

Inference for Product Competition and Separable Demand

Adam N. Smith, Peter E. Rossi, and Greg M. Allenby *

February 25, 2019

Abstract

This paper presents a methodology for identifying groups of products that exhibit similar patterns in demand and responsiveness to changes in price using store-level sales data. We use the concept of economic separability as the basis for establishing similarity between products, and build a weakly separable model of aggregate demand. A 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. We develop a methodology which allows the partition to be an estimated model parameter. In particular, we specify a log-linear demand system in which 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. We find evidence that the separable structure of demand can be inconsistent with category labels, which has implications for optimal category marketing strategies.

Keywords: price elasticities, category demand, dimension reduction, random partition models, Bayesian inference.

*Adam N. Smith, UCL School of Management, University College London, email: a.smith@ucl.ac.uk; Peter E. Rossi, Anderson School of Management, UCLA, email: perossichi@gmail.com; Greg M. Allenby, Fisher College of Business, Ohio State University, email: allenby.1@osu.edu. Researchers own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

This paper presents a methodology for measuring the dimension and boundaries of consumer demand using store-level scanner data. We formally define the boundaries of demand using the concept of economic separability: two product groups are said to be *weakly separable* if consumer preferences within each group are independent of consumption levels in the other group (Deaton and Muellbauer, 1980). For example, the demand for cereal would be weakly separable from the demand for milk only if consumer preferences between any two cereals (e.g., Cheerios and Frosted Flakes) do not depend on the type or amount of milk consumed. While separability assumptions remain widely used in empirical work, there is relatively little discussion of their practical viability (Chintagunta and Nair, 2011). In multi-category demand settings, for example, it is commonly assumed that preferences are separable across categories where the definition of categories is fixed according to product labels in the data. However, we believe that deciding on the separable structure of demand is an empirical question, especially when the goal is to measure substitution patterns within wide assortments or across multiple categories.

Our approach is to let the partition of products into separable groups be a parameter that can be estimated from the data. We specify an aggregate log-linear demand system in which the assumption of weakly separable preferences induces equality restrictions on a subset of cross-price elasticity parameters. Full inference in this model is non-trivial, as it requires quantifying the uncertainty in a partition, which is a discrete object defined over a high-dimensional and non-Euclidean space. To overcome this challenge, we take a Bayesian approach to inference and rely on recent advances in the literature on models for random partitions. Specifically, we use the location-scale partition model of Smith and Allenby (2018) to specify a coherent prior and efficiently navigate the posterior of both the partition and the lower-dimensional set of price elasticities. The novelty is that we are able to discover groups of separable products rather than just test whether a given set of groups is separable.

The proposed demand model is applied to weekly, store-level retail scanner data in two

product categories: (i) jams, jellies, and spreads and (ii) juice drinks. The jams, jellies, and spreads category contains two complementary subcategories – jams/jellies and spreads – which creates a strong prior on how to categorize products. That is, one may reasonably expect jams/jellies and spreads to constitute separable demand groups. This gives us the ability to examine the face validity of the partitions found by our method. In contrast, the juice category is comprised of wide assortments with many product lines spanning multiple flavor subcategories. The amount and type of variety makes it harder to define the nature of competition, and thus an appropriate product categorization, a priori.

Throughout both analyses, we find that partitions with high posterior probability can be markedly different from predefined category labels in both their granularity and composition. In the juice category, for example, we find evidence that demand is not separable by flavor categories (e.g., apple, orange). We then show how assuming the wrong structure of demand can lead to a misunderstanding of competition and biases in category promotion effects. An additional benefit of our methodology is that we can integrate over the partition uncertainty, which leads to improved estimates of cross-price effects and demand forecasts.

The idea of using economic separability to add structure to demand systems has a long history in marketing and economics. For example, much of the early empirical work in applied microeconomics imposes separability restrictions in order to facilitate estimation of demand models with highly aggregated goods, many parameters, and limited data (e.g., [Barten, 1964](#); [Byron, 1970](#); [Jorgenson and Lau, 1975](#)). While this additional structure provides gains in efficiency, it can also limit the set of admissible substitution patterns. [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 similar in spirit, as we use the data to find more suitable demand groups.

The availability of modern scanner data has since allowed for more disaggregate demand analysis. Researchers now have access to UPC-level price and movement data across many product categories, but for tractability reasons usually work with data from a single product

category or subcategory. Regardless of the scope of the data used, separability assumptions remain ubiquitous. For example, a common assumption in empirical work is that utility for the inside goods is separable from the utility of the outside good (Chintagunta and Nair, 2011). In multi-category demand models, it is also often assumed that utility is separable across pre-defined categories (e.g., Mehta, 2007; Thomassen et al., 2017). However, and as will be seen in the following section, the viability of these assumptions crucially depends on the researcher’s definition of categories. By allowing the composition of separable groups to be estimated from the data, we can offer insights into the boundaries of competition and if/where cross-category assumptions of separability appear unreasonable.¹

Separability has also shown to be useful for measuring price elasticities. For example, Montgomery and Rossi (1999) improve elasticity estimates in a log-linear demand system by shrinking the parameters towards values predicted by additive separability. The main difference between their approach and ours is the type of separability imposed on the model. In particular, they assume additive separability which requires all pairs of goods to be strong substitutes (Deaton and Muellbauer, 1980). This is sensible given their empirical focus on the demand for orange juice alone. We instead rely on weak separability which imposes a more flexible restriction across product groups and does not rule out complementarity a priori. This is motivated by our interest in modeling demand across wide assortments (e.g., the juice category), where products will be related but not necessarily substitutable.

More recently, there have been efforts to visualize and predict demand patterns for high-dimensional choice sets. Examples include France and Ghose (2016) and Ringel and Skiera (2016), who provide methods for visualizing competition for more than 1,000 products using consumer purchase and online search data, respectively. The output from both analyses is some form of competitive map rather than estimates of marketing responsiveness parameters. Jacobs et al. (2016) also study choice among large assortments, but focus on predicting future

¹Smith and Allenby (2018) also apply their partitioning methodology to a multi-category demand setting. In their demand model, the partition identifies *isolated* product groups exhibiting zero cross-price effects. While we utilize their random partition model, our work differs in that we focus on estimating *separable* demand models in which product groups have lower-dimensional, but not necessarily zero, cross-price effects.

purchases rather than identifying competitive effects. They leverage models in text analysis to improve large-scale customer purchase predictions. One fundamental difference in our work is the treatment of demand. While these methods abstract away from formal demand models, we embed the goal of identifying separable product groups within a traditional demand analysis where we estimate both the partition of goods and associated demand parameters. However, this comes at the cost of estimating a potentially large set of price effects, so the scalability of our model will depend on the type and amount of dimension reduction afforded by separability.

[Bajari et al. \(2015\)](#) provide an alternative dimension reduction solution for large-scale linear and logit demand models. In particular, they employ machine learning methods such as stepwise regression, support vector machines, and the lasso to reduce the number of relevant model covariates. However, many of these off-the-shelf methods provide a form of dimension reduction that lacks grounding in economic theory. For example, regularized regression estimators like the lasso can only accommodate shrinkage towards fixed constants that are usually set to zero. The benefits may thus be limited if the shrinkage points are misspecified. For instance, shrinkage to zero would be undesirable when applied to the price effect parameters of any Marshallian demand system, as zero substitution effects need not imply zero price effects. In this sense, separability offers a more flexible, theory-driven form of dimension reduction. Subsets of the full cross-price elasticity matrix are constrained to live in a lower-dimensional (and potentially non-zero) space of group-level parameters. This provides both adaptive and differential shrinkage, as elasticities are shrunk towards values consistent with theory, learned from the data, and that can change across product groups.

The remainder of the paper is organized as follows. [Section 2](#) develops a separable aggregate demand model. [Section 3](#) discusses probability models for random partitions. [Section 4](#) outlines our approach for posterior inference. [Section 5](#) presents the results of the two empirical studies. [Section 6](#) explores the implications of separable demand on product competition, brand positioning, and price promotions effects. [Section 7](#) 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 product 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 goods. 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 assumptions of separable preferences impose strong restrictions on demand elasticities when the partition is fixed, which means that the partition of products fully determines the nature of competition ([Deaton and Muellbauer, 1980](#)). 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.

2.1 Economic Separability

Let $\pi_n = (g_1, \dots, g_n)$ denote a 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 π_n if the utility of the consumption vector $q = (q_1, \dots, q_n)$ can be written as

$$u(q) = V(v_1(Q_1), \dots, v_K(Q_K)). \tag{1}$$

Here Q_k is the consumption set for products assigned to group k , $v_k(\cdot)$ is a subutility function, and $V(\cdot)$ is an aggregation function defined over K variables. To better understand the behavioral implications of separability, we discuss two additional (and ultimately equivalent) definitions.

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 .

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

Here $u_i(q)$ 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 complementary product groups: pasta and pasta sauce. The demand for pasta is weakly separable from the demand for sauce only if the consumer's preferences for two brands of pasta sauce (e.g., Prego and store brand) 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 consumers only prefer pairing certain types of pasta with high-quality brands of sauce, then 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.

$$S_{ij} = \phi_{k\ell}(q) \frac{\partial q_i}{\partial m} \frac{\partial q_j}{\partial m} \quad \text{for all } g_i = k, g_j = \ell, k \neq \ell \quad (3)$$

Here S_{ij} is a compensated price effect (i.e., the derivative of the Hicksian demand function for good i with respect to the price of good j), m is a monetary budget,² and $\phi_{k\ell}(q)$ is a

²In the separability literature, m is also commonly referred to as income (Goldman and Uzawa, 1964). This is likely because early empirical work concurrent to the development of separability focused on estimating complete demand systems with product sets spanning the entire space of commodities (e.g., food, housing, clothing) and a binding budget constraint on income. However, modern UPC-level scanner data have allowed

function summarizing the relationship between groups k and ℓ . In our empirical applications, we assume a constant function $\phi_{k\ell}(q) = \phi_{k\ell}$. 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, except as modified by budgetary effects. Again, if the demand of pasta is weakly separable from pasta sauce, then a decrease in the price of one sauce variety should have the same effect on the demand for all pasta varieties, barring budgetary effects. Goldman and Uzawa (1964) show that (1), (2), and (3) are all necessary and sufficient for each other, so that weakly separable preferences can be represented by any of the three.

A more restrictive form of separability is *strong* (or *block-additive*) separability, which assumes that the aggregation function $V(\cdot)$ in (1) is additive. There are two important empirical consequences of strong separability. The first is that the groups themselves cannot have special meaning, as the additivity of $V(\cdot)$ always permits the creation of new groups by arbitrarily combining other groups (Deaton and Muellbauer, 1980). Second, strong separability not only requires all cross-group price effects conform to (3), but also requires that $\phi_{k\ell}(q) = \phi(q)$ for all groups k and ℓ . However, like weak separability, strong separability still leaves the set of within-group elasticities unrestricted. In this sense, the partition parameter π_n is still meaningful in that it dictates which price effects are restricted.

