Inference for Product Competition and Separable Demand

by

Adam N. Smith
Fisher College of Business
Ohio State University
smith.6588@osu.edu

Peter E. Rossi
Anderson School of Management
UCLA
perossichi@gmail.com

Greg M. Allenby
Fisher College of Business
Ohio State University
allenby.1@osu.edu

June 2016

Inference for Product Competition and Separable Demand

Abstract

This paper presents a methodology for identifying demand groups and measuring competition for differentiated products using store-level sales data. We use the concept of economic separability as an identification condition for different product groups, and build a weakly separable model of aggregate demand. One common issue with separable demand models is that the partition of products into separable groups must be known a priori, which severely shrinks the set of admissible substitution patterns. This paper relaxes this assumption and allows the partition to be an estimated model parameter. We focus on estimating partitions within a log-log demand system where weak separability induces equality restrictions on a subset of cross-price elasticity parameters. An advantage of our approach is that we are able to find groups of separable products rather than just test whether a given set of groups is separable. Our method is applied to two aggregate, store-level data sets. The first is in the peanut butter and jelly category, where the nature of competition between products is straightforward. The second is within the juice category, where the structure of demand is less obvious. Our approach is useful for answering questions about brand positioning, and also provides gains in efficiency and predictive ability over unrestricted models.

Keywords: market structure, random partitions, price elasticities, Markov chain Monte Carlo, Bayesian estimation.

1 Introduction

Measuring product competition plays a central role in any demand analysis and is important for many marketing activities. For example, price elasticities convey degrees of substitutability between products and are commonly used to guide pricing and promotional strategies. Decisions to rebrand, reposition, or invest in new product design also require knowledge of how the brand is currently perceived relative to its competitors. Addressing standard marketing problems like these has become increasingly difficult, however, due to the granularity of modern data sets. Even the task of defining product segments or subcategories in today's marketplace is challenged by the proliferation of brands, forms, sizes, styles, and flavors.

This paper presents a methodology for identifying groups of products (termed "demand groups") that exhibit similar patterns in demand and responsiveness to changes in price. We do so in the context of an aggregate demand model for differentiated products using store-level sales data. We first use the concept of economic separability as a theoretical basis for defining product groups. Separable demand groups satisfy the condition that consumer preferences within each group are independent of consumption levels in the other groups (Deaton and Muellbauer, 1980b). For example, the demand for cereal would be separable from the demand for milk only if consumer preferences between two cereals like Cheerios and Frosted Flakes does not depend on the type or amount of milk purchased.

While this condition may trivially hold for broad categories like milk and cereal, it is less obvious if all cereals should constitute a single demand group, or if there exists a more refined structure of competition due to differences in brand, quality, flavor, or consumption context. Since this is fundamentally an empirical question, we let the partition of products be a model parameter that can be estimated from the data. In particular, we focus on estimating partitions within a log-log demand system where separability assumptions induce equality restrictions on a subset of cross-price elasticity parameters. An advantage of our approach is that we are able to discover groups of separable products rather than just test whether a given set of groups is separable.

The idea of adding structure to aggregate demand systems has a long history in marketing and economics. For example, much of the early work in applied microeconomics imposes separability restrictions in order to facilitate estimation of large demand models with many parameters and limited data (e.g., Barten, 1964; Byron, 1970). While this additional structure provides gains in efficiency, it can also limit the set of admissible substitution patterns, leading to highly restrictive demand models. Pudney (1981) suggests that empirical tests of separability are often rejected because there is no discussion of how to choose the grouping pattern imposed on the model. Our approach is thus in the spirit of Pudney (1981), as we use the data to find suitable demand groups.

Montgomery and Rossi (1999) also use separability to restrict price elasticities of a log-log demand system, but do so through the specification of the prior in a hierarchical Bayesian framework. They show that the added structure greatly improves price elasticity estimates by allowing for shrinkage towards values predicted by economic theory. A main difference between their approach and ours is the type of separability imposed on the model. In particular, they rely on an additively separable preference structure, which precludes the need to consider the partition of products into mutually exclusive groups. While additive separability may be appropriate for sets of substitutable products, we rely on weak separability in order to accommodate various types of economic relationships, including substitutes and complements.

Related work on cross-category demand typically imposes structure on price effect parameters or an underlying utility function that directly corresponds to the set of predefined product categories. For example, Song and Chintagunta (2006) estimate cross-category price and promotional effects using data on four laundry subcategories: liquid detergent, powder detergent, liquid softener, and sheet softener. In comparison, we seek to empirically find demand groups that satisfy some definition of competition (e.g., separability) rather than measure competition given a set of predefined groups. We believe that this approach will be especially useful when the relationship between predefined categories is unclear or if the goal is to uncover less obvious patterns in demand.

More recently, there have been efforts to estimate and visualize competition among high-dimensional product sets. Examples include France and Ghose (2016) and Ringel and Skiera (2016), who provide methods for visualizing competition among more than 1,000 products using demand and online search data, respectively. The output from both analyses is some form of competitive map rather than estimates of marketing responsiveness parameters. Our work differs in that the goal of identifying groups of similar products is embedded within a traditional demand analysis where we estimate both the partition of products and the associated model parameters (e.g., price elasticities). Another example is Bajari et al. (2015), who use various machine learning techniques to estimate large-scale linear and logit demand models. Many of these machine learning methods, like the LASSO, represent penalized estimators that add structure by shrinking regression coefficients to zero. In principle, we could apply a similar approach to restrict cross-price elasticities in a log-log model and identify isolated demand groups. However, even when substitution effects are zero for unrelated products, price elasticities can still be non-zero because of income effects.

We instead construct a separable aggregate demand model where certain cross-price elasticities are constrained to live in a lower dimensional space of group-level parameters. This provides a form of differential shrinkage across all elasticity parameters because the set of restricted elasticities can change with the partition. We take a Bayesian approach to inference and navigate the posterior distribution of partitions and other demand model parameters using Markov chain Monte Carlo (MCMC) methods. In particular, we use the location-scale random partitions model developed in Smith and Allenby (2016) to define a coherent prior and proposal distribution for the partition parameter. This allows us to efficiently navigate a high-dimensional, discrete, and non-Euclidean domain with little computational burden.

The proposed separable demand model is applied to weekly retail scanner data in two different product categories. The first application uses price and movement data for products in the jams, jellies, and spreads category. This particular category is chosen because of the strong prior on how to categorize products. That is, we would expect peanut butters to constitute one demand group and the jams and jellies to constitute another. Estimating the partition in this context allows us to examine the interpretability of our model. The second application uses demand data for breakfast juices, where it is harder to define an appropriate categorization a priori. Throughout both analyses, we find that partitions with high posterior probability tend to have more separable groups than predefined categories. This allows us to address a number of issues, including brand positioning and the existence of quality tiers. We also find that separable restrictions provide gains in efficiency and lead to better demand predictions at the store level.

The remainder of the paper is organized as follows. Section 2 reviews economic separability concepts, develops a separable aggregate demand model, and discusses the induced distribution for the data. Section 3 outlines our approach for estimating the partition of products along with other model parameters. A formal model for random partitions is also developed. Section 4 presents the results of the two empirical applications. Section 5 discusses possible extensions and concludes.

2 Separable Preferences and Aggregate Demand

The theory of economic separability dates back to Leontief (1947), Gorman (1959), Sono (1961), and Goldman and Uzawa (1964), who provide a framework for analyzing large or complicated demand systems under the assumption that consumer preferences conform to a lower-dimensional structure. In particular, separability requires that the set of products under study can be partitioned so that consumer preferences within each group can be described independently of consumption levels in other groups. This type of economic dimension reduction is attractive for both behavioral and econometric reasons. For example, consumers may naturally limit the amount or type of information they use when choosing among a high-dimensional set of products or attributes. Separability also limits the effects of price changes on the demand for products across groups, and therefore offers efficiency gains to the researcher wanting to estimate a demand system with many cross-price elasticities and limited data.

The problem is that the gains in efficiency or parsimony come at the cost of model flexibility. The partition of products fully determines the nature of competition, which means that assumptions of separable preferences impose strong restrictions on demand elasticities when the partition is fixed. Our approach is to let the partition be an estimated model parameter, which relaxes the assumption that the structure of demand is known. Rather, we only assume that some true structure exists and then use the data to learn about that structure.

This section outlines the construction of an aggregate demand model that is separable with respect to a partition. We first review various separability concepts and discuss three mathematical conditions for separable preferences. We then show how a restriction on price effect parameters can be used to impose separability on a log-log demand system. The induced likelihood function and other model properties are also discussed.

2.1 Economic Separability

The two main forms of separability are strong or additive separability and weak separability. We restrict our attention to weak separability because it allows meaning to be attached to the assignment of products into groups. Empirically, this implies that different partitions could yield different likelihood values, allowing us to interpret and potentially infer the partition from the data. In comparison, additive separability assumes that no special relationships among the groups exist, which makes the composition of groups arbitrary. A likelihood function induced by additively separable preferences (and within the class of demand models we consider) would be flat over the space of partitions. The adequacy of additive and weak separability assumptions also depends on the granularity of data available to the researcher. For example, additive separability has been found to be well-suited for modeling highly aggregated commodities, whereas weak separability is better for disaggregated commodities (Deaton and Muellbauer, 1980b). Since we intend to use separability to learn about the structure of demand with more granular data, weak separability will be a more appropriate concept. Going forward, we will use the terms separability and weak separability interchangeably.

Formally, let $\mathcal{g} = (g_1, \ldots, g_N)$ denote of partition of N items into $K \leq N$ groups, where g_i is an item-group indicator variable. Preferences are said to be weakly separable with respect to the partition $\mathcal{g}$ if the utility of the entire consumption vector $\boldsymbol{q} = (q_1, \ldots, q_N)$ can be written as

$$u(\boldsymbol{q}) = V(v_1(\boldsymbol{q}_1), \ldots, v_K(\boldsymbol{q}_K)). \tag{1}$$

Here $\boldsymbol{q}_k$ is the consumption vector for products assigned to group k, $v_k(\cdot)$ is a subutility function, and V is an aggregation function defined over K variables. While (1) may be useful for constructing or testing different utility functions, the intuition is limited. Therefore, we consider two additional definitions that provide more of a behavioral interpretation of separability. The first requires the marginal rate of substitution between any two products i and j in group k to be independent of consumption levels for a third product r in group ℓ.

$$\frac{\partial u_i(\boldsymbol{q})/u_j(\boldsymbol{q})}{\partial q_r} = 0 \quad \text{for all} \quad g_i = g_j = k,\ g_r = \ell,\ k \neq \ell \tag{2}$$

Here $u_i(\boldsymbol{q})$ denotes the marginal utility of good i. The consequence of (2) is that the consumption levels in group ℓ cannot affect the consumer's preference ordering for products in group k. Notice that this restriction does not say that the *demand* for products in group k is independent from the demand of products in group ℓ. Rather, separability can hold for product groups who are substitutes or complements in aggregate. For example, consider two product groups that are economic complements: pasta and pasta sauce. The demand for pasta is weakly separable from the demand for sauce only if the consumer's preferences for, say, Prego marinara sauce over the store brand marinara sauce are unaffected by the amount or type of pasta purchased. In the context of differentiated product markets, one feature that could "break" separability is the presence of quality tiers. For example, if the consumer only prefers pairing certain types of pasta with a high-quality national brand of marinara sauce, then her preferences between sauce varieties will depend on the type of pasta purchased, violating (2).

The last definition is based on restrictions to the compensated (Slutsky) price effects.

$$S_{ij} = \phi_{k\ell} \frac{\partial q_i}{\partial M} \frac{\partial q_j}{\partial M} \quad \text{for all} \quad g_i = k, \; g_j = \ell, \; k \neq \ell \tag{3}$$

Here $S_{ij} = \frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial M}$ is a compensated price effect, M is total market expenditure, and $\phi_{k\ell}$ is a parameter measuring the relationship between groups k and ℓ. Equation (3) implies that the effect of changing the price of product i in group k on the demand for product j in group ℓ must conform to the group-level price effect $\phi_{k\ell}$, except as modified by expenditure effects. Again, if the demand of pasta is weakly separable from pasta sauce, then a decrease in the price of a 23 oz. jar of Prego marinara sauce should have the same effect on the demand for all varieties of pasta, barring expenditure effects. Goldman and Uzawa (1964) show that Equations (1), (2), and (3) are all necessary and sufficient for each other, so that weakly separable preferences can be equally represented by any of the three conditions.

2.2 A Weakly Separable Log-Log Model

The next question we address is how the separability restrictions discussed above can be used to build a weakly separable aggregate demand model. Equations (1) and (3) appear to be particularly useful, in the sense that they relate to economic objects that can be manipulated by the researcher: utility functions and price effect parameters. Ideally, we would also want to apply these preference restrictions to an aggregate demand system that is consistent with the theory of utility maximization.

Examples of candidate demand systems include the Rotterdam model (Theil, 1965), the linear expenditure system (Pollak and Wales, 1969), the translog model (Christensen et al., 1975), and the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980a). These models all specify an underlying utility/cost function, and then solve for the associated demand functions using Roy's identity, Shephard's Lemma, or the Kuhn-Tucker conditions. The AIDS, Rotterdam, and translog models have the additional benefit of flexibility: the specified utility/cost function can approximate any underlying utility/cost function so that the resulting demand functions impose minimal assumptions on substitution patterns.

In principle, we could impose separability on any of these models by modifying the underlying objective function so that it conforms to the structure of (1). The problem with this approach is that the resulting demand functions may be intractible or complicated functions of prices and expenditure, which would pose problems for estimation and inference. A second way to impose separability is through restrictions on price effect/elasticity parameters in (3). However, none of the demand systems mentioned above explicitly contain price effect/elasticity parameters. Rather, elasticities must be solved for as non-linear functions of other model parameters.

