers.

Looking at the brand maps across retail formats, we find a number of similarities. The premium brands in the category, Ben & Jerry's and Häagen-Dazs, are found close to each other. In addition, note that brands in the south/southwest region of the maps tend to be fat-free whereas brands in the north/northeast region

[8] This was defined as households where more than three quarters of trips were made to the corresponding format.

Figure 2 Brand Maps for Three Store Formats, Ice Cream Category



all are non-fat-free. So this dimension of the maps seems to be a fat/fat-free dimension. The interpretation of the fat-free dimension in the map seems to hold fairly consistently across all formats. However, there are also systematic differences across these stores. For the EDLP map, we see a fairly tight clustering of all brands. On the other hand, competition in other formats seems to occur in distinct subgroups. For instance, at high-end stores, we find distinct clusters for the two premium brands, midrange non-fat-free brands Sealtest and Breyers, and a cluster of fat-free alternatives. A similar pattern occurs at regular stores. Overall, the strong cluttering of brands at the EDLP format suggests that brands compete much more intensely in this format. This could be because of the presence of several low-priced alternatives such as Kemps and Value Pak at this format along with our finding above that EDLP shoppers are more price sensitive.

The estimated brand maps for the paper towel category are shown in Figure 3. In this category, we observe fewer differences in the product assortment with Hi-Dri being the only brand that is available at the EDLP format but not at other stores. In the paper towel category, we again see significant selection effects: notice how the mass of the preference vectors move as we go from EDLP format to regular and high-end formats. Most of the households loyal to the high-end format prefer brands in the west/southwest region of the map. Households loyal to the regular format primarily prefer brands in the northwest region, whereas EDLP formats have segments who prefer brands in all regions. The differences across the three formats can be seen from the average preference vectors plotted in Figure 3(d). For this category, there are differences in both strength and direction of preference vectors.

Looking at the relative brand positioning across formats, we find a similar clustering effect in the EDLP format (with the exception of Scott) versus the regular and high-end format we saw in the ice cream category (although the effect is less extreme here). There are also several differences in the relative positioning of individual brands. Consider, for example, the Viva brand. In the regular and high-end format, this brand is positioned quite well, being the most preferred brand for the household segment with preferences toward the south region of the map. However, in the EDLP format, this brand competes closely with several brands including a cluster of value brands (private label, Sparkle, and Hi-Dri). Similarly, we find that Bounty is quite well positioned in regular grocery stores although it is perceived as similar to Brawny at other stores. From Brawny's perspective, its positioning at regular grocery stores seems problematic as it is perceived by consumers to be perceptually closer to the cheaper alternatives (Sparkle and Mardi Gras) than to Bounty. Not surprisingly, Bounty enjoys a 14% share advantage over Brawny at regular grocery stores compared to about 5% at the two other stores. Finally, looking at the position of the store brands at EDLP and regular stores (high-end stores have a negligible presence), we see that the private label of the EDLP store finds itself in the clutter of value brands, while regular supermarket chains have carved out a unique position closer to Scott. This positioning strategy seems to have worked well as the store brand captures a one-third market share in this store format.[9]

One feature of the maps that is hard to ascertain from Figures 2 and 3 is the degree to which the brand positions actually differ across format. The figures show only point estimates of each brand's location. In Figure 4, we have shown 100 posterior draws chosen at random for each of the four paper towel brands (Scott, Viva, Sparkle, and Mardi Gras). These figures clearly show the uncertainty associated with each brand's location. More important, they show that the maps clearly are different across retail formats. For example, the location of Sparkle is clearly different across formats (Figure 4(c)). This is direct empirical support for the need to allow perceptual brand maps to differ across retail formats.

4.3. Impact of New Product Entry

Our results above show significant differences in the underlying competitive structure across store formats. These differences will be important in applications such as analyzing the impact on market structure of new product or brand entry. In particular, given parameter estimates and information on the new product's position and marketing mix, the impact on existing brands may be analyzed. In the following, we wish to document the importance of controlling for retail format effects when analyzing the impact of new entry.

First, we define a *format configuration* as

$$\mathscr{C}_f = (J_f, \alpha_f, X_f).$$

The format configuration defines the number of brands in the format J_f, their positions (α_f), and an array of marketing mix values for each brand in the format (X_f). We can then derive household h's probability of choosing brand j (conditional on choice of format f) under configuration $\mathscr{C}_f$:

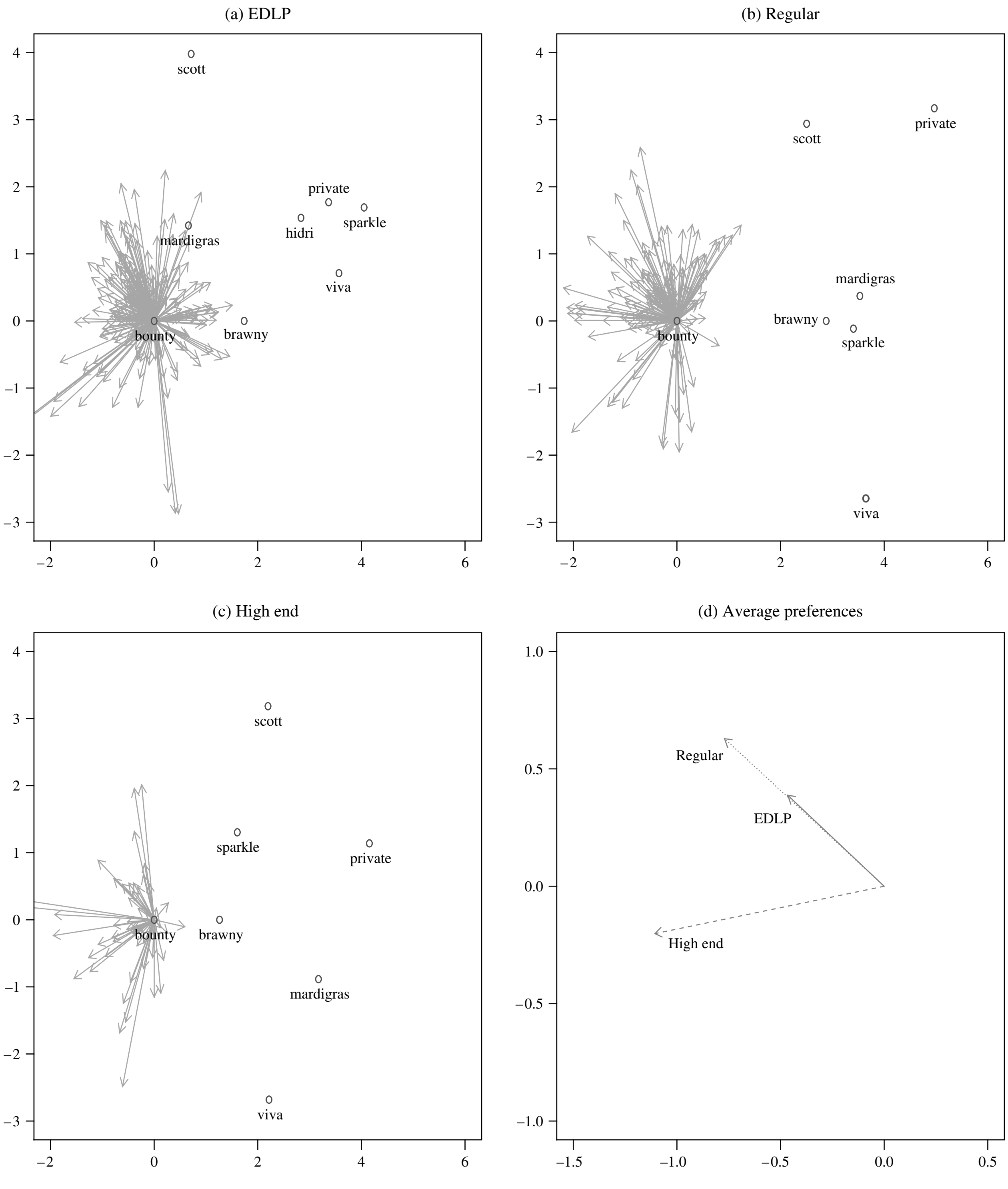
$$\Pr(D_{hj} = 1 \mid \alpha_f, \theta_h, D^F_{fh} = 1, \mathscr{C}_f) = \frac{\exp\{\alpha'_{fj}\psi_h + \beta'_h X_{fj}\}}{\sum_{j'=1}^{J_f} \exp\{\alpha'_{fj'}\psi_h + \beta'_h X_{fj'}\}}. \qquad (16)$$

From this, we can calculate the share of brand j in format f under configuration $\mathscr{C}_f$ by integrating over the $\theta_h = (\psi_h, \beta_h)$ distribution for format f loyal households:

$$s_{jf}(\mathscr{C}_f; \alpha_f) = \int \Pr(D_{hj} = 1 \mid \theta_h, D^F_{fh} = 1, \mathscr{C}_f) \cdot p(\theta_h \mid D^F_{fh} = 1)\, d\theta_h. \qquad (17)$$

[9] We also produced maps for each retail format using only the data for that format (maps are available from the authors). In terms of brand positions, these maps were, with a few exceptions, qualitatively similar to the maps above. However, note that this approach ignores the selection effects induced by households' retail format decisions. In other words, this approach assumes that the distribution of preferences for the two attribute dimensions of the map is identical across retail format. In light of the results above, this is clearly incorrect.