A special case of strong separability is when each product constitutes its own group. In this case, preferences are said to be *additive*³ and all pairs of goods must be strong substitutes. For both strong and additive separability, the additional restriction of constant group-level price effects is very limiting a priori and will only be suitable for narrowly defined product categories. Since our goal is to model demand for wide assortments with potentially ambiguous economic relationships, the flexibility of weak separability is preferred. Going forward, we will use the terms separability and weak separability interchangeably.

researchers to measure substitution patterns at a finer level of granularity for a subset of the commodity space (e.g., breakfast cereals). In these settings, partial demand systems are usually estimated in which the budget constraint binds on a preallocated monetary budget that need not be equal to income.

³The distinction between block-wise additivity and additivity is also made in Pollak (1971) and Deaton and Muellbauer (1980, chapter 5.3).

2.2 A Weakly Separable Log-Linear Demand Model

There are generally two strategies for constructing separable demand models. The first approach is to construct a separable utility function and then solve for the associated demand functions. However, the construction of weakly separable utility functions is challenging in practice. For instance, using the functional form restriction in (1) requires the researcher to specify sufficiently flexible subutility and aggregating functions $v_k(\cdot)$ and $V(\cdot)$ that also yield tractable demand equations. This is infeasible when taking the primal approach, given that direct utility functions must be additively separable in order to yield a tractable demand system (Mehta, 2015). An alternative is to work with indirect utilities and impose the restriction in (2) on the marginal rates of substitution of a flexible functional form. However, Blackorby et al. (1977) show that many flexible indirect utility functions such as the translog are “separability-inflexible” in that, once restricted, they are no longer flexible enough to be regarded as approximations to any arbitrary weakly separable utility function.

The second approach, which we follow in this paper, is to impose separability restrictions on the demand function directly. Specifically, we impose the restriction in (3) on the price elasticity parameters of a log-linear demand system.

$$\log q_{it} = \sum_{j=1}^n \beta_{ij} \log p_{jt} + \gamma_i \log m_t + z'_{it} \psi_i + \varepsilon_{it} \quad (4)$$

$$\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{nt}) \sim N(0, \Sigma)$$

Here q_{it} is the movement for product i at time t and p_{jt} is the price of product j at time t . The variable m_t represents total expenditure at time t and can be viewed as an aggregated sample analogue to the individual-level monetary budget term discussed in the previous section. It is important to include expenditure in the demand equation because (3) expresses price effects as a function of expenditure effects. We also include a d -dimensional vector of product-specific controls z_{it} such as seasonal effects or promotional activity. The demand equations are tied together contemporaneously through the $n \times n$ error covariance matrix Σ .

The log-linear functional form is attractive because it is linear in price and expenditure

elasticities and admits flexible substitution patterns. This has led to its widespread use for elasticity-based pricing in both industry and academia (Montgomery, 1997; DellaVigna and Gentzkow, 2017; Hitsch et al., 2017). For our purposes, this functional form also makes it straightforward to project β_{ij} onto the lower-dimensional space of separability parameters defined in (3). In fact, we later show that this restricted model retains its linearity in the lower-dimensional set of parameters, which enables efficient posterior computation.

Since separability is fundamentally a restriction on consumer preferences, it is also worth discussing the extent to which the demand equation in (4) is consistent with utility theory. In general, a demand system is said to be integrable (i.e., it can be rationalized by a model of constrained utility maximization) if it satisfies four properties: adding-up, homogeneity, Slutsky symmetry, and negativity (Deaton and Muellbauer, 1980). In a log-linear demand system, adding-up is the most straightforward to impose a priori as it requires all expenditure elasticities to equal one. The other three conditions place restrictions on functions of model parameters.⁴ However, requiring these conditions to hold everywhere yields a model in which a subset of own-price elasticities must equal negative one and all cross-price elasticities either equal zero or $1 + \beta_{ii}$ (LaFrance, 1986). Since this renders the log-linear model practically useless for measuring substitution patterns, we instead treat (4) as a flexible first-order approximation to a valid Marshallian demand system. In our empirical applications, we impose the adding-up restriction and then use the posterior distribution of cross-price effects to examine whether the remaining integrability conditions hold in the data.

Our separable log-linear model is then defined as (4) subject to the restriction in (3), which we translate from S_{ij} to β_{ij} and impose at the average expenditure share for each good \bar{w}_j (see Appendix A for details on the derivation).

$$\beta_{ij} = \begin{cases} \bar{w}_j(\theta_{k\ell}\gamma_i\gamma_j - \gamma_i) & \text{if } g_i \neq g_j \\ \eta_{ij} & \text{if } g_i = g_j \end{cases} \quad (5)$$

⁴Homogeneity: $\sum_j \beta_{ij} + \gamma_i = 0 \forall i$; Slutsky symmetry: $\beta_{ij}q_i/p_j + \gamma_i = \beta_{ji}q_j/p_i + \gamma_j \forall i, j$; and negativity: $\beta_{ii} \leq \gamma_i w_i \forall i$.

Here η_{ij} is an unrestricted within-group elasticity and $\theta_{k\ell}$ is a separability parameter that we further restrict to be symmetric in order to satisfy group-level Slutsky symmetry. For two normal goods i and j belonging to different groups k and ℓ , it is now that case that

$$\beta_{ij} > 0 \iff \theta_{k\ell} > \frac{1}{\gamma_j}. \quad (6)$$

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.

There is still a question of whether the log-linear model subject to the restriction in (5) can approximate a demand system generated by some weakly separable utility function. As shown in Blackorby et al. (1977), the concern is that the separability constraints are too restrictive in that the model no longer has enough parameters to approximate the underlying separable model. To address this issue, Driscoll et al. (1992) show that a model needs $1 + n + \sum_k n_k(n_k + 1)/2 + K(K - 1)/2$ parameters to constitute a flexible approximation to a weakly separable process, where n_k denotes the number of products assigned to group k . Since our model contains $\sum_k n_k^2 + K(K - 1)/2$ price effects, this condition will be satisfied as long as we include n product intercepts and $n_k > 1$ for at least one group.⁵

2.3 Related Models

Our aim in this paper is to study the structure of aggregate demand both within and across highly differentiated product categories. It is then necessary that our empirical demand specification allow for flexible substitution patterns. A natural starting point is to consider the unrestricted log-linear demand system in (4). The main drawback is that it requires n^2 price elasticity parameters to be estimated, which can be demanding on the data for even a moderate number of products. Moreover, the assessment of product competition is non-trivial, as it requires careful examination of a potentially large elasticity matrix.

Separability offers a dimension reduction solution through the introduction of product

⁵For example, if $\pi_n = (1, 1, 1, 2, 2, 2)$ then $n + \sum_k n_k^2 + K(K - 1)/2 = 6 + 18 + 1 = 25$ which exceeds the required $1 + n + \sum_k n_k(n_k + 1)/2 + K(K - 1)/2 = 1 + 6 + 12 + 1 = 20$.

groups, but this comes at the cost of flexibility. By treating the partition as a model parameter, we accomplish three goals: (i) we endogenize the separable structure of demand; (ii) we reduce the dimension of the parameter matrix; and (iii) we retain much of the flexibility of the unrestricted model a priori. Estimating our separable model would be preferred over an unrestricted model whenever the researcher has insufficient data to precisely estimate price effects or directly cares about inferring the boundaries of competition.

An alternative approach commonly taken in the differentiated products literature is to use a mixed or random coefficients logit demand model, which is flexible in the sense that its choice probabilities can approximate the choice probabilities from any underlying utility function (Berry et al., 1995; McFadden and Train, 2000). The added flexibility over the homogeneous logit model comes from the assumption that consumers are heterogeneous in their tastes for product characteristics. This allows the cross-price elasticity between two goods to be driven by their similarity in the characteristics space.

Although mixed logit models benefit from a formal treatment of utility theory and aggregation, there are a few reasons why they are not necessarily suitable for our purposes. First, the practical flexibility of the mixed logit is unclear, as the result of McFadden and Train (2000) only holds for *some* appropriately specified mixing distribution and when arbitrarily high-order polynomials of price are allowed to enter the logit expression. Second, mixed logit models assume that market shares result from the aggregation of discrete choices across consumers. When modeling demand for wide assortments or across categories, the assumption of discrete choice is unreasonable. Third, accurate measurement of cross-price effects depends on the availability of rich product characteristics data, which are not always available. The flexibility in our model is not driven by characteristics (although the model can be extended to include them if desired). Finally, it is not obvious how to identify separable product groups within a mixed logit demand model. While it is possible to use dimension reduction or clustering techniques on the induced price elasticity matrix, there is no reason to believe that those groups would actually reflect the separable structure of demand.

3 Models for Random Partitions

The separable demand model proposed in Section 2 is defined conditional on a partition π_n . This section outlines probability models for π_n that enable joint posterior inference. Since random partition models have seen little application in the field of marketing, we first present a general discussion of their foundation and connection with Bayesian nonparametrics. Then, we review the more recent work of [Smith and Allenby \(2018\)](#) who develop a location-scale partition model which we use as the proposal distribution in our posterior sampling routine.

The construction of models for random partitions is often based on sampling from a discrete random probability measure, which is a central topic of Bayesian nonparametric methods. 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 draws from a Dirichlet process. That is, if a set of random variables ϕ_1, \dots, ϕ_n 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 ([Figure 1](#)).

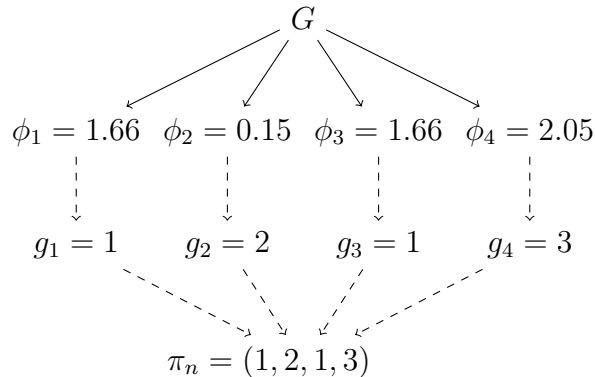


Figure 1: Four random variables are drawn from an almost surely discrete distribution G . The realized values are then used to create the vector item-group assignment variables g_1, \dots, g_4 that constitute a partition. The solid lines indicate probabilistic sampling, while the dashed lines indicate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

3.1 Random Partitions and the Dirichlet Process

Formally, consider sampling $\phi_i|G \sim G$ for $i = 1, \dots, n$ where $G \sim \text{DP}(\alpha, G_0)$. The precision parameter $\alpha > 0$ and base distribution G_0 are fixed hyperparameters. Blackwell and MacQueen's Pólya urn scheme provides a way to generate the vector of ϕ 's: draw $\phi_1 \sim G_0$ and then sample $\phi_i|\phi_{<i}$ for $i = 2, \dots, n$ according to

$$\phi_i|\phi_{<i} \sim w_0 G_0(\phi_i) + \sum_{k=1}^{K^{(i)}} w_k \delta_{\phi_k^*}(\phi_i) \quad (7)$$

where

$$w_0 = \frac{\alpha}{\alpha + i - 1} \quad \text{and} \quad w_k = \frac{n_k}{\alpha + i - 1}. \quad (8)$$

Here ϕ_k^* is the k th unique value in the set $\phi_{<i} = (\phi_1, \dots, \phi_{i-1})$, $\delta_{\phi_k^*}(\cdot)$ is a point-mass at ϕ_k^* , $K^{(i)}$ is the number of unique values in $\phi_{<i}$, and n_k is the number of ϕ 's equal to ϕ_k^* . The weights w_0 and w_k represent the probability that ϕ_i is a new draw from the base distribution or exactly equal to ϕ_k^* , respectively. Since our focus is on the induced partition rather than the distribution G , the choice of G_0 is immaterial since the value of each ϕ_i only serves as a group membership label.