To address these issues, we turn to a log-log demand system of the form

$$\log q_{it} = \beta_{0i} + \sum_{j=1}^{N} \beta_{ij} \log p_{jt} + \gamma_i \log M_t + \varepsilon_{it} \tag{4}$$

where i indexes products, t indexes time periods, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt}) \sim \mathrm{N}(0, \Sigma)$, and q_{it}, p_{jt}, and M_t represent movement, price, and total expenditure, respectively. While not formally derived from utility theory, the log-log specification is popular among both practictioners and researchers for its tractibility, flexibility, and ease of interpretation. The main advantage of this system for our purposes is that β_{ij} and γ_i are elasticity parameters. Therefore, the restriction on price effects in (3) can potentially be used to impose separability, to the extent that a log-log system can at least locally approximate a valid demand system.

In order to apply (3) to the log-log model above, we first need to translate the restriction on a compensated price effect to a restriction on an uncompensated price elasticity. It is straightforward to show that

$$S_{ij} = \phi_{k\ell} \frac{\partial q_i}{\partial M} \frac{\partial q_j}{\partial M} \implies \beta_{ij} = w_j(\theta_{k\ell}\gamma_i\gamma_j - \gamma_i) \tag{5}$$

where $w_j = p_j q_j / M$ and $\theta_{k\ell} = \phi_{k\ell}/M$. We let Θ denote the $K \times K$ matrix of separability parameters $\{\theta_{k\ell}\}$, and make the additional assumption that Θ is symmetric so that all information is contained in $\binom{K}{2}$ parameters. Since movement, price, and expenditure variables are observed over T time periods, we also replace w_j with an average expenditure share $\bar{w}_j = \frac{1}{T}\sum_{t=1}^{T} \frac{p_{jt}q_{jt}}{M_t}$. A weakly separable log-log demand model can then be defined by (4) subject to the condition that for product i in group k and product j in group ℓ

$$\beta_{ij} = \begin{cases} \bar{w}_j(\theta_{k\ell}\gamma_i\gamma_j - \gamma_i) & \text{if } k \neq \ell \\ \eta_{ij} & \text{if } k = \ell \end{cases} \tag{6}$$

where η_{ij} are unrestricted elasticities. Notice that the partition controls the grouping structure, and therefore dictates which cross-elasticities are restricted. That is, if two products belong to the same group, no restrictions are imposed. However, if two products belong to different groups, then the cross elasticities are populated by a lower-dimensional set of parameters. Separability thus offers a theoretically motivated approach for solving the explosion of parameters problem in a log-log demand system, which usually requires N^2 parameters to characterize the full elasticity matrix.

Estimates of the separability parameters are also informative of the economic relationship among product groups. The restriction in (6) implies that for normal goods ($\gamma_i > 0$, $\gamma_j > 0$)

$$\beta_{ij} > 0 \iff \theta_{k\ell} > \frac{1}{\gamma_j}. \tag{7}$$

Therefore, all products in group k will be substitutes to all products in group ℓ if $\theta_{k\ell}$ is big enough, and complements if $\theta_{k\ell}$ is small enough. However, we can no longer simply rely on the sign of $\theta_{k\ell}$ to infer the relationship, as is usually done with price elasticity parameters β_{ij}.

2.3 Likelihood

The log-log demand system has a convenient statistical representation. Let $\boldsymbol{Y}$ denote the $T \times N$ matrix of log movement variables, $\boldsymbol{X}$ the $T \times N$ matrix of log prices, $\boldsymbol{M}$ the $T \times 1$ vector of log expenditures, and $\boldsymbol{E}$ the $T \times N$ error matrix with rows defined by $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(0, \Sigma)$. Then (4) can be written in the form of a multivariate regression model

$$\boldsymbol{Y} = \mathbf{1}\boldsymbol{\beta}_0' + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{M}\boldsymbol{\gamma}' + \boldsymbol{E} \tag{8}$$

where $\mathbf{1}$ is an $N \times 1$ vector of ones, $\boldsymbol{\beta}_0$ is an $N \times 1$ vector of intercept parameters, $\boldsymbol{B}$ is an $N \times N$ matrix of price elasticities, and $\boldsymbol{\gamma}$ is an $N \times 1$ vector of expenditure elasticities. Under the assumption of multivariate normal errors, the distribution of the data will also be normal, yielding simple posterior sampling routines in the presence of conjugate priors (Rossi et al., 2005).

Estimating a log-log model subject to the separability restrictions in (6) poses additional challenges because the expenditure elasticity vector $\boldsymbol{\gamma}$ enters the mean function non-linearly. Even when γ is fixed, the dimension of unrestricted elasticities $\boldsymbol{\eta}$ and separability parameters $\boldsymbol{\theta}$ is always less than N^2 and can potentially change with the partition g. For example, if there are $N = 6$ products to be partitioned, then $g = (1, 1, 1, 2, 2, 2)$ yields a single separability parameter and 18 unrestricted elasticities. However, $g = (1, 2, 1, 2, 3, 3)$ yields 3 different separability parameters and 12 unrestricted elasticities. This complicates efficient posterior sampling, as the model in (8) is no longer linear in the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$.

To address these problems, we rewrite the proposed separable demand system as an SUR-like regression model that is linear in $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. That is, we show that conditional on $\boldsymbol{\gamma}$ and g, Equations (8) and (6) can be written as

$$\boldsymbol{y}^* = \boldsymbol{Z}_{g,\gamma}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{9}$$

Here $\boldsymbol{y}^* = \text{vec}(\boldsymbol{Y}^*)$, $\boldsymbol{Y}^*$ is equal to $\boldsymbol{Y}$ less a set of fixed parameters, $\boldsymbol{Z}_{g,\gamma}$ is a design matrix that depends on observed data and fixed parameters $(g, \boldsymbol{\gamma})$, $\boldsymbol{\varepsilon} = \text{vec}(\boldsymbol{E})$, and with a slight abuse of notation, $\boldsymbol{\beta}' = (\boldsymbol{\eta}', \boldsymbol{\theta}')$. This modified linear model will also induce a multivariate normal likelihood.

Formally, the restrictions in (6) allow us to rewrite the price elasticity matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = \boldsymbol{H} + \boldsymbol{W} \cdot (\boldsymbol{O} \cdot \boldsymbol{\gamma}\boldsymbol{\gamma}' - \boldsymbol{1}\boldsymbol{\gamma}') \tag{10}$$

where the $(\cdot)$ operator denotes element-wise matrix multiplication, $\boldsymbol{W} = \boldsymbol{1}\bar{\boldsymbol{w}}'$, $\bar{\boldsymbol{w}} = (\bar{w}_1, \ldots, \bar{w}_N)$, and $\boldsymbol{H}$ and $\boldsymbol{O}$ are sparse $N \times N$ matrices containing the unrestricted elasticities $\{\eta_{ij}\}$ and separability parameters $\{\theta_{k\ell}\}$, respectively. That is,

$$\boldsymbol{H}_{ij} = \begin{cases} \eta_{ij} & \text{for } g_i = k,\ g_j = \ell,\ k = \ell \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

and

$$\boldsymbol{O}_{ij} = \begin{cases} \theta_{k\ell} & \text{for } g_i = k,\ g_j = \ell,\ k \neq \ell \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Plugging (10) into (8) yields

$$\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{H} + \boldsymbol{X}(\boldsymbol{W} \cdot \boldsymbol{O} \cdot \boldsymbol{\gamma}\boldsymbol{\gamma}') + \boldsymbol{E} \tag{13}$$

where $\boldsymbol{Y}^* = \boldsymbol{Y} - \boldsymbol{1}\boldsymbol{\beta}_0' - \boldsymbol{W} \cdot \boldsymbol{1}\boldsymbol{\gamma}' - \boldsymbol{M}\boldsymbol{\gamma}'$. Since $\boldsymbol{O}$ and $\boldsymbol{H}$ are sparse matrices, we still need to transform (13) in order to estimate the underlying parameters $\{\theta_{k\ell}\}$ and $\{\eta_{ij}\}$. Consider a SUR-like transformation, where $\boldsymbol{Y}^*$ is vectorized and then regressed on a set of potentially different matrices.

$$\begin{pmatrix} \boldsymbol{y}_1^* \\ \boldsymbol{y}_2^* \\ \vdots \\ \boldsymbol{y}_N^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{U}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{U}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{U}_N \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{r}_{11} & \cdots & \boldsymbol{r}_{1n(\theta)} \\ \boldsymbol{r}_{21} & \cdots & \boldsymbol{r}_{2n(\theta)} \\ \vdots & & \vdots \\ \boldsymbol{r}_{N1} & \cdots & \boldsymbol{r}_{Nn(\theta)} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n(\theta)} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix} \tag{14}$$

The mean function contains two pieces: the first corresponds to the unrestricted elasticities and the second corresponds to the restricted elasticities. Let $n(\theta)$ denote the number of elements in the vector $\boldsymbol{\theta}$ and let $\boldsymbol{C}$ denote an $n(\theta) \times 2$ matrix with rows $(\boldsymbol{c}_1', \ldots, \boldsymbol{c}_{n(\theta)}')$ that index each element in the upper triangle of the $K \times K$ separability matrix Θ. Then $\boldsymbol{U}_i$ is the matrix formed by all columns m of $\boldsymbol{X}$ such that $g_i = g_m$ for $i = 1, \ldots, N$. Further, the restricted vectors $\boldsymbol{r}_{ij}$ are defined as

$$\boldsymbol{r}_{ij} = \begin{cases} 0 & \text{if } g_i \notin \boldsymbol{c}_j \\ \sum_{\{n: g_n \in \boldsymbol{c}_j \,\cap\, g_n \neq g_i\}} \bar{w}_n \gamma_i \gamma_n \boldsymbol{x}_n & \text{otherwise} \end{cases} \tag{15}$$

where $j = 1, \ldots, n(\theta)$, $n = 1, \ldots, N$, and $\boldsymbol{x}_n$ is the nth column of $\boldsymbol{X}$. Combining the first and second unrestricted and restricted matrices of (14) and stacking the vectors $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ yields

$$\boldsymbol{y}^* = \begin{pmatrix} \boldsymbol{U}_{g} & \boldsymbol{R}_{g,\gamma} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\theta} \end{pmatrix} + \boldsymbol{\varepsilon} = \boldsymbol{Z}_{g,\gamma}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{16}$$

where $\boldsymbol{\varepsilon} \sim \mathrm{N}(0, \Sigma \otimes I_T)$. The resulting likelihood is then

$$\boldsymbol{y}^* | \boldsymbol{Z}_{g,\gamma}, \boldsymbol{\beta}, g, \boldsymbol{\gamma}, \Sigma \sim \mathrm{N}(\boldsymbol{Z}_{g,\gamma}\boldsymbol{\beta}, \Sigma \otimes I_T). \tag{17}$$

We use the notation $\boldsymbol{Z}_{g,\gamma}$ to emphasize the dependence of the regression matrix $\boldsymbol{Z}$ on model parameters. The full set of model parameters is now $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}$, g, $\boldsymbol{\gamma}$, and Σ. The benefit of the form in (16) is that we are able to specify a conditionally conjugate prior for $\boldsymbol{\beta}$, which can greatly facilitate posterior sampling. A simple example that details the construction of $\boldsymbol{Z}_{g,\gamma}$ when $g = (1, 1, 2, 2, 3, 3)$ is provided in Appendix A.

3 Estimation

The separable demand model and associated likelihood proposed in Section 2 are specified with respect to a partition g. This section outlines an approach for estimating the partition parameter g along with other demand model parameters using MCMC methods. We depart slightly from the likelihood specified in subsection 2.3 and instead consider likelihood functions of the form

$$\boldsymbol{y}|\boldsymbol{Z}_g, \boldsymbol{\beta}, g, \Sigma \sim \mathrm{N}(\boldsymbol{Z}_g\boldsymbol{\beta}, \Sigma \otimes I_T). \tag{18}$$

Here $\boldsymbol{\beta}$ denotes a vector of parameters whose dimension depends on the partition g, $\boldsymbol{Z}_g$ is a design matrix that also depends on g, and Σ is a covariance matrix assumed to be known. The model in (17) takes the same form if we ignore $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}$ and fix Σ. The purpose of deviating from (17) is to isolate parameters that complicate estimation and highlight the nuances of our methodology.

We take a Bayesian approach and make inferences about $\boldsymbol{\beta}$ and g through the joint posterior distribution

$$p(\boldsymbol{\beta}, g|\boldsymbol{y}, \boldsymbol{Z}_g) \propto p(\boldsymbol{y}|\boldsymbol{Z}_g, \boldsymbol{\beta}, g)p(\boldsymbol{\beta}, g). \tag{19}$$

We build on existing MCMC methods to construct a Markov chain which yields samples from (19). A central feature of our approach is the location-scale partition distribution developed in Smith and Allenby (2016), which serves as a prior and proposal distribution for g. Our posterior sampling problem differs from the one considered in Smith and Allenby (2016), however, who assume structural *indepedence* between $\boldsymbol{\beta}$ and g so that the prior and proposal distributions can be factored as $p(\boldsymbol{\beta}, g) = p(\boldsymbol{\beta})p(g)$. We instead exploit the linearity of the model conditional on g, and suggest joint priors and proposals with the form $p(\boldsymbol{\beta}, g) = p(\boldsymbol{\beta}|g)p(g)$.