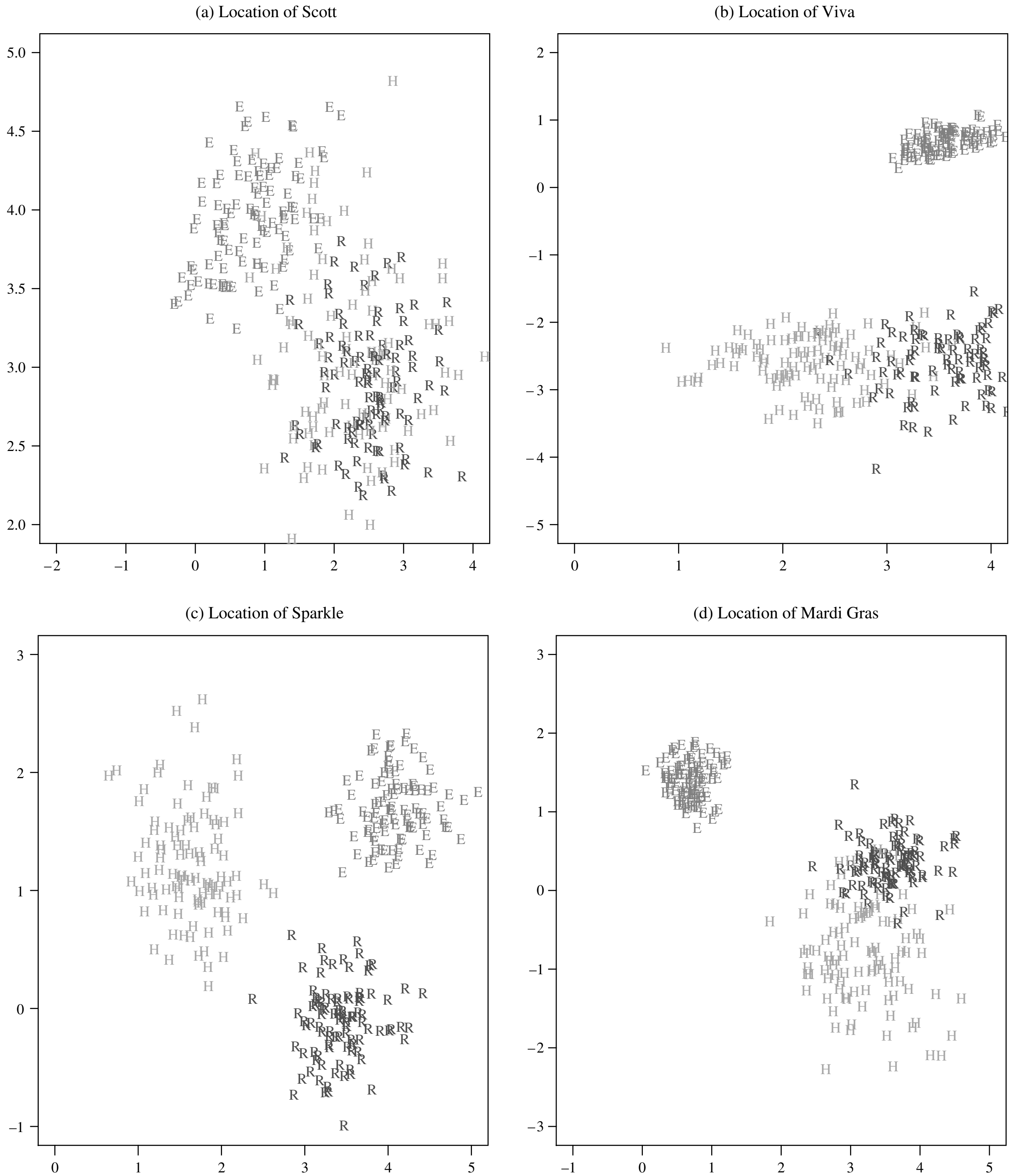
Figure 3 Brand Maps for Three Store Formats, Paper Towel Category



Note that this share depends on the matrix of unobservable brand perceptions α_f for format f. Rather than plugging in a point estimate of α_f, we integrate out these parameters in Equation (17) using their posterior distribution. This is easily done using the posterior draws of α_f obtained from our sampling algorithm and is quite standard in the literature (see, e.g., Rossi et al. 1996, Gonul and Ter Hofstede 2006).

It is then of interest to derive the market shares of one or several brands under different configurations. For example, a brand manager for a specific brand in the format might be interested in evaluating the outcome on his brand's share when a new brand enters the format. This can be done by defining a new configuration $\mathscr{C}_f^*$, which contains the original configuration plus the position and marketing mix of the

Figure 4 One Hundred Random Draws of Brand Locations of Four Brands, Paper Towel Category



Note. E, EDLP; R, regular; H, high end.

new brand. The share of brand j under the new configuration $s_{jf}(\mathscr{C}_f^*)$ can then be calculated using the formulas above. Note that the results of a calculation like this crucially depends on using the correct heterogeneity distribution for the format when computing the shares in Equation (17). If the format-specific distribution of unobservable heterogeneity differs from the population distribution, then relying incorrectly on the population distribution can lead to substantial biases. Note also that this calculation cannot be done format-by-format when a researcher relies only on an aggregate market map.