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 $\phi = (\phi_1, \dots, \phi_n)$ into a partition $\pi_n = (g_1, \dots, g_n)$ using the following mapping:

$$g_i = k \quad \text{if} \quad \phi_i = \phi_k^* \quad \text{for} \quad i = 1, \dots, n \quad \text{and} \quad k = 1, \dots, K \quad (9)$$

where K denotes the number of groups. Since the ϕ 's are generated stochastically, the Pólya urn scheme in (7) induces a probability distribution $p(\pi_n)$ defined over the space of possible partitions \mathcal{P}_n . The sequential nature of (7) also leads to simple probability calculations:

$$p(\pi_n) = \prod_{i=2}^n p(g_i|g_{<i}) \quad (10)$$

where $g_{<i} = (g_1, \dots, g_{i-1})$ and

$$p(g_i|g_{<i}) = \begin{cases} w_0 & \text{if } g_i \notin \{g_1, \dots, g_{i-1}\} \\ w_{g_i} & \text{otherwise.} \end{cases} \quad (11)$$

We next describe three desirable properties of random partition models based on Pólya urn sampling schemes.

Property 1 (Endogenous K). *If π_n is a random partition generated from a Pólya urn scheme, then the number of groups $K = \max\{g_1, \dots, g_n\}$ is also random. Consequently, the partitioning model $p(\pi_n)$ induces a probability model over the number of groups $p(K)$.*

In many existing models of categorization (e.g., models based on the multinomial distribution), the number of groups must be fixed by the researcher a priori. Partitioning models based on Pólya urn schemes instead allow us to make inference about partitions and the number of groups jointly.

Property 2 (Order Restriction). *If π_n is a random partition generated from a Pólya urn scheme, then the elements of π_n conform to the following order restriction:*

$$g_1 = 1 \quad \text{and} \quad g_i \in \{1, \dots, K^{(i)} + 1\} \quad \text{for } i = 2, \dots, n \quad (12)$$

where $K^{(i)} = \max\{g_1, \dots, g_{i-1}\}$.

This property prevents any issues of label switching, which are also common in models of categorization. For example, consider two equivalent partitions $\pi_n = (1, 1, 2, 1)$ and $\pi'_n = (2, 2, 1, 2)$. The order restriction only permits π_n to be generated from the model, not π'_n . We will let \mathcal{P}_n denote the set of all possible partitions of n items subject to this restriction.

Property 3 (Marginalization). *Let π_n be a random partition of n items and π_{n+1} be a random partition of $n + 1$ items. If π_n and π_{n+1} are generated from Pólya urn schemes, then the probability model for π_n arises from marginalizing the model for π_{n+1} .*

$$p(\pi_n) = \sum_{g_{n+1}=1}^{K+1} p(\pi_{n+1}) \quad (13)$$

This property ensures that a partitioning model for n items does not depend on whether we consider an additional item, at least a priori. Further discussion of the marginalization property can be found in [Müller et al. \(2015\)](#).

The problem with the specific probability model for π_n induced by the Dirichlet process is that it only depends on α and n_k . Consequently, it is hard for the researcher to control the partitioning process or form expectations about the composition of sampled partitions. 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 π_n through the single precision parameter 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 around the current state when generating a candidate partition.

3.2 The Location-Scale Partition Distribution

[Smith and Allenby \(2018\)](#) address these limitations by developing the location-scale partitioning model, denoted $\text{LSP}(\rho_n, \tau)$. The idea is to parameterize the group-assignment probabilities in (7) with a location partition $\rho_n = (s_1, \dots, s_n)$ and scale parameter $\tau > 0$.

$$\phi_i | \phi_{<i}, \rho_n, \tau \sim w_0(\rho_n, \tau) G_0(\phi_i) + \sum_{k=1}^{K^{(i)}} w_k(\rho_n, \tau) \delta_{\phi_k^*}(\phi_i) \quad (14)$$

Here $w_0(\cdot)$ and $w_k(\cdot)$ are positive functions that satisfy $w_0(\cdot) + \sum_k w_k(\cdot) = 1$. The composition of groups is now controlled by the information in ρ_n and τ . Holding τ fixed, two items that are grouped together in ρ_n are also more likely to be grouped together in π_n . The specific functional forms for $w_0(\cdot)$ and $w_k(\cdot)$ are given by

$$w_k(\rho_n, \tau) = \tilde{c}_i \cdot \frac{\tau + n_{S_k}^{s_i}}{\tau C^{(i)} + \tau + n_k} \quad (15)$$

$$w_0(\rho_n, \tau) = \tilde{c}_i \cdot \frac{\tau + \mathbf{I}(s_i = C^{(i)} + 1)}{\tau C^{(i)} + \tau + 1} \quad (16)$$

where $S_k = \{s_j : g_j = k \text{ and } j < i\}$, $C^{(i)} = \max\{s_1, \dots, s_{i-1}\}$, $n_{S_k}^c$ counts the number of elements in S_k equal to c , and \tilde{c}_i is a normalizing constant.

Since (14) is a modified Pólya urn scheme, the LSP probability model will automatically satisfy Properties 1-2 described above. The functional forms for $w_0(\cdot)$ and $w_k(\cdot)$ also guarantee that the LSP model satisfies Property 3 as well as the following consistency property (see [Smith and Allenby \(2018\)](#) for formal proofs).

Property 4 (Location-Scale Consistency). *If $\pi_n \sim LSP(\rho_n, \tau)$, then for any number of items n and location partition $\rho_n \in \mathcal{P}_n$,*

$$\lim_{\tau \rightarrow 0} \Pr(\pi_n = \rho_n | \rho_n, \tau) = 1. \quad (17)$$

This property ensures that the $LSP(\rho_n, \tau)$ distribution behaves like a continuous, univariate distribution belonging to a location-scale family. That is, ρ_n shifts the center of the distribution on \mathcal{P}_n and τ determines the amount of dispersion around ρ_n . As τ goes to zero, more probability mass is placed on ρ_n . As τ gets large, the mass is spread more evenly over \mathcal{P}_n . An illustration of the differences between the partition distribution induced by the Dirichlet process and the LSP distribution is provided in the online appendix.

For our purposes, the LSP distribution facilitates estimation and inference in two ways. First, it serves as a prior distribution for π_n , where prior knowledge can be directly imposed through the choice of location partition ρ_n 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 candidate partition π_n^* is proposed from an $LSP(\pi_n, v)$ distribution. Here π_n is the partition in the current state and v is a step size parameter. Candidate partitions will then be generated from a distribution centered around the partition in the current state, where the similarity between π_n^* and π_n is controlled by v . The resulting Markov chain will be able to navigate the partition space much more effectively than alternative independence proposal mechanisms.

4 Posterior Inference

4.1 Likelihood

The unrestricted log-linear demand system has a convenient statistical representation. While log-linear models usually take the form of a multivariate regression, we rewrite (4) as a seemingly unrelated regression (SUR) model to account for the product-specific control variables.

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (18)$$

$$y = X\Omega + \varepsilon \quad (19)$$

Here $y_i = \log q_i$ is a T -dimensional vector of log demand and $X_i = (P, M, Z_i)$ where P is a $T \times n$ matrix of log prices, M is a T -dimensional vector of log expenditure, and Z_i is a $T \times d$ matrix of product-specific controls. Further, let $\omega'_i = (\beta'_i, \gamma_i, \psi'_i)$ where β_i is an n -dimensional vector of price elasticities, γ_i is a scalar expenditure elasticity, and ψ_i is a d -dimensional vector of control coefficients. Finally, the T -dimensional error vector follows a multivariate normal distribution with an expanded covariance matrix $\varepsilon \sim N(0, \Sigma \otimes I_T)$, which induces a multivariate normal likelihood for the data.

4.2 Computation

In its unrestricted form, sampling from the posterior of the model in (19), subject to the specification of proper conjugate priors, can be achieved using the SUR Gibbs sampling routine outlined in Chapter 3 of Rossi et al. (2005). Posterior inference for the model in (19) subject to separability restrictions in (5) is more challenging. This is because the separability restrictions both (i) parameterize the conditional mean function as a nonlinear function of γ ,

which breaks the conjugacy that enables the use of Gibbs sampling; and (ii) project a subset of the elasticity parameters onto a lower-dimensional space whose dimension can change with the partition.

We address these challenges by rewriting the mean function:

$$y = X(\pi_n, \gamma; P, w)\beta_{\pi_n} + c(\pi_n, \gamma; P, w) + M\gamma + Z\psi + \varepsilon \quad (20)$$

where $\beta'_{\pi_n} = (\eta', \theta')$ denotes the joint vector of unrestricted elasticities η and restricted separability parameters θ . The first term presents the key result in that, once the suitable $X(\cdot)$ matrix is constructed, the model remains linear in β_{π_n} . The second term $c(\pi_n, \gamma; P, w)$ represents the remaining restrictions in (5) not contained in $X(\cdot)$. Specifically, let $c(\pi_n, \gamma; P, w) = \text{vec}(PC)$ where C is an $n \times n$ matrix with elements $C_{ij} = \bar{w}_j \gamma_i \mathbf{I}(g_i \neq g_j)$. The next two terms include log expenditure and product-specific controls, which are parameterized just as they are in (19) with $M = \text{diag}\{M, \dots, M\}$ and $Z = \text{diag}\{Z_1, \dots, Z_n\}$.

Next, we turn to the construction of $X(\cdot)$ which requires binding together the appropriate prices that are informative about η and θ , respectively. This gives rise to the following structure:

$$X(\pi_n, \gamma; P, w) = \begin{pmatrix} U_1(\pi_n) & 0 & \cdots & 0 & R_{11}(\pi_n, \gamma) & \cdots & R_{1n_\theta}(\pi_n, \gamma) \\ 0 & U_2(\pi_n) & \cdots & 0 & R_{21}(\pi_n, \gamma) & \cdots & R_{2n_\theta}(\pi_n, \gamma) \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & U_n(\pi_n) & R_{n1}(\pi_n, \gamma) & \cdots & R_{nn_\theta}(\pi_n, \gamma) \end{pmatrix} \quad (21)$$

where the block diagonal matrix in the first n columns constitutes the unrestricted covariates for η , and the remaining columns constitute the restricted covariates for θ . Specifically, $U_i(\pi_n)$ is the $T \times n_{\eta_i}$ matrix of log prices corresponding to the set of product i 's unrestricted elasticities η_i where $n_{\eta_i} = \dim(\eta_i)$. It is a function of π_n because it is formed by combining all columns j of P such that $g_i = g_j$. Each element of the remaining matrix is a $T \times 1$ vector

defined as

$$R_{ij}(\pi_n, \gamma) = \begin{cases} \sum_{\{r: g_r \in \mathcal{S}_j \cap g_r \neq g_i\}} \bar{w}_r \gamma_i \gamma_r \log p_r & \text{if } g_i \in \mathcal{S}_j \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

where $n_\theta = \dim(\theta)$ and $\mathcal{S}_j = \{k, \ell\}$ stores the pair of group indices (k, ℓ) corresponding to the j th element of the θ vector. For example, if $\pi_n = (1, 1, 2, 2, 3, 3)$, then $n_\theta = 3$ and $\mathcal{S}_1 = \{1, 2\}$, $\mathcal{S}_2 = \{1, 3\}$, and $\mathcal{S}_3 = \{2, 3\}$.