This section first provides details on the construction and resulting properties of the location-scale partition distribution. An estimation routine is then proposed to sample from (19) while accounting for the dependence between $\boldsymbol{\beta}$ and g.

3.1 A Model for Random Partitions

The construction of models for random partitions is often based on sampling from a discrete random probability measure, which is a central topic of nonparametric Bayesian statistics. For example, using the Pólya-urn scheme of Blackwell and MacQueen (1973) to sample from a random distribution G with a Dirichlet process prior (Ferguson, 1973) also induces a model for random partitions. This result is due to the almost sure discreteness of G. That is, if a set of random variables are sampled from G, then some of the draws will take on the same value with positive probability. The tied values can then be used to assign the items into mutually exclusive groups, which is equivalent to constructing a partition. An example of this process is shown in Figure 1.

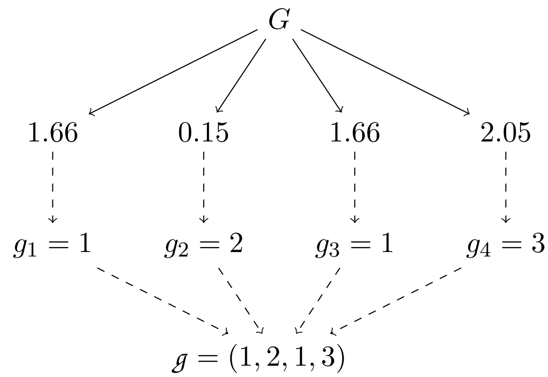


Figure 1: Four items are randomly drawn from a discrete distribution G. The realized values are then used to create the vector item-group assignment variables $g_1, \ldots, g_4$ that constitute a partition. The solid lines indicate probabilistic sampling, while the dashed lines indicate deterministic labeling to construct the partition.

Formally, consider sampling $\phi_i | G \overset{iid}{\sim} G$ for $i = 1, \ldots, N$ where $G \sim \text{DP}(\alpha, G_0)$. The concentration parameter α and base distribution G_0 are fixed by the researcher. Blackwell and MacQueen's Pólya-urn scheme provides a way to generate the vector of ϕ's: let $\phi_1 \sim G_0$ and sample $\phi_2, \ldots, \phi_N$ according to

$$\phi_i | \phi_1, \ldots, \phi_{i-1} \sim G_0 p_0 + \sum_{k=1}^{K^{(i)}} \delta_{\phi_k^*} p_k \tag{20}$$

where

$$p_0 = \frac{\alpha}{\alpha + i - 1} \quad \text{and} \quad p_k = \frac{n_k}{\alpha + i - 1}. \tag{21}$$

Here ϕ_k^* is the kth unique value in the set $\{\phi_1, \ldots, \phi_{i-1}\}$, $\delta_{\phi_k^*}$ is a point-mass at ϕ_k^*, $K^{(i)}$ is the number of ϕ_k^*'s, and n_k is the number of ϕ's equal to ϕ_k^*. The weights p_0 and p_k represent the probability that ϕ_i is a new draw from the base distribution or exactly equal to ϕ_k^*, respectively. Without loss of generality, we choose $G_0 = \text{N}(0, 1)$ since the value of each ϕ_i only serves as a group label when modeling partitions. The important feature of this sampling scheme is that ϕ_i can be exactly equal to one of the previous ϕ's with strictly positive probability. Therefore, we can translate the vector $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_N)$ into a partition $\mathcal{g} = (g_1, \ldots, g_N)$ using the following mapping:

$$\phi_i = \phi_k^* \implies g_i = k \quad \text{for} \quad i = 1, \ldots, N. \tag{22}$$

Random partition models based on Pólya-urn sampling schemes offer two main advantages. First, partitions are generated according to the order restriction

$$g_1 = 1 \quad \text{and} \quad g_i \in \{1, \ldots, \max(g_1, \ldots, g_{i-1}) + 1\} \quad \text{for} \quad i = 2, \ldots, N \tag{23}$$

which prevents any issues of label switching. We will let $\mathcal{P}_N$ denote the set of all possible partitions of N items subject to this order restriction. Second, the number of groups within the partition is random and need not be fixed by the researcher. The problem with the specific probability model induced by Dirichlet process sampling is that the composition of groups only depends on α and n_k. Further, the only parameter the researcher can use to control the model is α. This restrictive nature of the Dirichlet process partitioning model makes it unattractive as a prior because it is not obvious how to summarize our knowledge of $\mathcal{g}$ through α. It is also unattractive as a proposal distribution within an MCMC routine because we can only navigate the high-dimensional partition space using independent state transitions. That is, there is no way to "center" the partition distribution at the current state $\mathcal{g}^{\text{old}}$ when generating a candidate partition $\mathcal{g}^{\text{new}}$.

An alternative partitioning model is proposed by Smith and Allenby (2016), who parameterize the group-assignment probabilities in (20) with a location partition $\rho = (\rho_1, \ldots, \rho_N)$ and dispersion parameter τ.

$$(\phi_i | \boldsymbol{\phi}_{<i}, \rho, \tau) \sim G_0 p_0(\rho_i, \tau) + \sum_{k=1}^{K^{(i)}} \delta_{\phi_k^*} p_k(\{\rho_i, \boldsymbol{\rho}_k\}, \tau) \tag{24}$$

The composition of groups is now controlled by the information in ρ and τ. Holding τ fixed, two items that are grouped together in ρ are also more likely to be grouped together in $\mathcal{g}$. The functions $p_0(\rho_i, \tau)$ and $p_k(\{\rho_i, \boldsymbol{\rho}_k\}, \tau)$ serve as a formal distance metric for measuring the similarity among elements of ρ. The specific choice of functional form for $p_0(\rho_i, \tau)$ and $p_k(\{\rho_i, \boldsymbol{\rho}_k\}, \tau)$ follows from Smith and Allenby (2016) and is described in detail in Appendix B. The resulting location-scale partition distribution induced by (24) is denoted $\text{LSP}(\rho, \tau)$.

Operationally, the $\text{LSP}(\rho, \tau)$ distribution behaves like a continuous, univariate distribution belonging to a location-scale family. That is, ρ shifts the center of the distribution on $\mathcal{P}_N$ and τ

determines the amount of dispersion around ρ. As τ goes to zero, more probability mass is placed on ρ. As τ gets large, the mass is spread more evenly over $\mathcal{P}_N$. This feature is shown in Figure 2, which plots the LSP(ρ, τ) distribution function for $N = 6$, $\rho = (1, 1, 1, 2, 2, 2)$, and two different values of τ. In each plot, the partitions are ordered by probability and the red bar corresponds to the location partition.

(a) $\tau = 0.1$ **(b)** $\tau = 2$

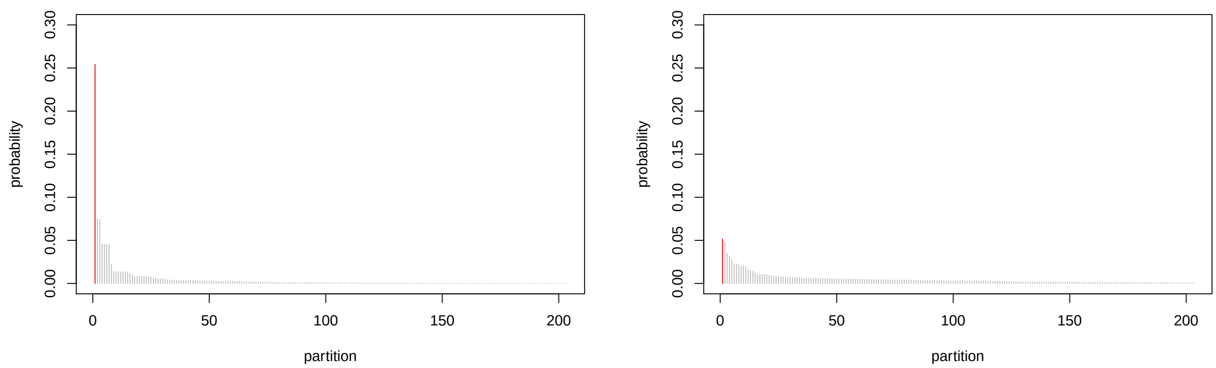


Figure 2: Two LSP(ρ, τ) distributions with $\rho = (1, 1, 1, 2, 2, 2)$ and $\tau = 0.1$ and 2, respectively. The red bar indicates the location partition ρ. The partitions are put in descending order according to their LSP probabilities.

The LSP(ρ, τ) distribution facilitates estimation and inference of partitions in two ways. First, it serves as a prior distribution for g, where prior knowledge can be directly imposed through the choice of location partition ρ and dispersion parameter τ. Second, it serves as a proposal distribution in an MCMC routine. Specifically, we implement a random-walk Metropolis-Hastings algorithm, where a candidation partition g^{new} is proposed using the LSP distribution centered around g^{old}. The resulting Markov chain will be able to navigate the partition space much more effectively than alternative independence proposal mechanisms.

3.2 MCMC Routine

Sampling from (19) is challenged by the high-dimensional, discrete domain of g and the structural dependence between $\boldsymbol{\beta}$ and g. For example, the conditional posterior distribution of g given the data and all other parameters does not belong to a known parametric family, which makes direct Gibbs sampling infeasible. Even if we replaced the Gibbs step for g with the Metropolis-Hastings step suggested by Smith and Allenby (2016), the dimension of $\boldsymbol{\beta}$ in the current state may not conform with the candidate partition, resulting in an intractible likelihood calculation. Therefore, rather than trying to iteratively sample from each parameter's full conditional distribution, we use a single Metropolis-Hastings step that jointly proposes $\boldsymbol{\beta}$ and g. The efficiency of a joint proposal will then crucially depend on the choice of proposal distributions, given the complex domain of g and potentially high dimension of $\boldsymbol{\beta}$.

We first consider generating values of $\boldsymbol{\beta}$ when the partition g is fixed. Since the underlying regression model in (18) is linear in $\boldsymbol{\beta}$ conditional on g, we are able to exploit normal conjugacy theory in order to construct efficient proposals. That is, if we specify a conditionally conjugate prior for $\boldsymbol{\beta}$

$$\boldsymbol{\beta}|g \sim \mathrm{N}(\bar{\boldsymbol{\beta}}_g, A_g^{-1}) \tag{25}$$

then the conditional posterior distribution of $\boldsymbol{\beta}$ takes the form

$$\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{Z}_g, g, \Sigma \sim \mathrm{N}(\tilde{\boldsymbol{\beta}}, (\tilde{\boldsymbol{Z}}_g' \tilde{\boldsymbol{Z}}_g + A_g)^{-1}) \tag{26}$$

where

$$\begin{aligned}
\Sigma &= U'U \\
\tilde{\boldsymbol{y}} &= ((U^{-1})' \otimes I_T)\boldsymbol{y} \\
\tilde{\boldsymbol{Z}}_g &= ((U^{-1})' \otimes I_T)\boldsymbol{Z}_g \\
\tilde{\boldsymbol{\beta}} &= (\tilde{\boldsymbol{Z}}_g' \tilde{\boldsymbol{Z}}_g + A_g)^{-1}(\tilde{\boldsymbol{Z}}_g' \tilde{\boldsymbol{y}} + A_g \bar{\boldsymbol{\beta}}_g).
\end{aligned}$$

Proposing values of $\boldsymbol{\beta}$ according to (26) while holding g fixed results in a highly efficient independence Metropolis-Hastings sampling routine, in the sense that the proposal distribution is exactly the conditional posterior distribution. In fact, sampling $\boldsymbol{\beta}$ according to (26) reduces to Gibbs sampling, so the acceptance ratio in the Metropolis-Hastings algorithm will always equal one.

Given a proposal distribution for $\boldsymbol{\beta}$ conditional on g, we then use the $\mathrm{LSP}(\rho, \tau)$ distribution to marginally propose values of g, as in Smith and Allenby (2016). That is, we generate $g^{\text{new}} \sim \mathrm{LSP}(g^{\text{old}}, v)$ where g^{old} is the value of the partition in the current state and v is a step size. The efficiency of this approach comes from being able to generate g^{new} from a distribution "centered" around the current partition, as in a random-walk Metropolis-Hastings algorithm. The main difference between our proposals and traditional random-walk proposals with normal errors is that our proposal distribution is highly asymmetric. We must then include the appropriate transition probabilities when evaluating the Metropolis-Hastings acceptance ratio.

The associated estimation routine is outlined below. The full MCMC routine that samples from the posterior distribution induced by the demand model in section 2 is provided in Appendix C. We use the superscript "old" to denote the parameter value at iteration $(r-1)$ throughout.

For each iteration $r = 1, \dots, R$

1. Generate the partition

$$g^{\text{new}} \sim q_1(g|g^{\text{old}}, v) = \text{LSP}(g^{\text{old}}, v).$$

Conditional on g^{new}, generate $\boldsymbol{\beta}^{\text{new}}$ from its full conditional distribution

$$\boldsymbol{\beta}^{\text{new}} \sim q_2(\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{Z}_{g^{\text{new}}}, g^{\text{new}}, \Sigma) = \text{N}(\tilde{\boldsymbol{\beta}}, (\tilde{\boldsymbol{Z}}'_{g^{\text{new}}}\tilde{\boldsymbol{Z}}_{g^{\text{new}}} + A_{g^{\text{new}}})^{-1}).$$

2. Set $(g^{(r)}, \boldsymbol{\beta}^{(r)}) = (g^{\text{new}}, \boldsymbol{\beta}^{\text{new}})$ with probability

$$\alpha = \min\left(1, \frac{p(\boldsymbol{y}|\boldsymbol{Z}_{g^{\text{new}}}, \boldsymbol{\beta}^{\text{new}}, g^{\text{new}}, \Sigma)p(\boldsymbol{\beta}^{\text{new}}|g^{\text{new}})p(g^{\text{new}})}{p(\boldsymbol{y}|\boldsymbol{Z}_{g^{\text{old}}}, \boldsymbol{\beta}^{\text{old}}, g^{\text{old}}, \Sigma)p(\boldsymbol{\beta}^{\text{old}}|g^{\text{old}})p(g^{\text{old}})} \times \frac{q_2(\boldsymbol{\beta}^{\text{old}}|g^{\text{old}})q_1(g^{\text{old}}|g^{\text{new}})}{q_2(\boldsymbol{\beta}^{\text{new}}|g^{\text{new}})q_1(g^{\text{new}}|g^{\text{old}})}\right).$$

Otherwise, set $(g^{(r)}, \boldsymbol{\beta}^{(r)}) = (g^{\text{old}}, \boldsymbol{\beta}^{\text{old}})$.