We analyze the impact of new product entry in the high-end format. Because we will be analyzing

Table 9 **Change in Brand Shares After New Brand Introduction, High-End Format**

	No selection	Selection
Bounty	−0.013	−0.022
Brawny	−0.008	−0.012
Scott	−0.005	−0.005
Viva	−0.023	−0.050
Sparkle	−0.008	−0.009
Mardi Gras	−0.009	−0.014
Private	−0.005	−0.005
New	0.07	0.12

brand entry effects, we first check the validity of our model to predict brand choice. When estimating the paper towel model, we held out a random sample of 50 households to assess the predictive fit of the model. We computed hit rates for these 50 households using as data only the households' demographic information—not purchase data. The hit rates were computed using the posterior obtained on the remaining 452 households. The average hit rate for the brand choice model in the holdout sample was 32%. This is a reasonable fit in light of it being obtained without using the purchase data for the holdout sample.

In Table 9, we analyze the impact of new product entry in the high-end format. We assume that a new product enters this format at the position $(\alpha_{\text{new}1}, \alpha_{\text{new}2}) = (2, -2)$. In addition, we assume that the new brand enters with price equal to \$1. This puts the new product fairly close to the Viva brand (located at $(2.2, -2.7)$) but at a lower price. Table 9 shows the change in brand share (pre minus post) after entry.[10] The first column (labeled "No selection") shows the effects of entry using the population distribution of heterogeneity when calculating the shares in Equation (17). In other words, this is the impact of entry we would expect if there were no selection effects; i.e., the average high-end households equals the average household in the population. The new brand achieves a market share of 7%, and not surprisingly, we see that the Viva brand suffers the biggest drop in share (3%). The second column of Table 9 shows the effects of entry when calculating the shares using the correct heterogeneity distribution, i.e., the distribution of θ_h among high-end loyal customers.[11] This leads to substantially different entry effects: The impact on several of the incumbent brands is substantially larger than the effects calculated using the population distribution. For example, the new brand's share is now 12% and Viva's drop in share is 5%. The reason for this disparity is evident from Figure 3: High-end loyal customers have a stronger preference for brands in the west and southwest direction of the map compared to the population. In other words, all other factors being equal, high-end loyal customers prefer products located to the left and lower left of the plot (because their preference vectors tend to point in this direction). The nonnegligible bias shown in Table 9 is likely to be even bigger if we compare the findings in the second column with those using an aggregate map that does not allow separate α loadings by format.

5. Conclusion

In this paper, we study how market structure within a product category varies across different retail formats. Our modeling approach builds on the literature on internal market structures and extends it in several directions. First, we use a hierarchical structure to incorporate consumer observables (e.g., demographics) and allow for valuation of brand attributes to be correlated with marketing mix variables. More important, by allowing the loading matrix of brand attributes to be retail format specific, we can recover market maps specific to each format. Finally, to allow for systematic differences in consumers that prefer a particular retail format, we estimate a two-stage model of store choice followed by brand choice. The approach allows us to obtain brand maps for different retail formats while controlling for the short-term marketing mix activities at these stores and the self-selection of households that frequent a particular retail format.

We apply the model to consumer panel data from two product categories. The panelists in our data are observed to make purchases from three different store formats: high-end grocery stores, traditional supermarkets, and larger warehouse-type EDLP formats. These stores differ in terms of the demographic profile of the households that patronize each format, product assortment, market shares of different brands, presence and market shares of private labels, pricing, and promotional activities. Our results show that the direction and size of household preference vectors for the unobservable brand attributes vary across formats, which in turn translates into significant differences in brand maps across retail formats. Interestingly, we find a tight clustering of all the brands at EDLP stores, whereas brands are found to compete in distinct subgroups such as premium and fat-free at

[10] In the calculation, we assumed that all brands had zero feature and display activity and the prices for the incumbent brands were set at their sample average. The exercise ignores competitive reactions and potential for market expansion because of the new product (Soberman and Gatignon 2005).

[11] We computed the integral in Equation (17) by integrating over households where more than half of trips were to the high-end store. For the "No selection" column, we integrated over all households.

high-end stores. In general, brands are found to compete more intensely in the EDLP stores, which could be driven by self-selection of more price-sensitive households to this format. Our results also show that consumer perceptions about interbrand similarities and rivalries as well as perception and positioning of private labels varies significantly across formats.