The full set of model parameters includes the vector of unrestricted elasticities and separability parameters β_{π_n} , the expenditure elasticities γ , the partition of goods π_n , the coefficients on the controls ψ , and the error covariance matrix Σ . Our posterior sampling strategy directly follows from the structure of the mean function in (20) and takes advantage of conditional conjugacy. We outline the two-step posterior sampling routine below.

Step 1 (Σ, ψ): To sample Σ and ψ from their posterior conditional on $(\beta_{\pi_n}, \pi_n, \gamma)$, let

$$y - X(\pi_n, \gamma; P, w)\beta_{\pi_n} - c(\pi_n, \gamma; P, w) - M\gamma = Z\psi + \varepsilon \quad (23)$$

which is a standard SUR model. Thus, we can sample from this conditional posterior directly, subject to the specification of conjugate priors: $\psi \sim N(\bar{\psi}, A_\psi^{-1})$ and $\Sigma \sim IW(\nu_0, V_0)$.

Step 2 ($\beta_{\pi_n}, \gamma, \pi_n$): To sample $(\beta_{\pi_n}, \pi_n, \gamma)$ from their posterior conditional on (Σ, ψ) , let

$$y - Z\psi = X(\pi_n, \gamma; P, w)\beta_{\pi_n} + c(\pi_n, \gamma; P, w) + M\gamma + \varepsilon. \quad (24)$$

This model is linear in β_{π_n} but not in γ or π_n , so direct sampling from the associated posterior is not possible. We instead rely on a Metropolis-Hastings step. The key is to construct efficient proposal distributions that ensure proper mixing and dimension matching between β_{π_n} and π_n .

We construct our joint proposal distribution as follows:

$$q(\beta_{\pi_n}^*, \pi_n^*, \gamma^* | \beta_{\pi_n}, \pi_n, \gamma) = q_1(\beta_{\pi_n}^* | \pi_n^*, \gamma^*) q_2(\pi_n^* | \pi_n) q_3(\gamma^* | \gamma) \quad (25)$$

where $(\beta_{\pi_n}^*, \pi_n^*, \gamma^*)$ denote the candidate values and $(\beta_{\pi_n}, \pi_n, \gamma)$ denote the current values.

Notice that conditional on values of π_n and γ , we can rewrite (24) as another SUR-like model which is linear in β_{π_n} .⁶ The benefit is that, subject to the specification of a conditionally conjugate prior $\beta_{\pi_n}|\pi_n \sim N(\bar{\beta}_{\pi_n}, A_{\beta}^{-1})$, we can directly sample from the associated conditional posterior (see Appendix B for details). This results in a highly efficient independence proposal given that $q_1(\cdot)$ is the exact conditional posterior of β_{π_n} . Two marginal random-walk proposals can then be used for π_n and γ . In particular, we choose $q_2(\cdot)$ to be the LSP distribution of Smith and Allenby (2018) so that $\pi_n^* \sim \text{LSP}(\pi_n, v)$ where v is a step size. The efficiency of this approach comes from being able to generate π_n^* from a distribution “centered” around the current partition, as in any ordinary random-walk step.

There are two benefits from the joint proposal strategy outlined above. First, the resulting ratio of likelihoods will always be tractable in the sense that the dimension of β_{π_n} is guaranteed to conform with π_n . In fact, recognizing that $q_1(\cdot)$ is the conditional posterior of β_{π_n} , it becomes straightforward to show the acceptance probability reduces to

$$\mathcal{A} = \min \left\{ 1, \frac{m^*(y|\pi_n^*, \gamma^*, \Sigma)p(\pi_n^*)p(\gamma^*)q_2(\pi_n|\pi_n^*)}{m(y|\pi_n, \gamma, \Sigma)p(\pi_n)p(\gamma)q_2(\pi_n^*|\pi_n)} \right\} \quad (26)$$

where

$$m(y|\pi_n, \gamma, \Sigma) = \int p(y|\beta_{\pi_n}, \pi_n, \gamma, \Sigma)p(\beta_{\pi_n}|\pi_n)d\beta_{\pi_n}. \quad (27)$$

Therefore, the proposed sampler is equivalent to a sampler that marginalizes over β_{π_n} when navigating the posterior of π_n .

The second benefit is that generating $\beta_{\pi_n}^*$ from its full conditional distribution independent of β_{π_n} 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 such as the reversible-jump MCMC sampler of Green (1995). However, these methods require complicated transition probability calculations and become inefficient in high dimensions. In contrast, we only use random-walk proposals to marginally navigate a high-dimensional (but fixed) partition space, and then use efficient independence

⁶The model is not exactly a SUR model because the design matrix $X(\pi_n, \gamma; P, w)$ is not block diagonal.

proposals for β_{π_n} . Therefore, the transition probabilities in the proposed routine only require evaluating a multivariate normal density and the LSP probability mass function.

4.3 Specification of Priors

Care must be taken when choosing hyperparameters in the prior distributions of β_{π_n} and π_n . This is because the acceptance probability in the proposed MCMC routine is driven by the ratio of marginalized likelihoods that are integrated over β_{π_n} . Given that the dimension of β_{π_n} depends on π_n , the integrals in the numerator and denominator of (26) may be of differing dimensions and will penalize models with many parameters in the presence of diffuse conditional priors on β_{π_n} (Lindley, 1957). In the context of our separable demand model where $n_\eta = \sum_i n_{\eta_i}$ grows faster than n_θ , this means that the ratio of marginalized likelihoods will naturally favor models with fewer unrestricted elasticity parameters and many groups.

In situations with limited data (where the prior dominates the likelihood) or when the researcher wants to penalize models with many groups, this effect can be countered through a judicious choice of $\bar{\beta}_{\pi_n}$, A_β , ρ_n , and τ . For example, researchers often specify diffuse priors in regression models by choosing $\bar{\beta}_{\pi_n} = 0$ and $A_\beta = .01I$. However, choosing $A_\beta = .1I$ may be more appropriate, as it reflects the knowledge that elasticities rarely fall outside the range $(-6, 6)$. One may also choose ρ_n to have few groups with a small value of τ . Doing so allows the marginal prior on π_n to penalize partitions with many groups. See the online appendix for a simulation study that further explores the effect of the prior on posterior inference.

4.4 Scalability

Although the SUR-like representation of the model likelihood enables efficient posterior inference, there are some limitations that arise when n is large. First, the joint proposal strategy outlined in Section 4.2 requires us to sample from the conditional posterior of β_{π_n} given all other parameters and data. The linearity of the model makes direct sampling possible, but potentially computationally expensive. This is because the dimension of this

conditional posterior is $n_{\beta_{\pi_n}} = n_{\eta} + n_{\theta} = \sum_k n_k^2 + K(K-1)/2$, which grows quadratically in both the number of unrestricted parameters and groups. For example, if there are $K = 10$ groups with each having $n_k = 10$ products, then $n_{\beta_{\pi_n}} = 1000 + 10(9)/2 = 1045$. In contrast, if there are only $K = 2$ groups with each having $n_k = 50$ products, then $n_{\beta_{\pi_n}} = 5000 + 1 = 5001$. Thus, the dimension of the conditional posterior can change dramatically with π_n .

As with any Bayesian SUR model, drawing from this conditional posterior distribution of β_{π_n} requires computing its posterior mean and the inverse of the Cholesky root of its posterior covariance matrix in each iteration of the Markov chain (Rossi et al., 2005). Since the posterior moments depend on the design matrix $X(\pi_n, \gamma; P, w)$, which in turn depends on π_n , we must also construct $X(\pi_n, \gamma; P, w)$ in each iteration. While this adds to the complexity, we avoid the cost of multiplying these large matrices by exploiting the partial sparsity of $X(\pi_n, \gamma; P, w)$ and instead construct its crossproduct directly.⁷

Figure 2 illustrates the computational cost of sampling from the conditional posterior of β_{π_n} . The computation time we measure includes the cost of constructing the crossproduct of $X(\pi_n, \gamma; P, w)$, computing the posterior mean, and taking the inverse of the Cholesky root of the posterior covariance matrix.⁸ In this exercise, we always define π_n to be a partition with two groups ($n_{\theta} = 1$), but vary the number of elements in each group $n_k \in \{5, 10, 15, \dots, 50\}$. This leads to models with $n \in \{10, 20, 30, \dots, 100\}$ products and $n_{\eta} \in \{50, 200, 450, \dots, 5000\}$ unrestricted elasticities. We plot the average computation time (across 100 replicates) against the number of unrestricted elasticities, where the number of observations is fixed at $T = 200$ across all models. We find that the computation cost increases nonlinearly with n_{η} .

⁷A similar strategy is used to sample from the posterior of the standard Bayesian SUR model in the `bayesm` package (Rossi, 2017) in R.

⁸All posterior sampling functions are written using the `Rcpp` package (Eddelbuettel and François, 2011) in R.

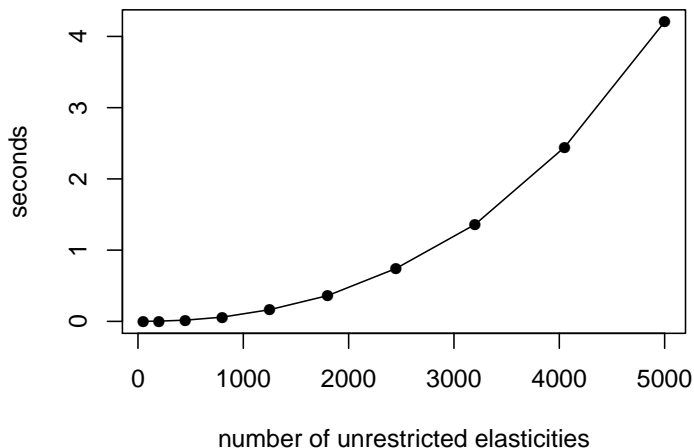


Figure 2: Average computation time in seconds (over 100 replicates) for a single draw from the conditional posterior of β_{π_n} for various model sizes.

Our model also requires navigating the space of partitions, which grows super-exponentially in n (Bell, 1934). LSP proposals allow us to navigate this space efficiently for moderate n , but are subject to the same scaling limitations as any random-walk Metropolis-Hastings algorithm. That is, there is a question of how to choose the step size of the LSP proposal distribution in order to ensure proper mixing as n gets large. Smith and Allenby (2018) find that using $v = 1/(n \log n)$ appropriately scales the dimension of \mathcal{P}_n and demonstrate practical convergence for up to $n = 100$ items. Higher-dimensional product sets can still be accommodated in our model, but may require stronger prior restrictions on the partition (e.g., assuming that a set of products must always belong to the same group).