There are two benefits from the joint proposal strategy outlined above. First, the resulting ratio of likelihoods will always be coherent in the sense that the dimension of $\boldsymbol{\beta}$ is guaranteed to conform with g. Second, proposing $\boldsymbol{\beta}^{\text{new}}$ from its full conditional distribution independent of $\boldsymbol{\beta}^{\text{old}}$ allows us to handle changes in dimensionality with little computational burden. Traditional attempts to accommodate trans-dimensional sampling problems within MCMC rely on random-walk proposals (e.g., reversible-jump MCMC), which require complicated transition probability calculations and become inefficient in high dimensions (Green, 1995). We only use random-walk proposals to marginally navigate a high-dimensional (but fixed) partition space, and then use efficient indepedence proposals for $\boldsymbol{\beta}$. Therefore, the transition probabilities in the proposed routine only require evaluating a multivariate normal density and the LSP distribution function.

3.3 Simulation Study

We perform a simulation study for two purposes. First, to show the effect of the separability restrictions in (6) on the resulting cross-price elasticity matrix. Second, to show that the proposed estimation routine can recover all model parameters. Data are generated from the separable demand model in section 2. The true parameter values and dimension of the data are chosen in order to loosely match the empirical application below. Let $N = 12$ be the total number of products and $T = 100$ the number of time periods. The matrix of log prices is generated from a Unif$(0, 1.5)$ distribution, the log expenditure variables are generated from a Unif$(0, 3)$ distribution, and the expenditure shares are generated from a Dirichlet(5) distribution.

Rather than pick the partition arbitrarily, we generate g from a LSP(ρ, τ) distribution with $\rho = (1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2)$ and $\tau = 1$. The realized value of the partition is

$$g^{\text{true}} = (1, 1, 1, 2, 2, 1, 3, 2, 4, 4, 4, 3).$$

Conditional on g^{true}, the rest of the model parameters are generated as follows:

$$\begin{aligned}
\eta_{ij}|g^{\text{true}} &\sim \text{N}(-3, .5) \quad \text{for all } (i,j) \text{ s.t. } g_i = g_j \text{ and } i = j \\
\eta_{ij}|g^{\text{true}} &\sim \text{N}(2, .5) \quad \text{for all } (i,j) \text{ s.t. } g_i = g_j \text{ and } i \neq j \\
\theta_n|g^{\text{true}} &\sim \text{N}(0, 3) \quad \text{for } n = 1, \dots, n(\theta) \\
\beta_{i0} &\sim \text{N}(0, 5) \\
\gamma_i &\sim \text{Unif}(.2, 2) \\
\Sigma &= 0.1 \cdot J_N + 5 \cdot I_N
\end{aligned}$$

where J_N is an $N \times N$ matrix of ones. The associated matrix of separability coefficients is

$$\Theta = \begin{pmatrix} - & -4.04 & 0.34 & -1.89 \\ - & - & 4.38 & 0.16 \\ - & - & - & -0.70 \\ - & - & - & - \end{pmatrix}.$$

The full matrix of price elasticity parameters β_{ij} induced by the separable demand model is shown in Figure 3. We see that the sign of $\theta_{k\ell}$ alone does not dictate the sign of β_{ij}. If $\theta_{k\ell}$ is large enough (e.g., $\theta_{23} = 4.38$), then the associated cross price elasticities will also be positive (see the elements of Figure 3 in blue). However, if $\theta_{k\ell}$ is positive but small (e.g., $\theta_{13} = 0.34$), the associated cross price elasticities can be negative depending on the magnitude of the expenditure elasticities.

$$\begin{pmatrix}
-2.95 & 1.68 & 1.95 & -0.15 & -0.30 & 1.70 & -0.03 & -0.18 & -0.14 & -0.22 & -0.05 & -0.03 \\
1.67 & -3.89 & 2.49 & -0.36 & -0.71 & 1.25 & -0.06 & -0.42 & -0.34 & -0.53 & -0.12 & -0.07 \\
2.14 & 1.82 & -3.55 & -0.51 & -1.00 & 2.38 & -0.09 & -0.59 & -0.48 & -0.74 & -0.17 & -0.10 \\
-0.11 & -0.39 & -0.23 & -2.95 & 2.20 & -0.08 & 0.24 & 2.52 & -0.02 & -0.02 & -0.01 & 0.08 \\
-0.39 & -1.36 & -0.78 & 1.75 & -2.70 & -0.29 & 0.83 & 2.39 & -0.07 & -0.08 & -0.05 & 0.29 \\
2.13 & 2.29 & 1.61 & -0.30 & -0.59 & -3.13 & -0.05 & -0.35 & -0.28 & -0.44 & -0.10 & -0.06 \\
-0.12 & -0.14 & -0.04 & 0.11 & 0.64 & -0.04 & -3.39 & 0.23 & -0.22 & -0.32 & -0.10 & 1.68 \\
-0.17 & -0.60 & -0.35 & 2.82 & 1.27 & -0.13 & 0.37 & -4.45 & -0.03 & -0.03 & -0.02 & 0.13 \\
-0.25 & -0.76 & -0.42 & -0.13 & -0.09 & -0.17 & -0.29 & -0.11 & -3.25 & 1.95 & 0.49 & -0.16 \\
-0.36 & -1.08 & -0.60 & -0.18 & -0.12 & -0.24 & -0.41 & -0.16 & 1.16 & -3.31 & 2.38 & -0.23 \\
-0.10 & -0.30 & -0.17 & -0.05 & -0.03 & -0.07 & -0.12 & -0.04 & 2.61 & 2.30 & -3.68 & -0.06 \\
-0.07 & -0.08 & -0.02 & 0.06 & 0.35 & -0.02 & 3.19 & 0.13 & -0.12 & -0.18 & -0.05 & -3.55
\end{pmatrix}$$

Figure 3: The true values of the price elasticity matrix induced by the separable demand model. The colors correspond to the lower dimensional matrix of separability matrix which populate many of the cross price elasticity parameters.

The proposed model is estimated using the MCMC routine outlined in Appendix C. We impose diffuse, conditionally conjugate priors on $\boldsymbol{\beta}_0$, $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, and Σ. A diffuse normal prior is used for the expenditure elasticities $\boldsymbol{\gamma}$. The LSP(ρ, τ) distribution is used as a prior for g where ρ assigns all products into the same group and $\tau = 1$. The Markov chain is run for $R = 200,000$ iterations and the first 10% are discarded as burn-in. We also set the step size of the random-walk proposals for g to be $v = .1$ for the first 5% of the draws and then $v = .001$ thereafter.

The trace plot of the partition parameter is shown in Figure 4. The x-axis indexes the iteration, and the y-axis labels the unique partitions visited in the Markov chain. We can see that the true partition (marked by the horizontal red line) is reached within the first 10,000 iterations. Moreover, a small set of partitions continue to be visited throughout the course of the R iterations, while still placing highest posterior mass on g^{true}.

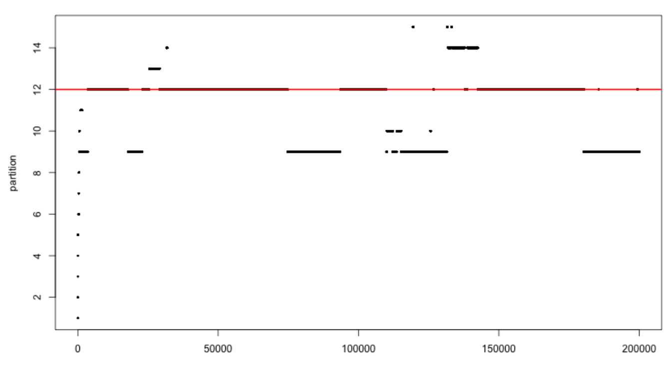


Figure 4: The trace plot for the partition parameter g. The horizontal red line represents g^{true}.

Figure 5a and Figure 5b plot the true values for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ against 95% credible intervals corresponding to the conditional posterior distributions $p(\boldsymbol{\theta}|\boldsymbol{y}, \boldsymbol{Z}, g = g^{\text{true}})$ and $p(\boldsymbol{\eta}|\boldsymbol{y}, \boldsymbol{Z}, g = g^{\text{true}})$. Figure 5c and Figure 5d plot the true values for $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}$ against the unconditional 95% credible intervals. We find that the all true values are covered by the 95% credible intervals, which shows that the proposed estimation routine is able to recover all true model parameters. While the values of Σ are also well-recovered, we suppress estimates of Σ for the sake of brevity.

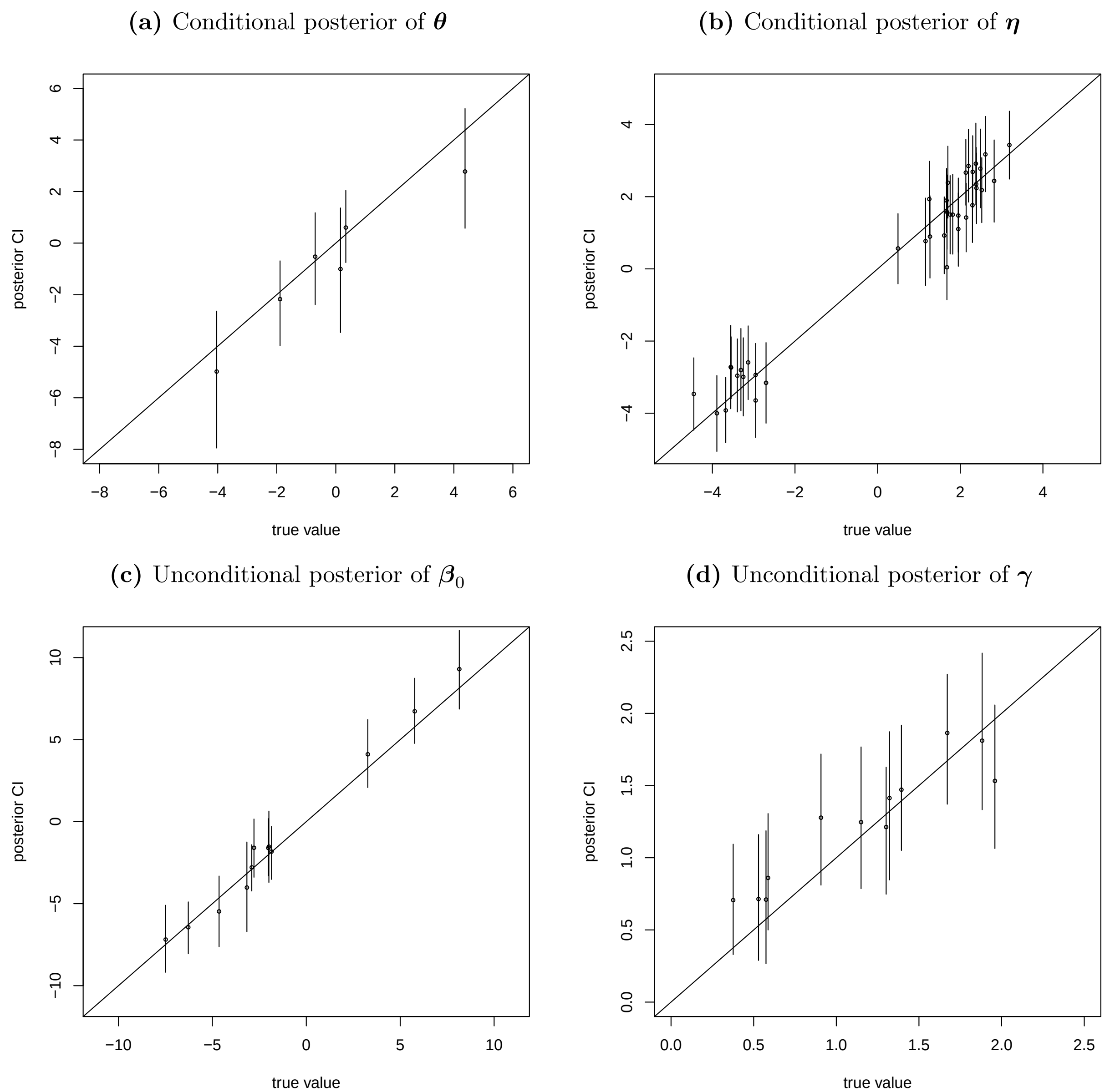


Figure 5: Posterior means and credible intervals are plotted against the true values of $\boldsymbol{\theta}$, $\boldsymbol{\eta}$, $\boldsymbol{\beta}_0$, and $\boldsymbol{\gamma}$.