There are several caveats to our analysis and directions of future research. From the modeling perspective, one could allow for marketing mix sensitivities within households to vary across formats, similar to the notion of context-specific sensitivities. For example, in the context of beer consumption, Yang et al. (2002) show how motivating conditions and environment can result in within-person heterogeneity in brand preferences. In our data, we do not observe sufficient store switching (see Table 1), with over 80% of the households showing loyalty to one specific format. With richer data (for example, with purchase histories from discount and regular supermarkets, both of which are frequented by a majority of the shoppers), an extension of the model in the spirit of Yang et al. (2002) could be quite useful. Similarly, because of data limitations, our focus in the current work is on three formats of grocery stores. As discussed in the introduction of the paper, food retailing in the United States has undergone a significant transition with the rapid growth of alternative retail formats ranging from value-oriented big box retailers to smaller high-end specialty stores. These formats compete with traditional supermarkets at opposite ends of the price and quality spectrum. In addition, depending on the product category, the competitive retail set for supermarkets could include small drugstores, convenience stores, dollar stores, price clubs, and so forth. More recent wand panel data sets collected by Information Resources, Inc. and ACNielsen track consumer purchase information from all such retail formats, and our modeling approach can easily be applied to these data. In our current paper, we find significant differences across the three types of grocery formats. Our conjecture is that such differences are likely to magnify if one were to look at the assortments, marketing mix activities, and consumer store choices across the full array of retail formats. Given the evolution of the food industry, understanding the perceptions and relative positioning of brands across retail formats could be important as this could not only impact short-term marketing activities but also have longer-term implications in terms of product line management, brand introductions, and repositioning strategies.

Appendix. Markov Chain Monte Carlo Algorithm

This section outlines the MCMC algorithm used to estimate the model.

For each household h, we observe

$$\{f_{ht}, d_{hft}, X_{htf}\}_{t=1}^{T_h},$$

where f_{ht} is the chosen format at purchase occasion t, d_{hst} is the chosen brand in format $f = f_{ht}$, and X_{htf} is the matrix of marketing mix variables in format $f = f_{ht}$ at purchase occasion t. The probability of choosing format f is

$$\Pr(F_{ht} = f \mid \eta_h, W_h) = \frac{\eta_{hf} + W_{hf}\gamma}{\sum_{j=1}^{3}\exp\{\eta_{hj} + W_{hj}\gamma\}},$$

where W_{hf} is household h's distance to format f. Conditional on format choice $F_{ht} = f$, household h chooses a brand according to

$$\Pr(D_{hft} = d \mid \psi_h, \beta_h) = \frac{\alpha'_{fd}\psi_h + X'_{htfd}\beta_h}{\sum_j^{J_f}\exp\{\alpha'_{fj}\psi_h + X'_{htfj}\beta_h\}}, \tag{18}$$

where J_f is the choice dimension in format f. We define

$$\theta_h \equiv (\psi_h, \beta_h, \eta_h).$$

The household-specific parameters are assumed to be distributed as

$$\psi_h = \Pi_\psi z_h + \nu_{h\psi},$$
$$\beta_h = \Pi_\beta z_h + \Gamma\nu_{h\psi} + \nu_{h\beta},$$
$$\eta_h = \Pi_\eta z_h + \Lambda_\psi\psi_h + \Lambda_\beta\beta_h + \nu_{h\eta},$$

where $\nu_{h\psi} \sim \mathrm{N}(0, I_2)$, $\nu_{h\beta}\mathrm{N}(0, \Omega_\beta)$, and $\nu_{h\eta} \sim \mathrm{N}(0, \Omega_\eta)$. This parameterization of the heterogeneity distribution is not optimal: Z_h, α_h, and β_h are all correlated with each other, and this might make it hard to directly estimate Π_η, Λ_ψ, and Λ_β (initial trials of our algorithm confirmed this). We therefore use a reparameterization: Substituting the ψ_h and β_h equation into the η_h equation leads to

$$\begin{aligned}\eta_h &= (\Pi_\eta + \Lambda_\psi\Pi_\psi + \Lambda_\beta\Pi_\beta)z_h + (\Lambda_\psi + \Lambda_\beta\Gamma)\nu_{h\psi} + \Lambda_\beta\nu_{h\beta} + \nu_{h\eta}\\ &= \Pi^*_\eta z_h + \Lambda^*_\psi\nu_{h\psi} + \Lambda^*_\beta\nu_{h\beta} + \nu_{h\eta},\end{aligned} \tag{19}$$

where

$$\begin{aligned}\Pi^*_\eta &= \Pi_\eta + \Lambda_\psi\Pi_\psi + \Lambda_\beta\Pi_\beta,\\ \Lambda^*_\psi &= \Lambda_\psi + \Lambda_\beta\Gamma,\\ \Lambda^*_\beta &= \Lambda_\beta.\end{aligned} \tag{20}$$

It is straightforward to verify that the Jacobian of the transformation from the original parameterization to the *-parameterization is 1 (proof available from the authors). We estimate the model in the *-parameterization. With generated posterior draws in this parameterization, it is then straightforward to generate posterior draws for the original parameterization.