Finally, there is a question of how the number of products n is allowed to scale relative to the number of observations T . In online retail, for example, there is a practical interest in “high-dimensional” regression-based demand models in which the number of products and associated price effects exceeds the number of observations. The challenge is in this setting is that the least squares estimator of the unrestricted price elasticity matrix is no longer well-defined. Note that while a Bayesian estimator of the same price elasticity matrix is not subject to this limitation because of the regularity imposed by the prior, the associated

posterior would still exhibit significant uncertainty. Since we take a Bayesian approach to inference and impose further parameter restrictions, our model will generate a well-defined posterior even when $n > T$. However, this high-dimensional setting is not our inherent focus for a few reasons.

First, our analysis is motivated by the fact that separability assumptions remain widely (and often implicitly) used in large and small-scale demand problems alike. One of our goals is to assess the viability of these assumptions by providing empirical insights into the separable structure of demand. Although the choice of separable groups becomes more difficult when product space is really large, to our knowledge there is no other work that attempts to identify separable demand groups even within a single product category. Instead of focusing on scale alone, we aim to provide a more comprehensive description of competition within broad and highly differentiated product categories. Moreover, in our empirical applications we focus on competition between brands and have found that the number of brands that constitute the majority of total category sales is small relative to the number of UPCs. Thus, for a brand-level analysis, the product set need not be extremely large in order to provide a complete description of category competition.

That being said, the core dimension reduction offered by separability remains attractive in very high-dimensional settings. Joint inference for π_n and β_{π_n} is impractical when $n > T$ due to the complexity of the partition space \mathcal{P}_n and the limited number of observations. However, if the goal is to simply forecast demand or generate estimates of price effects (which are otherwise intractable), then our recommendation is to fix both the expenditure elasticities (e.g., $\gamma_i = 1$ for all i) and partition. Doing so greatly reduces complexity of our sampling algorithm by removing the need to implement a Metropolis-Hastings step, which in turn makes posterior inference more practical when n is large (e.g., in the hundreds). In the online appendix, we also use a simulated data to show how the separable demand model performs in $n > T$ settings. Because the core focus of this paper is to uncover the separable structure of demand through posterior inference on π_n , we assume $T > n$ going forward.

5 Empirical Analysis

5.1 Data

Our data come from the Nielsen Retail Measurement System (RMS) database, which contains weekly movement, pricing, and store environment information for over 90 retail chains across the United States.⁹ Within the RMS database, UPCs are grouped into 10 departments (e.g., “dry grocery”), 125 product groups (e.g., “snacks”), and 1,100 product modules (e.g., “potato chips”). We construct two data sets from two distinct product groups: (*i*) jams, jellies, and spreads and (*ii*) juice drinks. The jams, jellies, and spreads product group consists of 9 product modules, while the juice product group consists of 18.¹⁰ For simplicity, we use data from a single store in Cincinnati, OH for which in-store promotional activity is available. Our sample period spans four years from 2008-2011, and we use a random sample of 52 weeks as a hold-out sample.

For each product group, we aggregate UPCs to the brand level due to high collinearity of prices within a brand (e.g., “JIF CRM H PLS 28” and “JIF CHK H PLS 28”). In the jams, jellies, and spreads category, we also combine the jams and jellies modules due to near perfectly collinear prices. We believe that a brand-level analysis is the most natural for considering the formation of separable demand groups. For example, it seems unreasonable to think that rate of substitution between an 18 oz. jar of Jif peanut butter and an 18 oz. jar of Skippy peanut butter would ever be independent of the demand for the exact same Jif offering in a 28 oz. jar. Our sample of products includes the majority of brands that have a within-module share of volume exceeding 1% and are present for each year in the sample. For each brand, we then create volume-weighted prices that are standardized prices to the most common pack-size in the product group.

⁹The data are provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business. <http://research.chicagobooth.edu/nielsen>

¹⁰*Jams, Jellies, and Spreads*: butter-fruit and honey, fruit spreads, garlic spreads, honey, jams, jelly, marmalade, peanut butter, preserves. *Juice Drinks*: apple, cider, clam, cranberry, fruit drinks, fruit drinks (can), grape, grapefruit, grapefruit (can), lemon/lime, nectars, orange, orange (can), pineapple, prune, tomato, vegetable (other), remaining.

Our analysis in the jams, jellies, and spreads product group uses data from the jams, jellies, preserves, and peanut butter modules. These four modules constitute 84% of total demand in the product group. Aggregating up to the brand level results in 18 unique products that are comprised of 224 UPCs and account for 85% of demand across the four product modules. Moreover, the distinct differences in and complementary relationship between peanut butter and jellies/preserves create a strong prior for how to categorize products. By comparing the separable demand groups we find in the data to the product module labels, we can examine the face validity of our methodology.

The second analysis from the juice drinks product group uses data from the the top eight product modules which constitute 90% of total demand for the entire juice product group. Aggregating up to the brand level results in 32 products that are comprised of 421 UPCs and account for 64% of total demand across the eight product modules. The relatively lower share of volume is due to the high fraction of small “on-the-go” package sizes, which we exclude in our analysis. Conditional on the set of package sizes in our sample (40-140 oz.), these 421 UPCs account for 83% of demand.

In both data sets, we include seasonal effects, display and feature advertising variables, and product intercepts as product-specific covariates. We also construct M_t using expenditure data at the product group level, which spans a broader set of product modules and UPCs than those used to form the items under study. Detailed descriptions of the products from both data sets are provided in Appendix C.

5.2 Prior and MCMC Settings

Until now, we have treated expenditure elasticities γ as model parameters. While inference for γ is possible, we choose to fix $\gamma_i = 1$ for all $i = 1, \dots, n$ as suggested by the adding-up condition. We feel this restriction is justified for two reasons. First, it amounts to placing a very dogmatic prior on γ that is centered around values suggested by economic theory. Doing so helps push estimates of remaining model parameters towards values consistent with

integrable demand. Second, specifying demand as a function of total expenditure raises a concern of endogeneity when estimating γ . The inclusion of expenditure in (4) is important to maintain the coherence of the restriction in (3). However, our focus is on understanding the structure of cross-price effects, so a formal treatment of expenditure endogeneity (e.g., by way of instrumental variables) only adds unnecessary complication to our analysis.

We proceed by estimating three demand models. The first is an unrestricted log-linear model. The second is a separable log-linear model where the partitioning of goods is fixed a priori based on the definition of product modules. The third is a separable log-linear model where π_n is estimated from the data. All models are estimated using MCMC methods discussed in Section 4. Each Markov chain is run for $R = 500,000$ iterations and thinned by keeping every 100th raw. We use the first 25% of draws as burn-in. When π_n is estimated, we choose the scale parameter of the proposal distribution to be $v = 1/(n \log n)$.

Conditionally conjugate but fairly diffuse priors are placed on all parameters in the unrestricted model: $\psi \sim N(0, 100I)$, $\beta \sim N(0, 10I)$, and $\Sigma \sim IW(n + 3, (n + 3)I)$. In both separable demand models, we place a $N(0, 10I)$ prior on the unrestricted elasticities η , a $N(0, 100I)$ prior on the separability parameters θ , and the same priors on ψ and Σ . When π_n is estimated, we use the $LSP(\rho_n, \tau)$ as a prior with ρ_n equal to the partition with one group and $\tau = 1/(n \log n)$. Doing so imposes a slight penalty on models with many groups.

5.3 Model Comparison

Table 1 provides posterior estimates of in-sample and predictive fit statistics. We report both the log marginal density (LMD) calculated using the Newton-Raftery approximation as well as the root mean squared error (RMSE). The in-sample fit across models is generally similar, with the unrestricted model fitting best in the jams, jellies, and spreads data and the more flexible separable model fitting best in the juice data. The advantages offered by separability are most noticeable in the predictive fit statistics. Specifically, reducing the dimension of the parameter space improves the accuracy and precision of demand forecasts.

Table 1: Model Fit Statistics

Model	In-Sample		Predictive	
	LMD	RMSE	LMD	RMSE
<i>Jams, Jellies, and Spreads</i>				
Unrestricted	-2494.675	0.786 (0.007)	-1205.749	1.047 (0.023)
Separable - Fixed π_n	-2477.253	0.790 (0.005)	-1092.224	0.974 (0.015)
Separable - Estimated π_n	-2471.433	0.789 (0.004)	-1082.494	0.960 (0.014)
<i>Juice Drinks</i>				
Unrestricted	-3126.433	0.420 (0.005)	-1237.058	0.492 (0.013)
Separable - Fixed π_n	-2938.950	0.418 (0.003)	-1016.541	0.422 (0.007)
Separable - Estimated π_n	-2867.157	0.398 (0.004)	-1009.781	0.409 (0.007)

Note: Posterior standard deviations of the RMSE statistic are given in parentheses.

The extent of the improvement between the two separable models depends on the product category and how closely the module labels match the demand structure in the data. For example, estimating π_n leads to larger improvements in predictive fit when the posterior places mass on product groups inconsistent with category labels. In the next section, we show that module labels do a reasonable job describing the separable structure of demand for jams, jellies, and spreads, but not juice drinks.

We can also test for differences between weak, strong, and additive separability using the nested structure of the elasticity restrictions as discussed in Section 2.1. For example, additive separability requires π_n to have n groups so we can test for additivity by computing the posterior probability that $K = n$ in the more flexible separable model. In both data sets, the posterior distribution of K places no mass on $K = n$, rejecting additivity.

Strong separability holds whenever the separability parameters are constant across groups. We can then test for strong separability by testing for equality among the $\theta_{k\ell}$ parameters. We carry out this test using the separable model that estimates π_n because it fits the data best.¹¹ The only complication is that when π_n is estimated, the dimension of the θ vector potentially changes across iterations. We therefore isolate the partitions visited more than once in the Markov chain and compute 95% credible intervals for each $\theta_{k\ell}$ in the associated

¹¹We also computed $1 - \alpha$ credible intervals for each $\theta_{k\ell}$ in the separable model with a fixed partition π_n . We find evidence of non-overlapping credible intervals at the 10% level in juice and the 20% level in jams, jellies, and spreads.

θ vector. We test for equality among the $\theta_{k\ell}$'s by then checking to see if any of the credible intervals do not overlap, which is the case 99% of the time in both data sets. Thus, the magnitude of the separability parameters provides empirical evidence that supports weak rather than strong separability.

Finally, we use the posterior distribution of cross-price effects β_{ij} to examine the extent to which the integrability conditions are satisfied in the data. Adding-up is satisfied by assumption. The posterior distributions for the homogeneity and negativity conditions are shown in the online appendix. We find that both hypotheses cannot be rejected for nearly all products in the data. We also compute the posterior distribution of the symmetry condition and find that it cannot be rejected 97% of the time in the jams, jellies, and spreads data and 98% of the time in the juice data.

5.4 Inference for Partitions

The posterior distribution of π_n is summarized in [Figure 3](#), which plots the posterior pairwise similarity matrix from each data set. The similarity matrix captures the proportion of times two items are grouped together throughout all post-burn-in iterations of the Markov chain, where darker colors imply higher clustering probabilities. We also show the grouping structure implied by product modules in the bottom right corner of each similarity matrix to be able to compare how the partitions supported by the data compare to product categories.