4 Empirical Analysis

4.1 Data Description

Our data come from the Nielsen Retail Scanner data set, which contains weekly movement, pricing, and store environment information for over 90 retail chains across the United States.[1] For simplicity, we focus on a single grocery retailer in Cincinnati, Ohio and use price and movement data from the store with the highest volume. The data cover a four-year period ($T = 208$) spanning January 2008 through December 2011. We use three years of data for estimation and one year of data for out-of-sample validation. To address concerns of endogeneity of the total expenditure and expenditure share variables in (4) and (6), we construct proxies using price and movement data from the remaining retail stores within the Cincinnati designated market area (DMA). If there are a total of S_{dma} stores (excluding the one with highest volume) in the Cincinnati DMA, then the revenue of product j at time t is formed using a share-weighted average of revenue across all stores: $R_{tj} = \sum_{s=1}^{S_{\text{dma}}} w_s p_{tjs} q_{tjs}$. Here w_s is an expenditure share for store s. Total market expenditure and expenditure shares are then calculated as $M_t = \sum_{j=1}^{N} R_{tj}$ and $\bar{w}_j = \frac{1}{T} \sum_{t=1}^{T} \frac{R_{tj}}{M_t}$.

In order to illustrate the benefits of our methodology, we use data from two product categories that differ based on prior knowledge of market structure. We first consider the jams, jellies, and spreads category and study the demand for a set of peanut butter and jelly (PBJ) offerings. This allows us to assess both the interpretability and validity of our model, given the distinct differences of and complementary relationship between peanut butter and jelly. That is, we may expect the model to separate peanut butters from jellies in the estimated partition and then indicate group-wise complementarity in the estimates of the separability coefficients. Our second application uses data from the juice category, where the majority of products are substitutes. However, the extensive number of types, forms, brands, and flavors makes the nature of product competition less clear. In this case, our methodology provides a parsimonious way to learn about the structure of demand.

[1] The data are provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business. http://research.chicagobooth.edu/nielsen

4.1.1 Peanut Butter and Jelly Data

Our first application uses price and movement data from the peanut butter, jelly, and jam subcategories. The top three brands from each subcategory are chosen based on within-group market share. The resulting brands include four national brands (Jif peanut butter, Skippy peanut butter, Welch's jelly/jam, Smucker's jelly/jam) and a store brand with products in all three groups. A time series plot of UPC-level prices (Figure 6) shows that line pricing exists both within and across subcategories. For example, Jif coordinates prices of creamy and chunky peanut butter, while Welch's coordinates prices of grape jam and jelly.

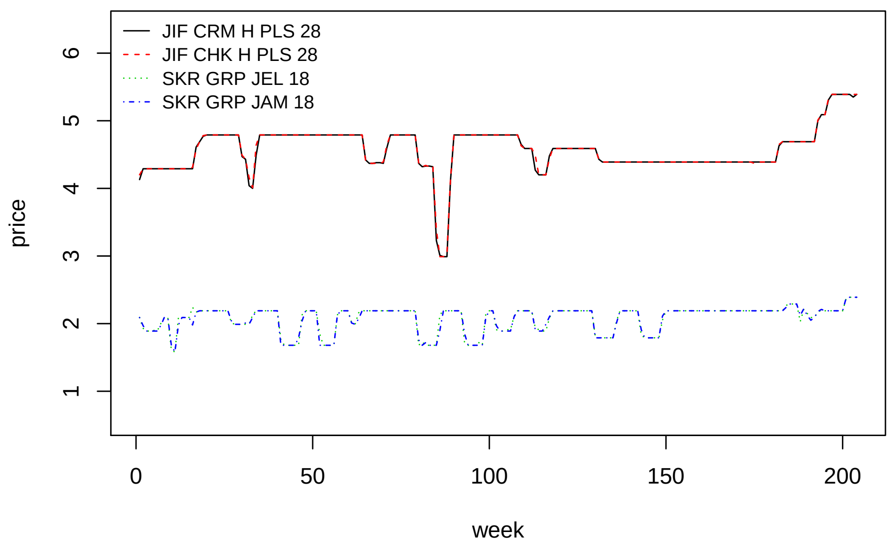


Figure 6: UPC-level prices for Jif peanut butter and Smucker's grape jelly and jam are plotted over the 208 week sample period.

We therefore aggregate UPCs up to the brand-size combination (termed "item" going forward) and consider both small and large package sizes. For peanut butter, this means that each item is aggregated over different types (e.g., creamy, chunky, reduced-fat). We further require each peanut butter item to contain at least one creamy variety and one chunky variety. Jelly and jam items are aggregated across product groups, so we focus on grape varieties and require each item to have at least one jelly UPC and one jam UPC. The resulting $N = 11$ items included in our analysis account for 62.3% of total demand within the peanut butter, jelly, and jam subcategories. More detailed descriptions of the items including market shares and average prices are provided in Table 1.

Table 1: PBJ Data Description

	Brand	Size (oz.)		Share (%)	Average Price ($)
1	Jif	SM	16-18	27.09	2.73
2	Jif	LG	28	15.09	4.57
3	Skippy	SM	16-18	3.56	2.98
4	Private Label	SM	16-18	19.72	1.76
5	Private Label	LG	28	11.24	3.13
6	Welch's	SM	16-22	3.96	2.41
7	Welch's	LG	30-32	6.62	2.37
8	Smucker's	SM	18	2.52	2.07
9	Smucker's	LG	32	3.29	2.31
10	Private Label	SM	18	2.14	1.66
11	Private Label	LG	32	4.77	1.78

4.1.2 Juice Data

Our second application uses price and movement data from the apple and orange juice product categories. We consider the top four brands for each flavor, which includes five national brands (Mott's, Simply, Nestle Juicy Juice, Tropicana, Minute Maid) and a store brand. Simply and the store brand are the only two brands to offer products in both flavor categories. Juices are further differentiated based on storage type: shelf-stable (S), refrigerated (R), and frozen (F). While non-frozen apple and orange juice are traditionally sold in shelf-stable and refrigerated containers, respectively, Simply Apple and Simply Orange are both refrigerated products.

We again aggregate UPCs up to brand-size combinations, as line pricing exists within many juice products as well (e.g., pulp and low-pulp orange juice). The resulting $N = 17$ items account for 91.9% of total demand across the frozen and non-frozen apple and orange juice categories. Detailed descriptions of the items are provided in Table 2.

Table 2: Juice Data Description

	Brand	Flavor	Type	Size (oz.)	Share (%)	Average Price ($)
1	Mott's	Apple	S	64	4.77	2.69
2	Mott's	Apple	S	128	1.77	5.06
3	Simply Apple	Apple	R	59	3.28	3.57
4	Nestly Juicy Juice	Apple	S	46	6.20	2.86
5	Private Label	Apple	S	64	12.59	1.85
6	Private Label	Apple	S	96	2.43	3.65
7	Tropicana	Orange	R	54	9.48	3.09
8	Tropicana	Orange	R	89	5.11	5.59
9	Minute Maid	Orange	R	59	4.51	2.77
10	Minute Maid	Orange	R	89	2.52	5.67
11	Simply Orange	Orange	R	59	9.61	3.52
12	Simply Orange	Orange	R	89	4.07	6.35
13	Private Label	Orange	R	59	11.72	1.86
14	Private Label	Orange	R	128	16.81	3.30
15	Private Label	Apple	F	12	1.02	1.24
16	Minute Maid	Orange	F	12	1.47	2.22
17	Private Label	Orange	F	12	2.61	1.29

4.2 Estimation Results

The proposed separable demand model is estimated on both data sets along with an unrestricted log-log demand model. The MCMC routine described in Appendix C is an extension of the routine outlined in subsection 3.2 and is used to estimate parameters of the proposed model. A standard Gibbs sampler for the multivariate regression model (e.g., Rossi et al., 2005) is used for the unrestricted model. We impose diffuse normal priors on $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, and a diffuse inverse-Wishart prior on Σ. The LSP(ρ, τ) distribution is used as a prior for g with $\tau = 1$ and ρ chosen based on prior knowledge of market structure. For the PBJ data, the location partition assigns all peanut butters into one group and the jams and jellies into another. For the juice data, the location partition assigns all apple juices into one group and the orange juices into another. We find that our empirical results are robust to the choice of ρ and τ, however. Each Markov chain is run for $R = 200,000$ iterations and the first 20% and 40% of draws are discarded as burn-in for the PBJ and juice data sets, respectively. The step size of the random-walk proposals for g is set to $v = .1$ for the first 5% of the draws and then $v = .001$ thereafter.

Figure 7a provides posterior estimates of the in-sample and predictive root mean squared error (RMSE) for both models across both data sets. We find that the unresticted model performs well in-sample, but has inferior predictive ability. Specifically, the restrictions imposed by separability not only lead to better predictions, on average, but also more precise predictions. This result is highlighted in Figure 7b and Figure 7c, which plot the entire predictive RMSE distribution for both models and data sets. In comparison to the separable models, the unrestricted models produce predictive RMSE distributions that are shifted right and have longer upper tails.

(a) Model Fit Statistics

		In-Sample RMSE		Predictive RMSE	
Data Set	Model	Mean	SD	Mean	SD
PBJ	Unrestricted	0.344	0.002	0.370	0.010
	Separable	0.359	0.003	0.349	0.004
Juice	Unrestricted	0.304	0.002	0.362	0.007
	Separable	0.328	0.003	0.348	0.004

(b) PBJ **(c)** Juice

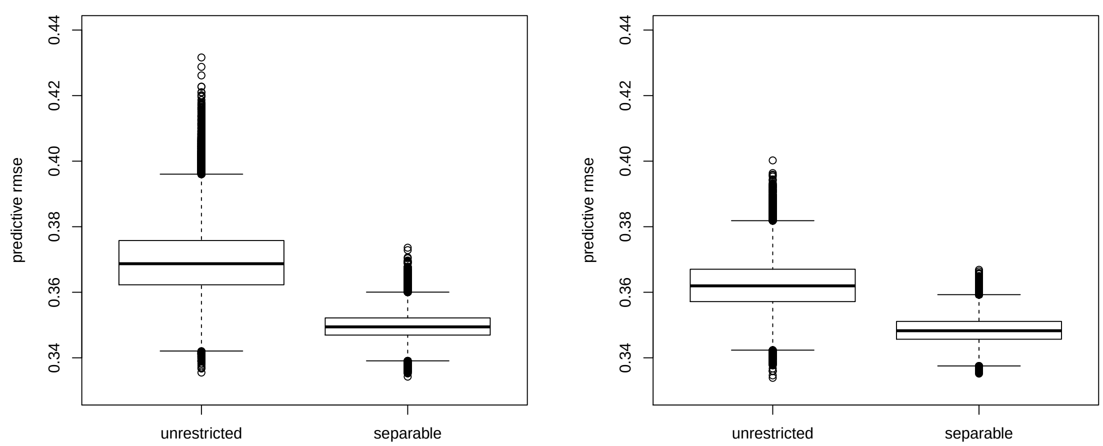
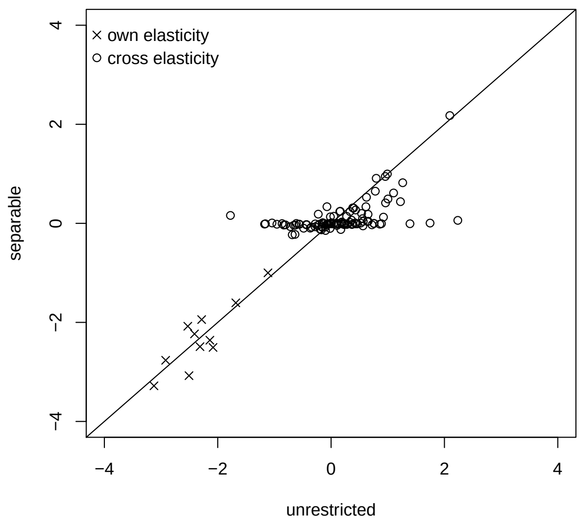


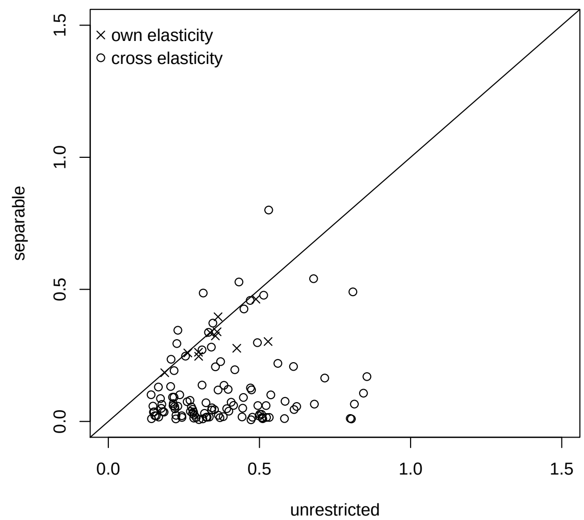
Figure 7: In-sample and predictive RMSE for the PBJ and juice data sets.

Next, we compare models using estimates of the demand model parameters. Estimates of product intercepts and expenditure elasticities for the proposed model are provided in Appendix D. Figure 8 plots the posterior means and standard deviations of the full price elasticity matrix for the unrestricted model against the induced elasticities from the separable model. While we find a high degree of correlation between elasticity estimates, the precision of the estimates is almost uniformly worse for both unrestricted models. This is shown in Figure 8b and Figure 8d, where the majority of points fall to the right of the 45 degree line. This provides evidence that separability restrictions are still able to provide gains in efficiency when the partition is also estimated.