Sample θ_h

We sample θ_h using a Metropolis-Hastings step. The conditional distribution of θ_h is

$$p(\theta_h)\prod_{t=1}^{T_h}\Pr(D_{hft} = d \mid \psi_h, \beta_h)\Pr(F_{ht} = f_{ht} \mid \eta_h, W_h),$$

where $p(\theta_h)$ is the joint prior distribution of $\theta_h = (\psi_h, \beta_h, \eta_h)$. To sample θ_h, we use a proposal distribution centered on the previous draw of θ_h with normal distributed jump sizes, scaled to deliver an adequate acceptance rate.

Sample $\Pi_\psi, \Pi_\beta, \Pi^*_\eta, \Gamma, \Lambda^*_\alpha, \Lambda^*_\beta, \Omega_\beta, \Omega_\eta$
These parameters only enter the joint posterior through the heterogeneity distribution $p(\theta_h)$ and prior. So the contribution from the heterogeneity distribution is simply

$$\prod_{h=1}^{H} p(\eta_h, \beta_h, \psi_h).$$

The joint distribution of $(\psi_h, \beta_h, \eta_h)$ is

$$\begin{pmatrix} \psi_h \\ \beta_h \\ \eta_h \end{pmatrix} \sim \mathrm{N}(\Pi_{\psi\beta\eta} z_h, \Sigma_{\psi\beta\eta}),$$

where

$$\Sigma_{\psi\beta\eta} \equiv \begin{pmatrix} I_2 & & \\ \Gamma & \Omega_\beta & \\ \Lambda^*_\psi & \Lambda^*_\psi \Gamma' + \Lambda^*_\beta \Omega_\beta & \Lambda^*_\psi \Lambda^{*\prime}_\psi + \Lambda^*_\beta \Omega_\beta \Lambda^{*\prime}_\beta \end{pmatrix}$$

and

$$\Pi_{\psi\beta\eta} \equiv \begin{pmatrix} \Pi_\psi \\ \Pi_\beta \\ \Pi^*_\eta \end{pmatrix}.$$

We can sample $\Pi_{\psi\beta\eta}$ conditional on $\Sigma_{\psi\beta\eta}$. We use a normal prior on $\pi_{\psi\beta\eta} \equiv \mathrm{vec}(\Pi_{\psi\beta\eta}) \sim \mathrm{N}(0, V_0 = 10 * I)$. The resulting conditional for $\pi_{\psi\beta\eta}$ is

$$\pi_{\psi\beta\eta} \sim \mathrm{N}\big(\Omega_{\pi_{\psi\beta\eta}} [\Sigma^{-1}_{\psi\beta\eta} \otimes I] M_{z\psi\beta\eta}, \Omega_{\pi_{\psi\beta\eta}}\big),$$

where

$$\Omega_{\pi_{\psi\beta\eta}} = [(\Sigma^{-1}_{\psi\beta\eta} \otimes M_{zz}) + V_0^1]^{-1},$$

$$M_{zz} = \sum_{h=1}^{H} z_h z'_h,$$

$$M_{z\psi\beta\eta} = \mathrm{vec}\bigg(\sum_{h=1}^{H} z_h (\psi'_h \ \ \beta'_h \ \ \eta'_h)\bigg).$$

To sample Ω_β, note that we can factor the joint distribution as $p(\eta \mid \beta, \psi) p(\beta \mid \psi) p(\psi)$ and Ω_β only enters

$$\beta_h \mid \psi_h \sim \mathrm{N}(\Pi_\beta z_h + \Gamma(\psi_h - \Pi_\psi z_h), \Omega_\beta).$$

We can write this as

$$\tilde{\beta}_h = \Gamma \tilde{\psi}_h + \epsilon_{h,\beta},$$

where

$$\tilde{\beta}_h \equiv \beta_h - \Pi_\beta z_h,$$

$$\tilde{\psi}_h \equiv \psi_h - \Pi_\psi z_h,$$

$$\epsilon_{h,\beta} \sim \mathrm{N}(0, \Omega_\beta).$$

Again, this is in the form of a multivariate normal regression model. We assign a Wishart prior to Ω^{-1}_β: $\Omega^{-1}_\beta \sim \mathrm{W}(\nu, S)$ with $\nu = 5$ and $S = \nu I_5$. This leads to a conditional posterior for Ω^{-1}_β of the form:

$$\Omega^{-1}_\beta \sim \mathrm{W}\bigg(\nu + H, S + \sum_{h=1}^{H} [\tilde{\beta}_h - \Gamma \tilde{\psi}_h][\tilde{\beta}_h - \Gamma \tilde{\psi}_h]'\bigg).$$