In the jams, jellies, and spreads data set, we find that partitions with high posterior probability are fairly consistent with the definition of product modules. That is, the estimated demand groups tend to separate peanut butter from jellies, jams, and preserves. Our results do indicate some co-clustering across the jellies/jams and preserves modules. However, the majority of clustering occurs within the product module. We also find a slightly more refined structure of demand within the jellies/jams module, in which grape and strawberry offerings are usually separated from blackberry and raspberry offerings. However, the model fit statistics suggest that making the mistake of defining a category too broadly is still less

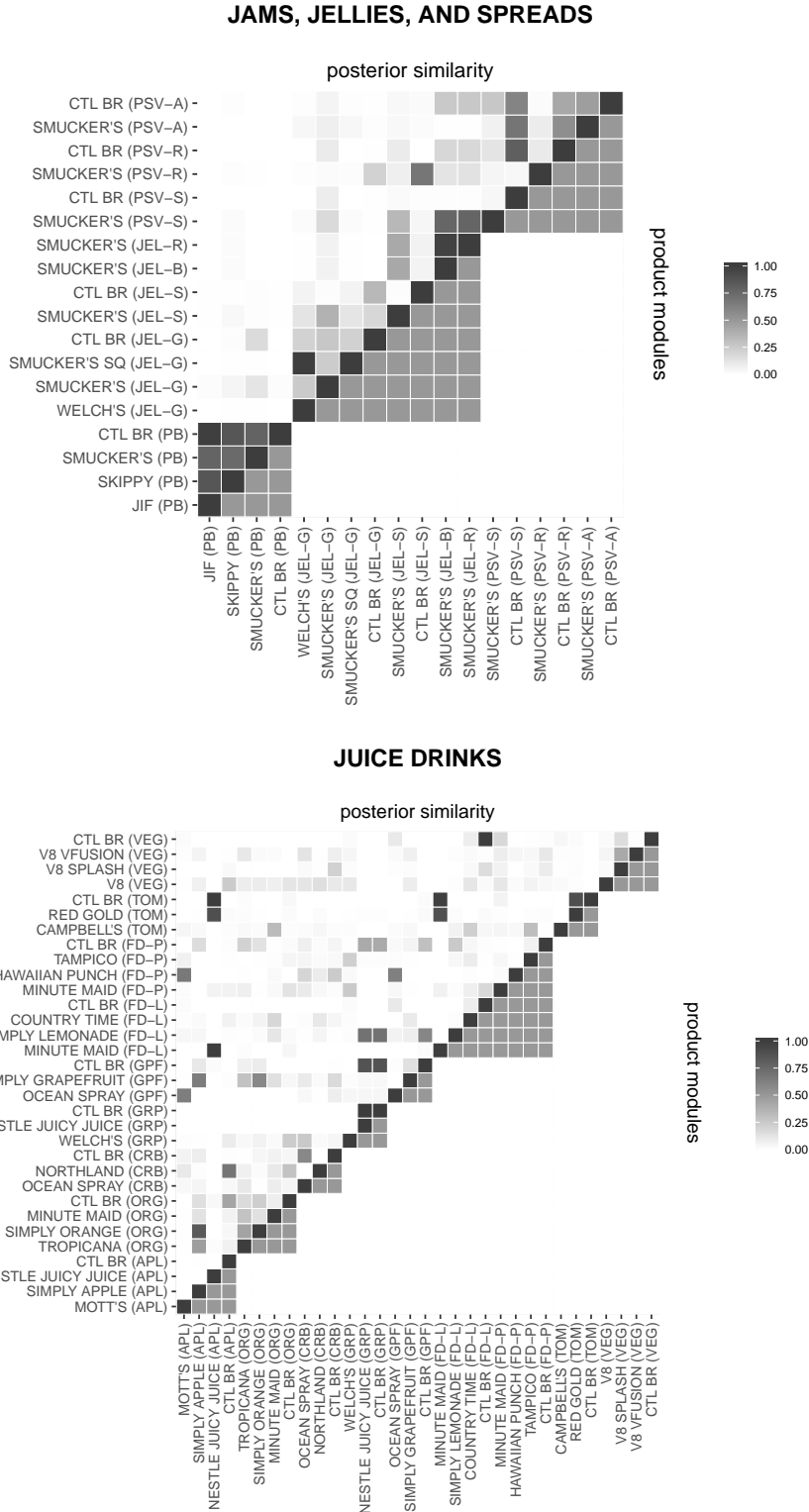


Figure 3: Posterior pairwise similarity matrices are plotted for each data set. The bottom right corner of each plot shows the grouping structure implied by product modules.

egregious than defining categories that are too narrow.

In the juice data set, we find stark differences between the separable structure of demand and the predefined module categorization. In fact, partitions with appreciable posterior mass always contain demand groups that span different flavor subcategories. For example, we find that three Simply juices (apple, orange, grapefruit) exhibit high co-clustering probabilities (83% for apple-orange, 64% for apple-grapefruit, and 58% for orange-grapefruit).

The non-separability of demand for Simply juices may be attributed to marketing activities. For example, Simply products all have identical branding and are displayed together in the refrigerated section of the grocery store.¹² The shelf placement of Simply Apple is especially intriguing given that most all apple juice products are usually shelf-stable and physically located in a different aisle. Activities like these could help Simply, a much younger brand than its competitors, create a local monopoly in the broader juice category.

5.5 Inference for Separability and Elasticity Parameters

The estimates of the separability parameters provide information about the relationship between separable groups. In the jams, jellies, and spreads data set, for example, one may not only expect peanut butter and jelly to be in separable groups, but also that the separability parameters indicate group-wise complementarity. Given that the a priori relationship between different demand groups in the juice category is less obvious, we focus on inference for separability parameters for the jams, jellies, and spreads category.

Marginal inference for separability parameters is challenging when π_n is estimated because the dimension of θ changes with π_n . Since we are interested in whether the estimates of θ generally pick up complementarity between the peanut butter and jellies/preserves product groups, we summarize the posterior of θ as follows. We first condition on states in the posterior such that all peanut butters belong to different groups from the jellies and preserves. Then, we compute the average value of θ both across peanut butters and jellies/preserves

¹²<https://www.simplyorangejuice.com/faq>

and within jellies/preserves. The full distribution of θ in each case is shown in [Figure 4](#).

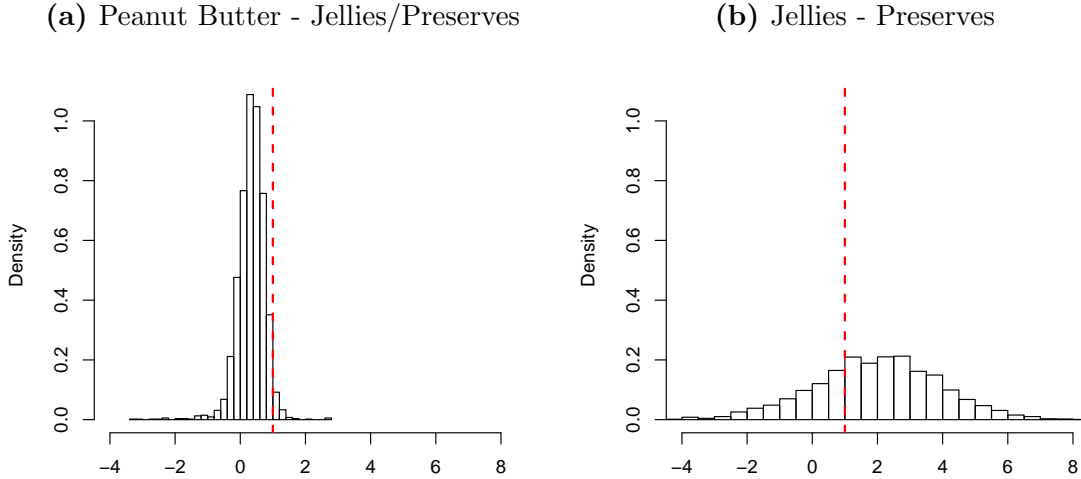
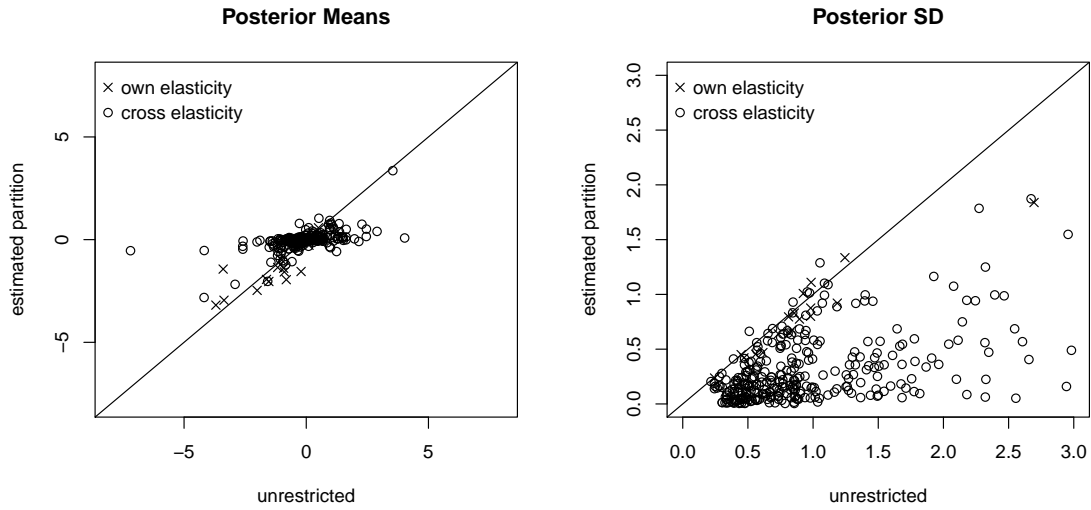


Figure 4: The distribution of the average cross-group separability parameter is plotted for (a) peanut butter and jellies/preserves and (b) jellies and preserves. The dashed vertical line indicates the value of θ for which products are complements ($\theta < 1$) or substitutes ($\theta > 1$).

A priori, one may expect θ to induce complementarity between peanut butter and jellies/preserves, but substitutability between jellies and preserves. We can assess product relationships using (6), which shows that complementarity holds whenever $\theta < 1$. [Figure 4](#) shows that the distribution of θ generally falls below one for peanut butter - jellies/preserves, and above one for jellies-preserves. This provides evidence that when π_n is estimated, the model not only tends to assign peanut butter and jellies/preserves into separable groups, but also produces values of θ that reflect complementarity between peanut butter and jellies/preserves and substitutability between jellies and preserves.