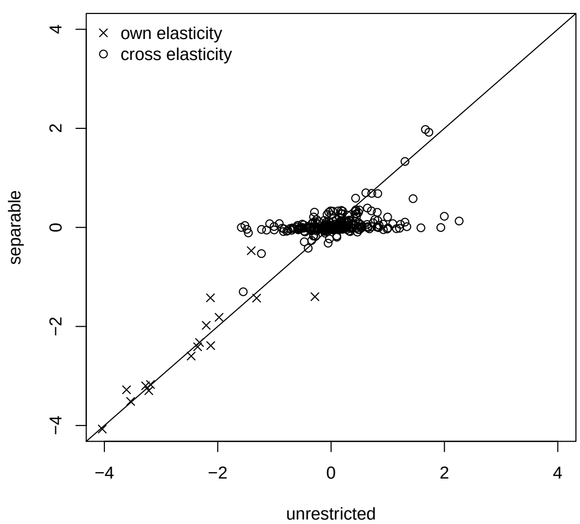
(a) PBJ: Posterior Means



(b) PBJ: Posterior Standard Deviations



(c) Juice: Posterior Means



(d) Juice: Posterior Standard Deviations

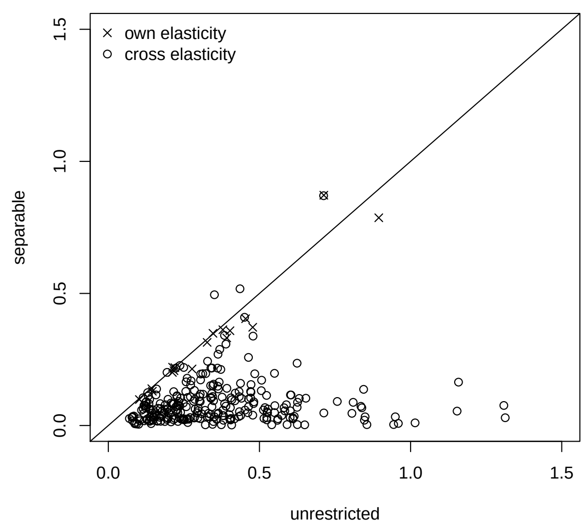


Figure 8: The posterior means and standard deviations of price elasticity parameters for both the PBJ and juice data sets.

We also find that many of the cross-price elasticities are estimated to be zero in the proposed model. One explanation is that some of the demand groups are not only separable, but also isolated.

The fact that our model can accommodate both cases is an advantage over penalized regression models that simply shrink many of the cross-price elasticities to zero.

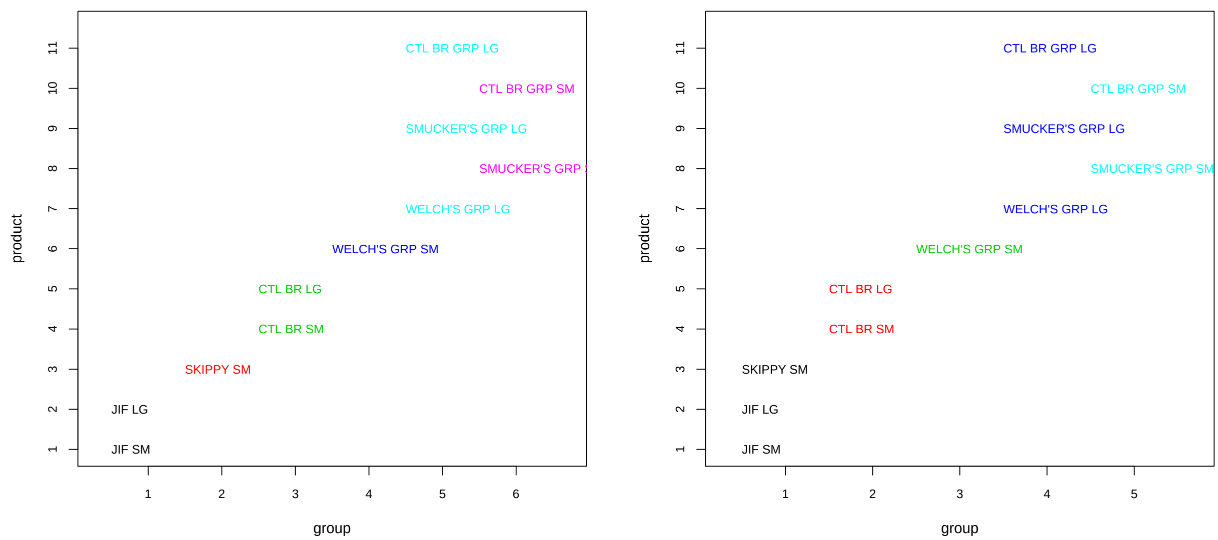
The posterior distribution of the partition g can be used to learn about the structure of demand. Appendix E provides trace plots for g in each data set, which show the Markov chain exploring the marginal posterior distribution of g. We find that a small set of partitions continues to be visited over the course of 200,000 iterations. In the juice data set, for example, 88% of the posterior mass is concentrated on 10 unique partitions, which is a dramatic reduction from the roughly 83 billion elements in the set of all possible partitions of $N = 17$ items. This provides us with some confidence in saying the chain has converged. Our sampling routine is also able to move around neighborhoods of these high probability partitions, which allows us to learn about the variability in the posterior distribution of g.

The two partitions with highest posterior probability from both data sets are plotted in Figure 9. Across both analyses, we find that partitions consistent with the data contain many groups relative to the number of products. In principle, the number of groups can be controlled through the prior for g. For example, if the location partition ρ contains two groups, then the LSP distribution will place higher mass on partitions with two groups. As stated before, however, we have found that the likelihood overwhelms the prior for even small to moderately sized data sets like ours.

In the PBJ data set, we find that partitions with high posterior probability do separate peanut butter from jelly/jam, but are also much more refined than a partition that assigns all peanut butter products into one group and all jellies and jams into another. For example, the maximum a posteriori (MAP) estimate of g contains six separable groups: three consist of only peanut butter and the other three consist of only jelly/jam. The composition of these groups reveals information about the more granular nature of competition. For example, Jif and store brand peanut butter are assigned into separate groups in nearly all partitions with positive posterior probability. This result could be explained by the presence of quality tiers among the peanut butter offerings. It would then be of interest for brand managers of Skippy, a lower-market share national brand, to know how their product is perceived relative to their competitors. We find that Skippy peanut butter is grouped together with Jif 31% of the time, with the store brand only 2% of the time, and on its own 67% of the time. This suggests that Skippy is sufficiently differentiated from a potentially lower quality store brand and perceived to be more similar to its national brand competitor.

(a) PBJ

First partition (26.6% posterior probability) Second partition (12.9% posterior probability)



(b) Juice

First partition (39.7% posterior probability) Second partition (17.6% posterior probability)

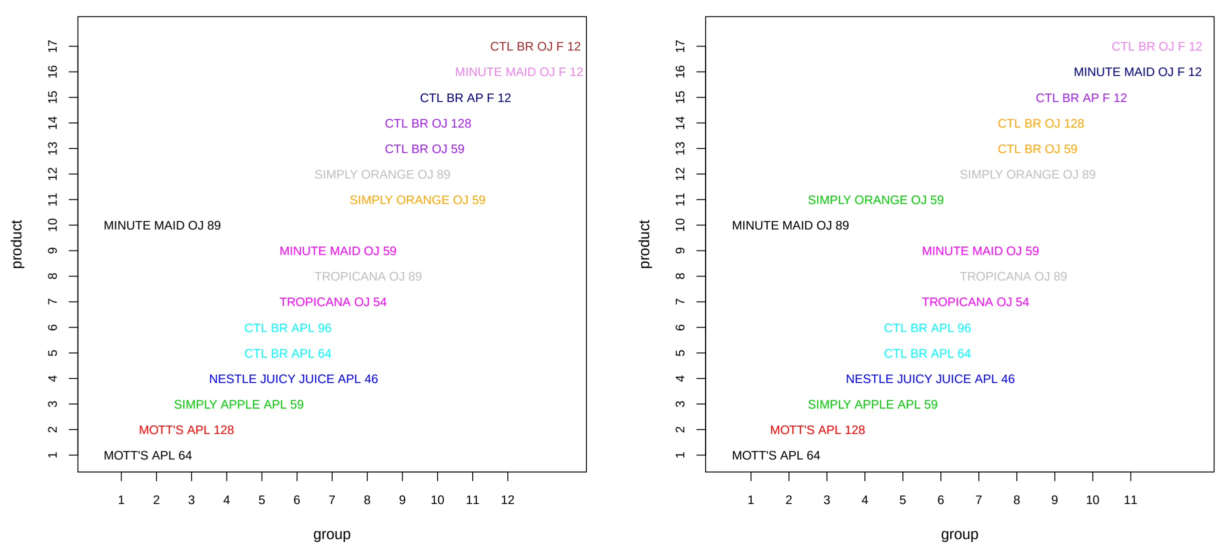


Figure 9: The two partitions with highest posterior probability are plotted for each data set. The y-axis labels the N items and the x-axis labels the $K \leq N$ groups. Two items that are vertically aligned belong to the same group.

In the juice data set, we find that flavor should not be used as a basis for defining separable demand groups. For example, the MAP estimate of g contains 12 groups, where one group contains both Mott's apple juice and Minute Maid orange juice. In fact, partitions with appreciable posterior mass always place an apple juice item and orange juice item in the same group. Another example is the partition with the second highest posterior probability, where Simply Apple and Simply Orange are grouped toghether. This means that consumer preferences for Simply Apple and other varieties of apple juice depend on the consumption levels of Simply Orange. The non-separability of the demand for Simply Apple and Simply Orange could be attributed to marketing activities. For example, Simply Apple and Simply Orange have identical branding and are displayed together in the refrigerated section of the grocery store. Both activities could help Simply, a much younger brand than its competitors, create a local monopoly in the juice category. We also find that the demand for store brand items is almost always separable from the demand for nationally branded items, just as in the PBJ analysis.

Finally, estimates of the separability parameters $\boldsymbol{\theta}$ provide information about the relationship between separable groups due to the result in Equation (7). In the PBJ data set, for example, we may not only expect peanut butter and jelly to be in separable groups, but we may also expect the separability parameters to indicate group-wise complementarity. Given that the a priori relationship between different demand groups in the juice category is less obvious, we only report estimates of $\boldsymbol{\theta}$ for the PBJ analysis (Figure 10).

We find that the estimates of $\boldsymbol{\theta}$ largely conform with the expectation that peanut butter and jelly and complementary goods. That is, the separability parameters tend to be smaller for groups across subcategories and larger for groups within a subcategory (e.g., different brands of peanut butter). In the MAP estimate of $\mathcal{g}$, for instance, the estimated separability parameter measuring the relationship between Jif peanut butter and a small jar of Welch's jelly/jam is -1.37, while the estimated parameter for Jif and Skippy peanut butter is 4.59. We also find that appropriately measuring the group-level relationships can improve the estimates of the full price elasticity matrix. For example, there is a 22% increase in the number of cross-category elasticities estimated to be negative in the proposed model relative to the unrestricted log-log model.

First partition ($K = 5$)

–	4.59	3.72	-1.37	-0.80	0.18
	–	3.08	1.22	0.13	-1.10
		–	1.49	2.20	-0.09
			–	1.03	0.02
				–	1.37
					–

Second partition ($K = 4$)

–	4.29	4.11	-1.40	-0.80
	–	3.04	1.08	0.27
		–	1.71	1.73
			–	0.72
				–

Figure 10: Posterior means of the separability coefficients conditional on the two partitions with highest posterior probability in the PBJ analysis.

5 Conclusion

This paper develops a separable aggregate demand model for the purpose of identifying demand groups and measuring product competition. Rather than fix the partitioning of products into separable groups a priori, we let the partition be a model parameter that can be estimated from the data. In particular, we estimate demand groups within a log-log demand system where weak separability induces equality restrictions on a subset of cross-group price elasticities. An immediate benefit is that we are able to use the data to find potentially less obvious grouping structures.

The proposed model is estimated on two data sets that differ based on prior knowledge of market structure. We first analyze the demand for a set of peanut butter and jelly offerings, and find that partitions with high posterior probability separate the peanut butters from the jellies, but also contain refined groups within each subcategory. For example, the demand for store brand peanut butter is always separable from national brand peanut butter, suggesting the presence of quality tiers. Our second application analyzes the demand for apple and orange juice, where it is harder to define demand groups a priori due to the variety of brands, flavors, and forms. We find that multiple flavors are often present within demand groups, suggesting that the demand for orange juice is not completely separable from the demand for apple juice. This result would be of interest to brands like Simply, who offer products in both flavor subcategories. The fact that Simply Apple and Simply Orange are often grouped together suggests that the brand has been able to differentiate itself from its competitors and meet the demand of a niche market of consumers. Across both analyses, we also find that separability assumptions lead to more precise elasticity estimates and demand predictions.

There are many possible extensions of the current work. For example, we could modify the

LSP(ρ, τ) prior to include additional information about product attributes, including unstructured data such as text or images from product packaging and advertisements. This would result in a partitioning model that, a priori, groups together products that are more similar in attribute space. Moreover, we could use data from multiple retail stores within the same chain to understand within-market heterogeneity of the demand model parameters. The LSP(ρ, τ) prior could then be used as a distribution of heterogeneity for the store specific partition parameters, where ρ would represent an aggregate measure of competition across all stores. Finally, we could consider a time series model of separable demand with time-varying partition parameters. This would allow us to address the dynamic aspect of product competition and ultimately measure how the structure of demand changes over time. We leave these topics for future research.