To sample Γ, note that

$$\tilde{\beta} \mid \tilde{\eta}, \tilde{\psi} \sim \mathrm{N}(\Gamma \tilde{\psi}_h + \Omega_\beta \Lambda^{*\prime}_\beta A(\tilde{\eta}_h - \Lambda^*_\psi \tilde{\psi}_h), B),$$

where $\tilde{\eta}_h = \eta_h - \Pi^*_\eta z_h$ and

$$A = (\Lambda^*_\beta \Omega_\beta \Lambda^{*\prime}_\beta + \Omega_\eta)^{-1},$$

$$B = \Omega_\beta - \Omega_\beta \Lambda^{*\prime}_\beta A \Lambda^*_\beta \Omega_\beta.$$

This leads to the regression model for Γ:

$$\tilde{\beta} - \Omega_\beta \Lambda^{*\prime}_\beta A(\tilde{\eta}_h - \Lambda^*_\psi \tilde{\psi}_h) = \Gamma \tilde{\psi}_h + \epsilon,$$

where $\mathrm{V}[\epsilon] = B$. We assign a normal prior on $\mathrm{vec}(\Gamma) \sim \mathrm{N}(0, V_0 = 10 * I)$ and sample $\mathrm{vec}(\Gamma)$ as above.

To sample Π^*_η, Λ^*_ψ, Λ^*_β, and Ω_η, note that

$$\eta_h \mid \beta_h, \psi_h \sim \mathrm{N}(\Pi^*_\eta z_h + \Lambda^*_\psi \tilde{\psi}_h + \Lambda^*_\beta \tilde{\beta}_h, \Omega_\eta).$$

Again, this is in the form of a multivariate regression model with regressors (z_h, ψ_h, β_h). We use a Wishart/normal prior as above and sample the parameters in an analogous fashion.

Sample γ
To sample γ, note that the likelihood contribution to the posterior for γ is

$$\prod_{h=1}^{H} \prod_{t=1}^{T_h} \Pr(F_{ht} = f_{ht} \mid \eta_h, W_h).$$

We use a normal prior $\gamma \sim \mathrm{N}(0, 10)$, and sample γ according to a Metropolis-Hastings step with a normal proposal distribution centered at the previous draw of γ.

Sample α
The likelihood contribution to the conditional posterior for α_f, $f = 1, 2, 3$, is

$$\prod_{h=1}^{H} \prod_{t:\, f_{ht} = f} \Pr(D_{hft} = d \mid \psi_h, \beta_h).$$

We use a normal prior for α_f: $\mathrm{vec}(\alpha_f) \sim \mathrm{N}(0, V_0 = 10 * I)$. In addition to the zero restrictions discussed in the text to pin down a unique rotation, for one format we restrict the first loading for one brand to be positive and the second loading for another brand to be positive. Formally, we do this by transforming these loadings to log space: $\log(\alpha_{1j}) \sim \mathrm{N}(0, 10)$, $\log(\alpha_{2j'}) \sim \mathrm{N}(0, 10)$. This is to guarantee identification of the model—factor analytic models are only identified up to an arbitrary *reflection* of loadings and factor scores (i.e., the likelihood stays unchanged by multiplying all loadings and factor scores by minus one). In the application, this is done for the EDLP format. We sample α_f using a Metropolis-Hastings step with a normal proposal distribution centered at the previous draw of α_f.

References

Allenby, G. M. 1989. A unified approach to identifying, estimating and testing demand structures with aggregate scanner data. *Marketing Sci.* **8**(3) 265–281.

Allenby, G., P. Lenk. 1994. Modeling household purchase behavior with logistic normal regression. *J. Amer. Statist. Assoc.* **89**(428) 1218–1231.

Allenby, G., P. Rossi. 1999. Marketing models of consumer heterogeneity. *J. Econometrics* **89**(1–2) 57–78.

Bell, D., J. Lattin. 1998. Shopping behavior and consumer preference for store price format: Why large basket shoppers prefer EDLP. *Marketing Sci.* **17**(1) 66–88.

Bell, D., T.-H. Ho, C. S. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *J. Marketing Res.* **35**(33) 352–369.

Bradlow, E. T., Y.-H. Park. 2007. Bayesian estimation of bid sequences in Internet auctions using a generalized record-breaking model. *Marketing Sci.* **26**(2) 218–229.

Chintagunta, P. 1994. Heterogeneous logit model implications for brand positioning. *J. Marketing Res.* **31** 304–311.

Chintagunta, P. 1998. Inertia and variety seeking in a model of brand purchase timing. *Marketing Sci.* **17**(3) 253–272.