We can also project estimates of the lower-dimensional separability parameters back up to the space of the $n \times n$ price elasticity matrix. [Figure 5](#) 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 with estimated π_n . While there is a high degree of correlation between elasticity estimates, 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

(a) Jams, Jellies, and Spreads



(b) Juice Drinks

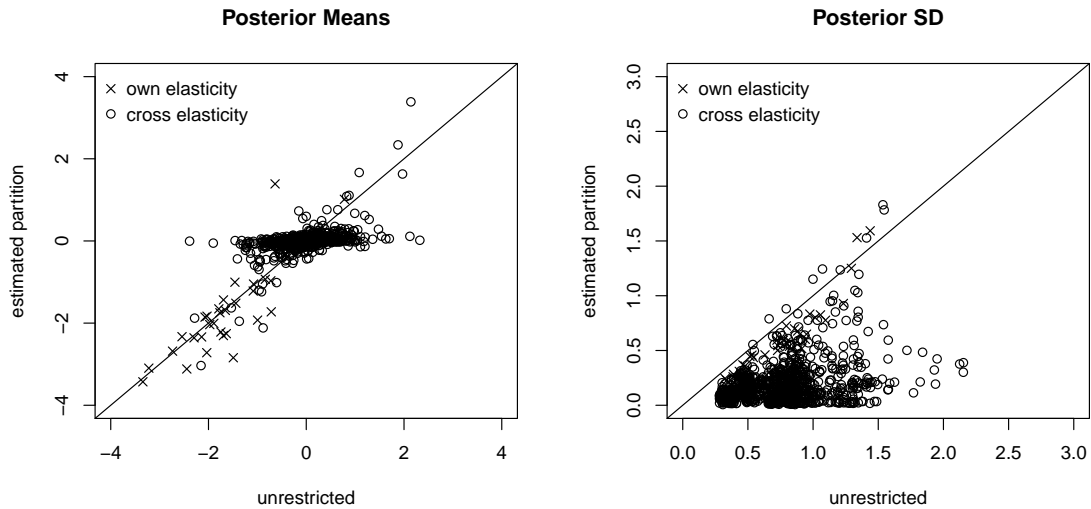


Figure 5: The posterior means and standard deviations of price elasticity estimates are plotted for the separable and unrestricted models across both data sets.

is an advantage over other regularization methods that only impose shrinkage to zero.

Moreover, the precision of the estimates tends to be worse for both unrestricted models. This is shown in the right column of Figure 5, 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.

6 Implications

This section explores the implications of separable demand on product competition, brand positioning, and the measurement category promotion effects. We investigate these topics using results from the juice data set, where the variety of flavors and brands generally complicates the analysis of market structure.

6.1 Competition

An advantage of economic demand models is that they allow for a formal analysis of product competition as measured by the sign and magnitude of cross-price effects. For instance, brands with higher cross-price effects compete more intensely and are therefore closer substitutes. Separability offers a third dimension for assessing competition: the *dimension* of the price elasticity matrix. By allowing π_n to be estimated from the data, we can learn about the boundaries of consumer preferences as defined by (2).

Empirically, we find brands like Simply in the juice category appear to break the notion of separability by flavor. This means that the rate at which consumers substitute between brands within a flavor category (e.g., apple juice) may very well depend on their consumption in other flavor categories (e.g., orange juice). In a sense, this finding corroborates the goals of many marketing activities like umbrella branding and advertising, which seek to create information spill-overs in the consumer’s choice process. This result also implies that the boundaries of competition may be broader than previously thought, at least within the juice category. Allowing for a formal connection between marketing variables (e.g., shelf-collocation or advertising) to the separable structure of demand is something we leave for future work.

In addition to the boundaries of competition, we can also explore the induced effects of separability restrictions on the sign and magnitude of cross-price effects. In a separable demand model, cross-group price effects are a function of budget shares \bar{w}_j , separability

parameters θ , and expenditure elasticities γ . Since we fix expenditure elasticities a priori, the sign and magnitude of these cross-group price effects will be driven by θ . In many cases, θ will need to be large in absolute value in order to overcome small budget shares and produce non-zero cross-effects.

We start by plotting the distribution of within-group and across-group price elasticities for both separable models (Figure 6). We find that imposing separability generally shrinks the cross-group price effects to a small neighborhood around zero, which suggests that products within a group are most influenced by price changes within the group than across groups. This is consistent with previous work on the effects of separability on competition (e.g., Rubinfeld, 2000), which have treated this property as a shortcoming of separability since the form of competition is in part fixed by the researcher’s partitioning of goods. However, this is not a concern when π_n is estimated from the data. In fact, we need the cross-group restrictions to be dogmatic in order to separate true cross-effects from random noise.

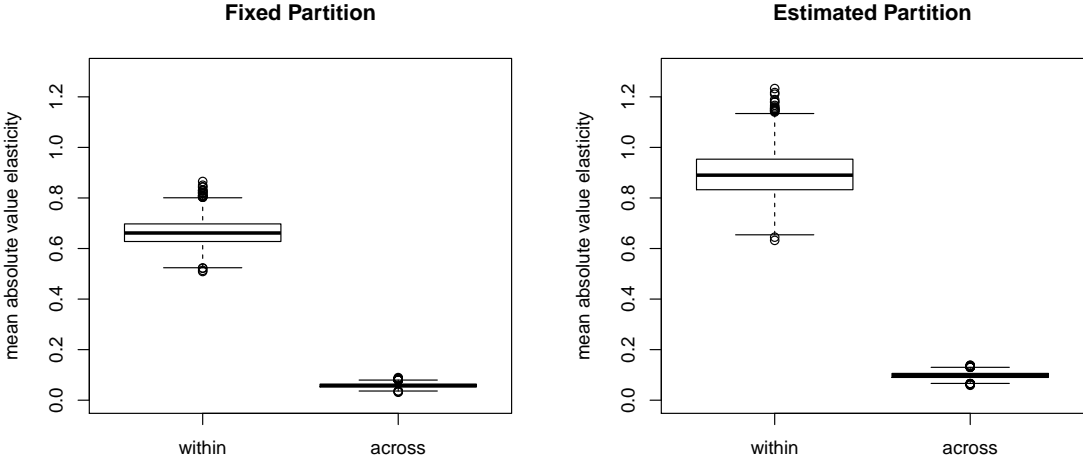


Figure 6: The distributions of the mean absolute value within-group elasticities and across-group elasticities are plotted for both separable demand models in the juice category.

To see the impact of separability at a more granular level, we also decompose the within-group and across-group elasticities by flavor category. Table 2 reports the average own elasticity, within-group elasticity, across-group elasticity, and absolute value of the ratio of

within to across for each product module. The ratio illustrates the extent to which the magnitude of within-group elasticities exceeds that of the cross-group elasticities.

Table 2: Average Price Elasticities by Product Module

Module	Fixed π_n				Estimated π_n			
	Own	Within	Across	Ratio	Own	Within	Across	Ratio
Apple	-0.773	0.079	-0.027	9.448	-0.439	0.012	-0.044	1.147
Orange	-2.010	0.185	-0.022	9.799	-2.085	0.062	0.016	3.816
Cranberry	-2.146	0.321	-0.007	82.190	-2.021	0.303	0.008	22.014
Grape	-1.458	0.024	-0.045	13.858	-1.498	-0.416	-0.039	12.238
Grapefruit	-1.368	0.012	0.010	38.589	-0.798	-0.039	-0.064	5.518
Fruit Drinks	-1.853	-0.080	0.012	12.794	-2.064	-0.004	0.015	8.592
Tomato	-1.833	0.281	-0.014	19.803	-2.029	0.174	-0.012	20.758
Other Vegetable	-2.087	0.189	-0.006	38.455	-2.097	0.025	-0.006	7.533

We find that when the partition is fixed, the average ratio of within to across-group elasticities tends to be extremely large. However, the ratios shrink significantly when π_n is estimated. Even though each partition still places strong restrictions on the cross-group effects, the reported elasticities are averaged over the posterior distribution of π_n . This induced differential shrinkage allows the separable model to admit more flexible substitution patterns, obviating the concern that separable assumptions fix the form of product competition.

6.2 Brand Positioning

Given the relationship between the partitioning of goods into separable groups and the induced patterns of competition, we next investigate how this information can influence brand positioning strategies. To illustrate, again consider the case of Simply juices. Our empirical analysis showed that Simply juices often belong to one group rather than separable groups. The retailer may then wonder whether this local concentration of demand for Simply hurts its ability to compete with its store brand juice offerings.

If brand positioning is a determinant of the separable structure of demand, then the retailer could consider repositioning the line of store brand juices in a way that moves store brand juices into the cluster of Simply juices. We explore this type of counterfactual with a

simple example with three products: Simply Apple, Simply Orange, and store brand orange juice. If Simply Apple and Simply Orange belong to one group and the store orange juice belongs to another, then Simply Orange and store brand orange will likely be less competitive (as measured by the magnitude of their cross price effect).

By exploiting the uncertainty in the posterior distribution of π_n , we can offer insights relating to this counterfactual. That is, in the posterior of π_n we observe Simply Orange and Simply Apple belonging to the same cluster 83% of the time, while we observe Simply Orange and store brand orange belonging to the same cluster 23% of the time. We then look at cases when Simply Orange and store brand orange are either grouped together or not, and compute the average cross-price elasticity between (i) Simply Orange and store brand orange and (ii) Simply Apple and store brand orange. The distributions of the the mean absolute value cross-price elasticities are shown for each case in [Figure 7](#).

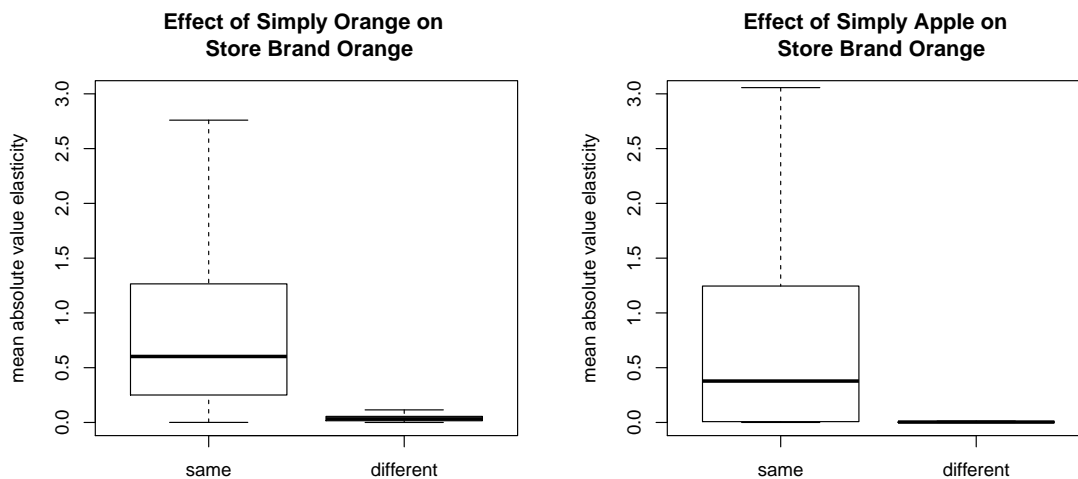


Figure 7: Cross-price elasticities between (i) Simply Orange and store brand orange and (ii) Simply Apple and store brand orange are plotted. “Same” indicates Simply Orange and store brand orange belong to the same group while “different” indicates they belong to different groups.

We find that the Simply-store brand cross-price effects are roughly zero when the two belong to different groups, but non-zero when they belong to the same group. This is even

true for the effect of a price change on Simply Apple on the demand for store brand orange juice. The intuition is that joining the same cluster as Simply has the potential to generate positive cross-category spillover effects for the store brand.