A Example: Constructing $\boldsymbol{Z}_{\mathscr{g},\gamma}$

Suppose $N = 6$ and $\mathscr{g} = (1, 1, 2, 2, 3, 3)$. We wish to show that the multivariate regression model in (4) can be transformed into a SUR-like model that is linear in $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. The demand model in (4) subject to the restrictions in (6) is given by the following system of equations.

$$
\begin{array}{llllllll}
\boldsymbol{y}_1 = \eta_{1,1}\boldsymbol{x}_1 + \eta_{1,2}\boldsymbol{x}_2 & + & \bar{w}_3(\theta_{12}\gamma_1\gamma_3 - \gamma_1)\boldsymbol{x}_3 + \bar{w}_4(\theta_{12}\gamma_1\gamma_4 - \gamma_1)\boldsymbol{x}_4 & + & \bar{w}_5(\theta_{13}\gamma_1\gamma_5 - \gamma_1)\boldsymbol{x}_5 + \bar{w}_6(\theta_{13}\gamma_1\gamma_6 - \gamma_1)\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_1 \\
\boldsymbol{y}_2 = \eta_{2,1}\boldsymbol{x}_1 + \eta_{2,2}\boldsymbol{x}_2 & + & \bar{w}_3(\theta_{12}\gamma_2\gamma_3 - \gamma_2)\boldsymbol{x}_3 + \bar{w}_4(\theta_{12}\gamma_2\gamma_4 - \gamma_2)\boldsymbol{x}_4 & + & \bar{w}_5(\theta_{13}\gamma_2\gamma_5 - \gamma_2)\boldsymbol{x}_5 + \bar{w}_6(\theta_{23}\gamma_2\gamma_6 - \gamma_2)\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_2 \\
\boldsymbol{y}_3 = \bar{w}_1(\theta_{12}\gamma_3\gamma_1 - \gamma_3)\boldsymbol{x}_1 + \bar{w}_2(\theta_{12}\gamma_3\gamma_2 - \gamma_3)\boldsymbol{x}_2 & + & \eta_{3,3}\boldsymbol{x}_3 + \eta_{3,4}\boldsymbol{x}_4 & + & \bar{w}_5(\theta_{23}\gamma_3\gamma_5 - \gamma_3)\boldsymbol{x}_5 + \bar{w}_6(\theta_{23}\gamma_3\gamma_6 - \gamma_3)\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_3 \\
\boldsymbol{y}_4 = \bar{w}_1(\theta_{12}\gamma_4\gamma_1 - \gamma_4)\boldsymbol{x}_1 + \bar{w}_2(\theta_{12}\gamma_4\gamma_2 - \gamma_4)\boldsymbol{x}_2 & + & \eta_{4,3}\boldsymbol{x}_3 + \eta_{4,4}\boldsymbol{x}_4 & + & \bar{w}_5(\theta_{23}\gamma_4\gamma_5 - \gamma_4)\boldsymbol{x}_5 + \bar{w}_6(\theta_{23}\gamma_4\gamma_6 - \gamma_4)\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_4 \\
\boldsymbol{y}_5 = \bar{w}_1(\theta_{13}\gamma_5\gamma_1 - \gamma_5)\boldsymbol{x}_1 + \bar{w}_2(\theta_{23}\gamma_5\gamma_2 - \gamma_5)\boldsymbol{x}_2 & + & \bar{w}_3(\theta_{23}\gamma_5\gamma_3 - \gamma_5)\boldsymbol{x}_3 + \bar{w}_4(\theta_{23}\gamma_5\gamma_4 - \gamma_5)\boldsymbol{x}_4 & + & \eta_{5,5}\boldsymbol{x}_5 + \eta_{5,6}\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_5 \\
\boldsymbol{y}_6 = \bar{w}_1(\theta_{13}\gamma_6\gamma_1 - \gamma_6)\boldsymbol{x}_1 + \bar{w}_2(\theta_{23}\gamma_6\gamma_2 - \gamma_6)\boldsymbol{x}_2 & + & \bar{w}_3(\theta_{23}\gamma_6\gamma_3 - \gamma_6)\boldsymbol{x}_3 + \bar{w}_4(\theta_{23}\gamma_6\gamma_4 - \gamma_6)\boldsymbol{x}_4 & + & \eta_{6,5}\boldsymbol{x}_5 + \eta_{6,6}\boldsymbol{x}_6 & + & \boldsymbol{\varepsilon}_6
\end{array}
$$

Next, we split the mean function into two pieces. The first matrix corresponds to the unrestricted elasticities and the second corresponds to the restricted elasticities.

$$
\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \\ \boldsymbol{y}_4 \\ \boldsymbol{y}_5 \\ \boldsymbol{y}_6 \end{pmatrix}
=
\underbrace{\begin{pmatrix}
(\boldsymbol{x}_1 \;\; \boldsymbol{x}_2) & 0 & 0 & 0 & 0 & 0 \\
0 & (\boldsymbol{x}_1 \;\; \boldsymbol{x}_2) & 0 & 0 & 0 & 0 \\
0 & 0 & (\boldsymbol{x}_3 \;\; \boldsymbol{x}_4) & 0 & 0 & 0 \\
0 & 0 & 0 & (\boldsymbol{x}_3 \;\; \boldsymbol{x}_4) & 0 & 0 \\
0 & 0 & 0 & 0 & (\boldsymbol{x}_5 \;\; \boldsymbol{x}_6) & 0 \\
0 & 0 & 0 & 0 & 0 & (\boldsymbol{x}_5 \;\; \boldsymbol{x}_6)
\end{pmatrix}}_{\boldsymbol{U}_{\mathscr{g}}}
\begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \\ \boldsymbol{\eta}_4 \\ \boldsymbol{\eta}_5 \\ \boldsymbol{\eta}_6 \end{pmatrix}
$$

$$
+ \underbrace{\begin{pmatrix}
\bar{w}_3\gamma_1\gamma_3\boldsymbol{x}_3 + \bar{w}_4\gamma_1\gamma_4\boldsymbol{x}_4 & \bar{w}_5\gamma_1\gamma_5\boldsymbol{x}_5 + \bar{w}_6\gamma_1\gamma_6\boldsymbol{x}_6 & 0 \\
\bar{w}_3\gamma_2\gamma_3\boldsymbol{x}_3 + \bar{w}_4\gamma_2\gamma_4\boldsymbol{x}_4 & \bar{w}_5\gamma_2\gamma_5\boldsymbol{x}_5 + \bar{w}_6\gamma_2\gamma_6\boldsymbol{x}_6 & 0 \\
\bar{w}_1\gamma_3\gamma_1\boldsymbol{x}_1 + \bar{w}_2\gamma_3\gamma_2\boldsymbol{x}_2 & 0 & \bar{w}_5\gamma_3\gamma_5\boldsymbol{x}_5 + \bar{w}_6\gamma_3\gamma_6\boldsymbol{x}_6 \\
\bar{w}_1\gamma_4\gamma_1\boldsymbol{x}_1 + \bar{w}_2\gamma_4\gamma_2\boldsymbol{x}_2 & 0 & \bar{w}_5\gamma_4\gamma_5\boldsymbol{x}_5 + \bar{w}_6\gamma_4\gamma_6\boldsymbol{x}_6 \\
0 & \bar{w}_1\gamma_5\gamma_1\boldsymbol{x}_1 + \bar{w}_2\gamma_5\gamma_2\boldsymbol{x}_2 & \bar{w}_3\gamma_5\gamma_3\boldsymbol{x}_5 + \bar{w}_4\gamma_5\gamma_4\boldsymbol{x}_4 \\
0 & \bar{w}_1\gamma_6\gamma_1\boldsymbol{x}_1 + \bar{w}_2\gamma_6\gamma_2\boldsymbol{x}_2 & \bar{w}_3\gamma_6\gamma_3\boldsymbol{x}_5 + \bar{w}_4\gamma_6\gamma_4\boldsymbol{x}_4
\end{pmatrix}}_{\boldsymbol{R}_{\mathscr{g},\gamma}}
\begin{pmatrix} \theta_{12} \\ \theta_{13} \\ \theta_{23} \end{pmatrix}
- C_\gamma +
\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \boldsymbol{\varepsilon}_4 \\ \boldsymbol{\varepsilon}_5 \\ \boldsymbol{\varepsilon}_6 \end{pmatrix}
$$

Here C_γ is a vector of constants given $\boldsymbol{\gamma}$. It follows that

$$
\boldsymbol{y} = U_{\mathscr{g}}\boldsymbol{\eta} + R_{\mathscr{g},\gamma}\boldsymbol{\theta} - C_\gamma + \boldsymbol{\varepsilon} \implies \boldsymbol{y} - C_\gamma = (\boldsymbol{U}_{\mathscr{g}} \;\; \boldsymbol{R}_{\mathscr{g},\gamma}) \begin{pmatrix} \boldsymbol{\eta} \\ \boldsymbol{\theta} \end{pmatrix} + \boldsymbol{\varepsilon}
$$

$$
\implies \boldsymbol{y}^* = \boldsymbol{Z}_{\mathscr{g},\gamma}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.
$$

B Similarity Functions

The functions $p_0(\rho_i, \tau)$ and $p_k(\{\rho_i, \boldsymbol{\rho}_k\}, \tau)$ measure the similarity among elements of ρ and govern the partitioning process of (24). A key property of $p_k(\cdot)$ is that, for some fixed τ, it must be increasing as ρ_i and $\boldsymbol{\rho}_k$ become more similar. The problem is that there is no natural distance metric for categorical variables. To address this issue, we follow the approach taken in Park and Dunson (2010), Müller et al. (2011), and Smith and Allenby (2016).

The idea is to measure distances between ρ_i and $\boldsymbol{\rho}_k$ using an auxiliary probability model. Formally, define

$$p_0(\rho_i, \tau) \propto \int f(\rho_i|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\tau)d\boldsymbol{\xi} \tag{27}$$

$$p_k(\{\rho_i, \boldsymbol{\rho}_k\}, \tau) \propto \int f(\rho_i|\boldsymbol{\xi})\pi^*(\boldsymbol{\xi}|\boldsymbol{\rho}_k, \tau)d\boldsymbol{\xi} \tag{28}$$

where $f(\rho_i|\boldsymbol{\xi})$ is a Categorical distribution with probability vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{C+1})$ and C is the number of unique values in the set $\{\rho_1, \ldots, \rho_{i-1}\}$. Moreover, $\pi(\boldsymbol{\xi}|\tau)$ is a conjugate Dirichlet prior with fixed hyperparameters τ. $\pi^*(\boldsymbol{\xi}|\boldsymbol{\rho}_k, \tau)$ is also a conjugate Dirichlet prior on $\boldsymbol{\xi}$, but the hyperparameters τ are now updated with the category counts in $\boldsymbol{\rho}_k$.

The benefit of a conjugate Dirichlet-Categorical model is that the similarity functions have closed form solutions. Further, the marginal probabilities are increasing functions in the distance between ρ_i and $\boldsymbol{\rho}_k$. This results in a partitioning model that acts like a continuous location-scale probability distribution. That is, $p(g|\rho, \tau)$ places higher mass on ρ as τ goes to 0, but spreads the mass more evenly across all partitions as τ gets large.

C MCMC Routine

Step 1. *Metropolis-Hastings step for $\mathcal{g}, \boldsymbol{\beta}, \boldsymbol{\gamma}$.*

(a) Generate $\mathcal{g}^{\text{new}} \sim q_1(\mathcal{g}|\mathcal{g}^{\text{old}}, v) = \text{LSP}(\mathcal{g}^{\text{old}}, v)$ and $\gamma^{\text{new}} \sim q_2(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{\text{old}}, s^2) = \text{N}(\gamma^{\text{old}}, s^2)$. Conditional on $\mathcal{g}^{\text{new}}$ and γ^{new}, generate $\boldsymbol{\beta}^{\text{new}}$ from its full conditional distribution:

$$\boldsymbol{\beta}^{\text{new}} \sim q_3(\boldsymbol{\beta}|\boldsymbol{y}^*, \boldsymbol{Z}_{\mathcal{g}^{\text{new}},\gamma^{\text{new}}}, \mathcal{g}^{\text{new}}, \boldsymbol{\gamma}^{\text{new}}, \boldsymbol{\beta}_0^{(r)}, \Sigma^{(r)}) = \text{N}(\tilde{\boldsymbol{\beta}}, (\tilde{\boldsymbol{Z}}'_{\mathcal{g}^{\text{new}},\gamma^{\text{new}}} \tilde{\boldsymbol{Z}}_{\mathcal{g}^{\text{new}},\gamma^{\text{new}}} + A_{\mathcal{g}})^{-1}).$$

Set $(\mathcal{g}^{(r)}, \boldsymbol{\beta}^{(r)}, \boldsymbol{\gamma}^{(r)}) = (\mathcal{g}^{\text{new}}, \boldsymbol{\beta}^{\text{new}}, \boldsymbol{\gamma}^{\text{new}})$ with probability

$$\alpha = \min\left(1, \frac{p(\boldsymbol{Y}|\boldsymbol{X}, \mathcal{g}^{\text{new}}, \boldsymbol{\beta}^{\text{new}}, \boldsymbol{\gamma}^{\text{new}}, \boldsymbol{\beta}_0^{(r)}, \Sigma^{(r)}) p(\boldsymbol{\beta}^{\text{new}}|\mathcal{g}^{\text{new}}) p(\mathcal{g}^{\text{new}}) p(\boldsymbol{\gamma}^{\text{new}})}{p(\boldsymbol{Y}|\boldsymbol{X}, \mathcal{g}^{\text{old}}, \boldsymbol{\beta}^{\text{old}}, \boldsymbol{\gamma}^{\text{old}}, \boldsymbol{\beta}_0^{(r)}, \Sigma^{(r)}) p(\mathcal{g}^{\text{old}}|\boldsymbol{\beta}^{\text{old}}) p(\mathcal{g}^{\text{old}}) p(\boldsymbol{\gamma}^{\text{old}})} \times \frac{q_3(\boldsymbol{\beta}^{\text{old}}|\mathcal{g}^{\text{old}}, \boldsymbol{\gamma}^{\text{old}}) q_1(\mathcal{g}^{\text{old}}|\mathcal{g}^{\text{new}}) q_2(\boldsymbol{\gamma}^{\text{old}}|\boldsymbol{\gamma}^{\text{new}})}{q_3(\boldsymbol{\beta}^{\text{new}}|\mathcal{g}^{\text{new}}, \boldsymbol{\gamma}^{\text{new}}) q_1(\mathcal{g}^{\text{new}}|\mathcal{g}^{\text{old}}) q_2(\boldsymbol{\gamma}^{\text{new}}|\boldsymbol{\gamma}^{\text{old}})}\right)$$

Otherwise, set $(\mathcal{g}^{(r)}, \boldsymbol{\beta}^{(r)}, \boldsymbol{\gamma}^{(r)}) = (\mathcal{g}^{\text{old}}, \boldsymbol{\beta}^{\text{old}}, \boldsymbol{\gamma}^{\text{old}})$.