Chintagunta, P., J.-P. Dubé, V. Singh. 2002. Market structure across stores: An application of a random coefficients logit model with store-level data. P. H. Franses, A. L. Montgomery, eds. *Econometric Models in Marketing*: *Advances in Econometrics*, Vol. 16. JAI Press, Amsterdam, 191–221.

Chintagunta, P., J.-P. Dubé, V. Singh. 2003. Balancing profitability and customer welfare in a supermarket chain. *Quant. Marketing Econom.* **1**(1) 111–117.

Cooper, L. G. 1988. Competitive maps: The structure underlying asymmetric cross elasticities. *Management Sci.* **34**(6) 707–723.

Day, G. S., A. D. Shocker, R. K. Srivastava. 1979. Customer-oriented approaches to identifying product-markets. *J. Marketing* **43**(4) 8–20.

Elrod, T. 1988. Choice map: Inferring a product market map from panel data. *Marketing Sci.* **7**(1) 21–40.

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

Elrod, T., G. J. Russell, A. D. Shocker, R. L. Andrews, L. Bacon, B. L. Bayus, J. D. Carroll et al. 2002. Inferring market structure from customer response to competing and complementary products. *Marketing Lett.* **13**(3) 221–232.

Erdem, T. 1996. A dynamic analysis of market structure based on panel data. *Marketing Sci.* **15**(4) 359–378.

Erdem, T. 1998. An empirical analysis of umbrella branding. *J. Marketing Res.* **35**(3) 339–351.

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

Erdem, T., R. Winer. 1999. Econometric modeling of competition: A multi-category choice-based mapping approach. *J. Econometrics* **89**(1–2) 159–175.

Fox, E. J., A. L. Montgomery, L. M. Lodish. 2004. Consumer shopping and spending across retail formats. *J. Bus.* **77**(2) S25–S60.

Gelfand, A., A. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**(410) 398–409.

Gogoi, P. 2006. A supermarket war in the store. *BusinessWeek Online* (January 24), http://www.businessweek.com/bwdaily/dnflash/jan2006/nf20060124_6416_db016.htm?campaign_id=tbw.

Gonul, F., F. Ter Hofstede. 2006. How to compute optimal catalog mailing decisions. *Marketing Sci.* **25**(1) 65–74.

Gupta, S., P. Chintagunta. 1994. On using demographic variables to determine segment membership in logit mixture models. *J. Marketing Res.* **31**(1) 128–136.

Hoch, S., X. Drèze, M. E. Purk. 1994. EDLP, Hi-Lo, and margin arithmetic. *J. Marketing* **58**(4) 16–27.

Horsky, D., S. Misra, P. Nelson. 2006. Observed and unobserved preference heterogeneity in brand-choice models. *Marketing Sci.* **25**(4) 322–335.

Kamakura, W., G. Russell. 1989. Probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(4) 379–390.

Kannan, P. K., G. P. Wright. 1991. On testing competitive market structures. *Marketing Sci.* **10**(4) 338–347.

Keane, M. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *J. Bus. Econom. Statist.* **15**(3) 310–327.

Lal, R., R. Rao. 1997. Supermarket competition: The case of every-day low-pricing. *Marketing Sci.* **16**(1) 60–80.

McCulloch, R., P. Rossi. 1994. An exact likelihood approach to analysis of the multinomal probit model. *J. Econometrics* **64**(1–2) 207–240.

Messinger, P., C. Narasimhan. 1997. A model of retail formats based on consumers economizing on shopping time. *Marketing Sci.* **16**(1) 1–23.

Ramaswamy, V., W. S. DeSarbo. 1990. SCULPTRE: A new methodology for deriving and analyzing hierarchical product-market structures from panel data. *J. Marketing Res.* **27**(4) 418–427.

Rossi, P., R. McCulloch, G. Allenby. 1996. On the value of household purchase history information in target marketing. *Marketing Sci.* **15**(4) 321–340.

Shugan, S. M. 1987. Estimating brand positioning maps using supermarket scanning data. *J. Marketing Res.* **24**(1) 1–18.

Singh, V., K. Hansen, R. Blattberg. 2006. Market entry and consumer behavior: An investigation of a Wal-Mart supercenter. *Marketing Sci.* **25**(5) 457–476.

Singh, V., K. Hansen, S. Gupta. 2005. Modeling preferences for common attributes in multi-category brand choice. *J. Marketing Res.* **42**(2) 195–209.

Soberman, D., H. Gatignon. 2005. Research issues at the boundary of competitive dynamics and market evolution. *Marketing Sci.* **24**(1) 165–174.

Whole Foods Market. 2006. Whole Foods Market 2006 Annual Report. Whole Foods Market, Inc., Austin, TX.

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

Zhu, T., V. Singh, A. Dukes. 2005. Local competition, entry, and agglomeration. Working paper, Carnegie Mellon University, Pittsburgh.