Step 2. *Gibbs sampling step for $(\boldsymbol{\beta}_0, \Sigma)$.*

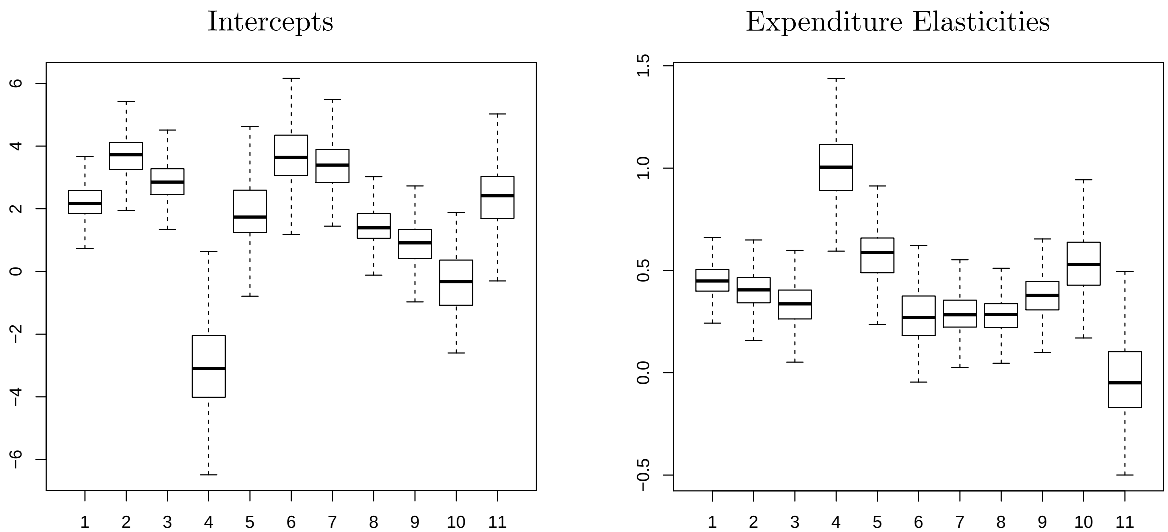
Draw $\boldsymbol{\beta}_0$ and Σ from the multivariate regression

$$\boldsymbol{Y}^* = \mathbf{1}\boldsymbol{\beta}_0' + \boldsymbol{E}$$

where $\boldsymbol{Y}^* = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}_r - \boldsymbol{M}\boldsymbol{\gamma}'$, $\boldsymbol{B}_r$ is restricted elasticity matrix depending on $\boldsymbol{\beta}$, $\mathcal{g}$, and $\boldsymbol{\gamma}$, and the rows of $\boldsymbol{E}$ are given by $\boldsymbol{\varepsilon}_t \sim \text{N}(0, \Sigma)$.

D Intercept and Expenditure Elasticity Estimates

(a) PBJ



(b) Juice

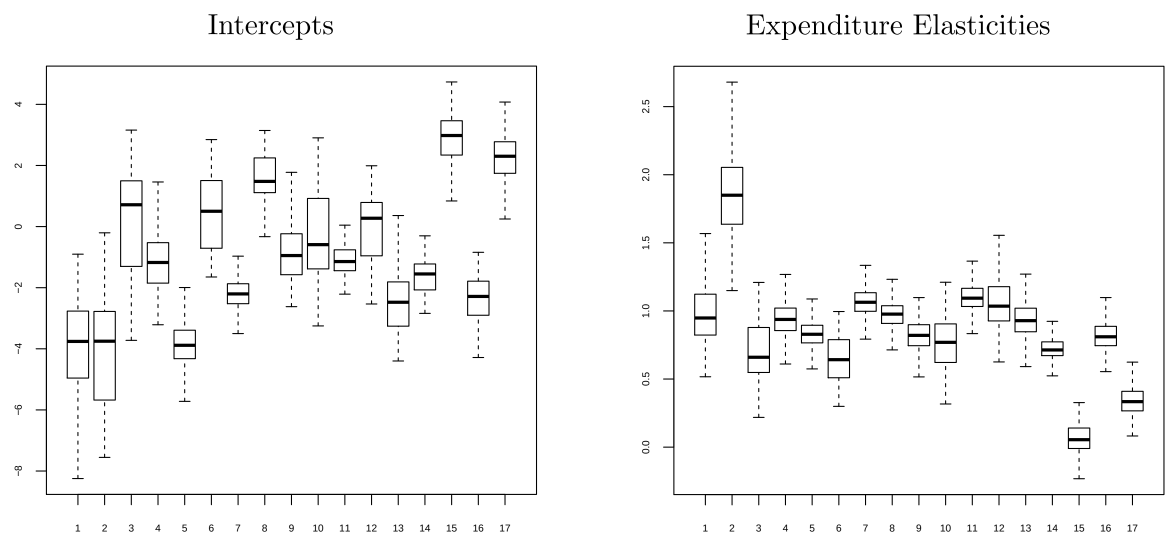


Figure 11

E Trace Plots

E.1 Peanut Butter and Jelly Data

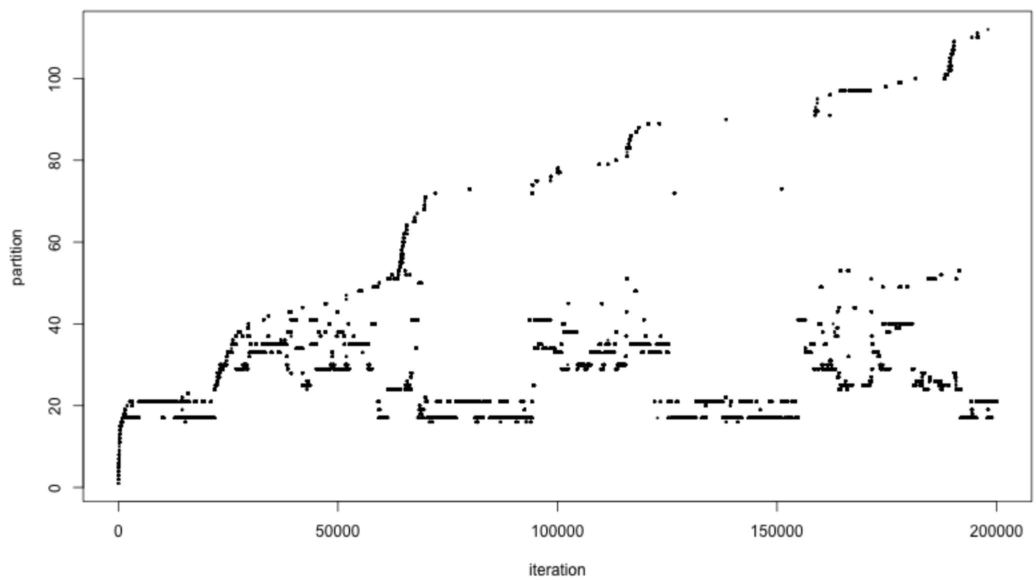


Figure 12: Trace plot for the partition parameter. The x-axis indexes iterations, and the y-axis labels the unique partitions visited by the Markov chain.

E.2 Juice Data

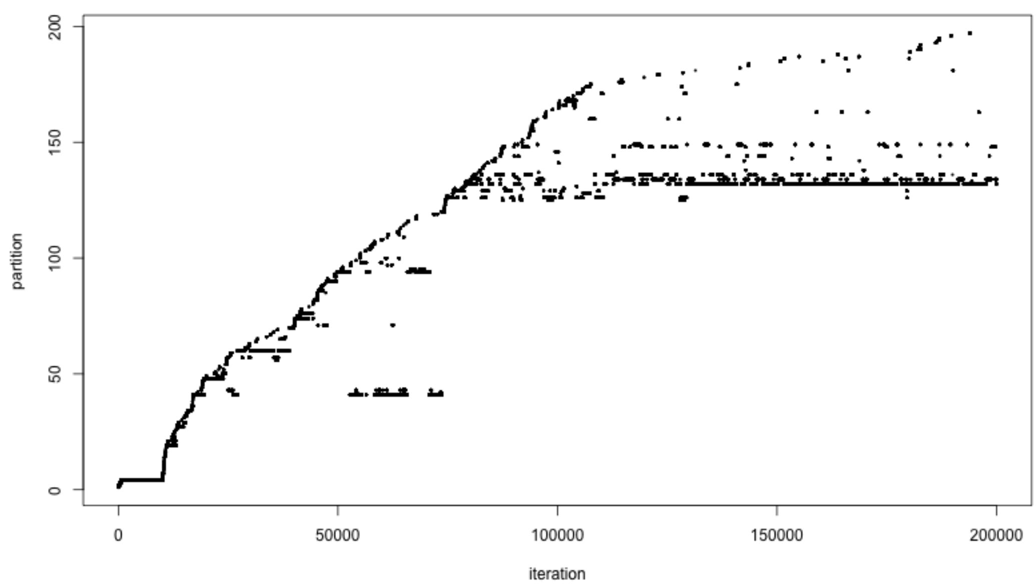


Figure 13: Trace plot for the partition parameter. The x-axis indexes iterations, and the y-axis labels the unique partitions visited by the Markov chain.

References

Bajari, P., Nekipelov, D., Ryan, S. P., and Yang, M. (2015). *Demand Estimation with Machine Learning and Model Combination.* Working paper.

Barten, A. (1964). Consumer Demand Functions under Conditions of Almost Additive Preferences. *Econometrica*, *32*(1-2), 1-38.

Blackwell, D., and MacQueen, J. B. (1973). Ferguson Distributions Via Polya Urn Schemes. *The Annals of Statistics*, *1*(2), 353-355.

Byron, R. (1970). A Simple Method for Estimating Demand Systems Under Separable Utility Assumptions. *The Review of Economic Studies*, *37*(2), 261-274.

Christensen, L. R., Jorgenson, D. W., and Lau, L. J. (1975). Transcendental Logarithmic Utility Functions. *The American Economic Review*, *65*(3), 367–383.

Deaton, A., and Muellbauer, J. (1980a). An Almost Ideal Demand System. *The American Economic Review*, *70*(3), 312–326.

Deaton, A., and Muellbauer, J. (1980b). *Economics and Consumer Behavior*. Cambridge University Press.

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, *1*(2), 209-230.

France, S. L., and Ghose, S. (2016). An Analysis and Visualization Methodology for Identifying and Testing Market Structure. *Marketing Science*.

Goldman, S., and Uzawa, H. (1964). A Note on Separability in Demand Analysis. *Econometrica*, *32*(3), 387-398.

Gorman, W. (1959). Separable Utility and Aggregation. *Econometrica*, *27*(3), 469-481.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, *82*(4), 711-732.

Leontief, W. (1947). Introduction to a Theory of the Internal Structure of Functional Relationships. *Econometrica*, *15*(4), 361-373.

Montgomery, A. L., and Rossi, P. E. (1999). Estimating Price Elasticities with Theory-Based Priors. *Journal of Marketing Research*, *36*(4), 413–423.

Müller, P., Quintana, F., and Rosner, G. L. (2011). A Product Partition Model With Regression on Covariates. *Journal of Computational and Graphical Statistics*, *20*(1), 260-278.

Park, J. H., and Dunson, D. B. (2010). Bayesian Generalized Product Partition Model. *Statistica Sinica*, *20*, 1203-1226.

Pollak, R. A., and Wales, T. J. (1969, October). Estimation of the Linear Expenditure System. *Econometrica*, *37*(4), 611–628.

Pudney, S. E. (1981). An Empirical Method of Approximating the Separable Structure of Consumer Preferences. *The Review of Economic Studies*, *48*(4), 561-577.

Ringel, D. M., and Skiera, B. (2016). Visualizing Asymmetric Competition Among More Than 1,000 Products Using Big Search Data. *Marketing Science*, *35*(3), 511–534.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing.* John Wiley and Sons Ltd.

Smith, A. N., and Allenby, G. M. (2016). *Demand Models with Random Partitions.* Working paper.

Song, I., and Chintagunta, P. K. (2006). Measuring Cross-Category Price Effects with Aggregate Store Data. *Management Science*, *52*(10), 1594–1609.

Sono, M. (1961). The Effect of Price Changes on the Demand and Supply of Separable Goods. *International Economic Review*, *2*(3), 239-271.

Theil, H. (1965, January). The Information Approach to Demand Analysis. *Econometrica*, *33*(1), 67-87.

Wedel, M., and Zhang, J. (2004). Analyzing Brand Competition across Subcategories. *Journal of Marketing Research*, *41*(4), 448–456.