6.3 Promotion Effects and Category Revenue

Category managers are often tasked with setting pricing and promotional schedules across a wide assortment of goods. In these settings, it becomes especially important to estimate category demand systems that yield flexible, yet precise demand elasticities in order to accurately measure the impact of price promotions on category sales. To illustrate the implications of separable (or non-separable) demand on marketing actions, we consider the following promotional forecasting problem: What is the effect of a price promotion of one product on total category sales? In general, the answer is ambiguous as it depends on the relationship between a product’s own and cross-price effects. It is then important to measure these price effects accurately and precisely, both within and across product modules.

Let $\mathcal{R}_{\mathcal{J}}(\Omega^m|X) = \sum_{j \in \mathcal{J}} p_j \cdot q_j(\Omega^m|X)$ denote the total revenue from product set \mathcal{J} predicted by model m with unobservables Ω^m and covariates X . The demand equation for each model takes the general form:

$$\begin{aligned} \log q_i &= h^m(X, \omega_i^m) + \varepsilon_i \\ \varepsilon &= (\varepsilon_1, \dots, \varepsilon_n) \sim N(0, \Sigma) \end{aligned} \tag{28}$$

where $h^m(X, \omega_i^m)$ denotes the conditional mean function for model m which depends on covariates X and parameters ω_i^m . We let Ω^m denote the full collection of model parameters. For example, $\Omega^m = (\beta, \gamma, \psi, \Sigma)$ in the unrestricted model while $\Omega^m = (\beta_{\pi_n}, \pi_n, \gamma, \psi, \Sigma)$ in the separable model when estimating π_n . When the error vector is multivariate normal, the expected demand for good j (integrating over ε) has a simple closed-form expression:

$$q_j(\Omega^m|X) = E^\varepsilon \left[q_j(\Omega^m, \varepsilon|X) \right] = \exp \left(h^m(X, \omega_j^m) + 0.5 \Sigma_{jj} \right). \tag{29}$$

Expected revenue can then be written as an integral of $\mathcal{R}_{\mathcal{J}}(\Omega^m|X)$ over the posterior distribution of all model parameters Ω^m , corresponding to the posterior predictive distribution of Ω^m .

$$\begin{aligned} E^{\Omega^m}[\mathcal{R}_{\mathcal{J}}(\Omega^m|X)] &= E^{\Omega^m} \left[\sum_{j \in \mathcal{J}} p_j \cdot q_j(\Omega^m|X) \right] \\ &= \int \sum_{j \in \mathcal{J}} p_j \cdot \exp(h^m(X, \omega_j^m) + 0.5\Sigma_{jj}) dF(\Omega^m|X) \end{aligned} \tag{30}$$

We first provide an illustrative example in which we predict the effects of a \$0.50-off price promotion for Simply Orange on category revenue. The depth of this promotion reflects a move from the 75th percentile of the Simply Orange price distribution to the 25th percentile of the distribution. Choosing Simply Orange also allows us to demonstrate the consequences of misspecifying the separable structure of demand since we find evidence that Simply Orange may not be separable from other Simply offerings. For each of the separable models, we compute expected demand under two price sets: (i) prices of all products are fixed at their average price; and (ii) prices of all products are fixed at their average price except for Simply Orange which is \$0.50 less than its average price.

When π_n is fixed, the model predicts a null effect on category revenue, with the 95% credible interval ranging from -3.6% to 3.3%. However, when π_n is estimated, the model predicts a 2.8% increase on category revenue, with the 95% credible interval ranging from -2.4% to 11.2%. The difference is likely due to the misspecified shrinkage structure of cross-price effects. For instance, the non-separability of demand coming from a brand like Simply may generate a form of brand complementarity. In this case, a price promotion of Simply in one flavor subcategory may lead to increased demand for Simply in other subcategories. However, this effect cannot be captured when π_n is fixed and misspecified.

To empirically validate the previous example and demonstrate that biases exist beyond Simply Orange alone, we use price variation in our hold-out sample to predict category sales. Since price promotions are not labeled in the data, we must first construct a set of base and promotion prices for each product. We define the base price as any price in the top 25th

percentile of the product’s price distribution in a given year, and the promotion price as any price in the bottom 25th percentile of the same distribution.¹³ We then use the unrestricted and separable models to predict demand under each product’s base and promotion price set, compute the associated percent change in category revenue, and then compare with the observed percent change in sales in our hold-out sample.

For each product $i = 1, \dots, n$, we use (30) to compute the percent change in expected category revenue generated by a price promotion of product i . This generates estimates of the category promotion effect for each good, which we then compare to the actual percent change in category revenue observed in the hold-out sample. Table 3 reports the average bias in absolute value and standard deviation of these estimates for each model, where the averaging is done across products.

Table 3: Effect of Promotion on Category Revenue

Model	Bias	SD
Unrestricted	0.051	0.055
Separable - Fixed π_n	0.058	0.026
Separable - Estimated π_n	0.051	0.026

Estimates from the unrestricted model exhibit less bias but larger variance than those from the separable model with fixed π_n . The larger variance is an artifact of the number of parameters in the unrestricted model. When π_n is estimated, however, the bias relative to the unrestricted model disappears and we still retain the relative gains in efficiency.

7 Discussion

This paper develops a methodology that can be used to learn about the separable structure of demand from store-level scanner data. Traditional work on separable demand has treated the partitioning of products as fixed. Instead, we let the partition be a model parameter that can be estimated and propose a Bayesian method of joint inference for the partition

¹³Other definitions of promotions, such as a percent reduction from the base price, did not induce any material changes to the results.

and other demand parameters. An immediate benefit is that we are able to use the data to find potentially less obvious grouping structures.

Using data from two distinct product categories, we find that the separable structure of demand can be markedly different from predefined category labels. For example, in the juice category, we find that multiple flavors are often present within demand groups, suggesting that the demand is not separable by flavor. This result also suggests that the boundary of competition between subcategories like apple and orange juice may extend beyond that which is defined by subcategory labels. We then show how misspecifying these boundaries impact inference for product competition and promotion effects.

There are limitations of the current work which serve as a motivation for possible extensions. For example, our treatment of aggregate demand ignores many important aspects of consumer choice, such as store choice, category choice, purchase timing, and household inventories. Ideally, we would derive aggregate demand from a micro-model of utility maximization to account for these factors. However, this becomes challenging when the underlying utility deviates from an additively separable structure. We have also used a store-level demand system which does not attempt to formally model the distribution of consumer preferences. If consumers are heterogeneous in the partitions which define their separable preferences, then some of the non-separabilities we find at the store level could be due to this source of heterogeneity. However, it is well known that consumer heterogeneity is only weakly identified with aggregate data. It remains to be investigated whether heterogeneity in separability-based groupings of products can be identified using even relatively disaggregate store-level data.

Our work may also have implications for formal retailer pricing policies. Traditionally, retailers are assumed to set prices by maximizing category profits, where categories are defined at a more disaggregate level, such as orange juice (e.g., [Montgomery, 1997](#); [Dubé et al., 2008](#)). However, our results suggest demand may not be separable across categories like apple and orange juice, especially in the presence of strong multi-product national brands.

Formalizing the consequences on coordinated pricing with non-separable category demand remains an open area for future work.

Finally, we have shown that both the structure of the log-linear model and the sheer dimension of the partition space both limit the scalability of our model. At the same time, very high-dimensional product sets are becoming more common in empirical work, so it would be useful to see how the core dimension reduction offered by separability can be used when the product space is extremely large. To the extent that separability can formally connect flexible and scalable statistical models with economic theory, this extension will especially relevant when firms want to estimate demand parameters for the purpose of optimizing marketing actions. We leave these topics for future research.

Appendix

A Separability Restrictions for Price Elasticities

Let $S_{ij} = \frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial m}$ denote the effect of a unit increase in the price of good j on the compensated (Hicksian) demand of good i . Weak separability imposes a restriction on cross-group compensated price effects. That is, if i and j are in different groups, then

$$S_{ij} = \phi_{k\ell}(q) \frac{\partial q_i}{\partial m} \frac{\partial q_j}{\partial m}. \quad (\text{A.1})$$

Since log-linear demand models are parameterized in terms of uncompensated price elasticities (i.e., elasticities with respect to Marshallian demand functions), we need to translate the restriction on S_{ij} into a restriction on β_{ij} . We first rewrite the original separability restriction in terms of expenditure elasticities.

$$S_{ij} = \phi_{k\ell}(q) \frac{\partial q_i}{\partial m} \frac{\partial q_j}{\partial m} \quad (\text{A.2})$$

$$= \frac{\phi_{k\ell}(q)}{m} \cdot \frac{\partial q_i/q_i}{\partial m/m} \cdot \frac{\partial q_j/q_j}{\partial m/m} \cdot \frac{q_i q_j}{m} \quad (\text{A.3})$$

$$= \frac{\phi_{k\ell}(q)}{m} \gamma_i \gamma_j \frac{q_i q_j}{m} \quad (\text{A.4})$$

$$= \theta_{k\ell}(q) \gamma_i \gamma_j \frac{q_i q_j}{m} \quad (\text{A.5})$$

We make the further assumption that the function $\theta_{k\ell}(q)$ only contains an intercept term and is thus independent of demand: $\theta_{k\ell}(q) = \theta_{k\ell}$. Then by the elasticity form of the Slutsky equation, it follows that

$$S_{ij} \cdot \frac{p_j}{q_i} = \beta_{ij} + \gamma_i w_j \implies \theta_{k\ell} \gamma_i \gamma_j \frac{q_i q_j}{m} \cdot \frac{p_j}{q_i} = \beta_{ij} + \gamma_i w_j \quad (\text{A.6})$$

$$\implies \theta_{k\ell} \gamma_i \gamma_j \frac{q_j p_j}{m} = \beta_{ij} + \gamma_i w_j \quad (\text{A.7})$$

$$\implies \beta_{ij} = w_j (\theta_{k\ell} \gamma_i \gamma_j - \gamma_i). \quad (\text{A.8})$$

B Details for the Conditional Posterior of β_{π_n}

To derive the posterior distribution of β_{π_n} conditional on all other parameters and data, we first isolate β_{π_n} and rewrite the model in (20) as

$$y^* = X(\pi_n, \gamma; P, w)\beta_{\pi_n} + \varepsilon \quad (\text{B.1})$$

where $y^* = y - c(\pi_n, \gamma; P, w) - M\gamma - Z\psi$ and $\varepsilon \sim \text{N}(0, \Sigma \otimes I_T)$. Although $X(\cdot)$ is not block diagonal as in a standard SUR model, the likelihood function still exhibits a SUR-like representation due to the vectorized response, expanded mean function, and correlated errors.

Next, we translate this system into a system of uncorrelated errors:

$$\tilde{y}^* = \tilde{X}_{\pi_n}\beta_{\pi_n} + \tilde{\varepsilon} \quad (\text{B.2})$$

where

$$\Sigma = U'U \quad (\text{B.3})$$

$$\tilde{y} = ((U^{-1})' \otimes I_T)y \quad (\text{B.4})$$

$$\tilde{X}_{\pi_n} = ((U^{-1})' \otimes I_T)X(\pi_n, \gamma; P, w) \quad (\text{B.5})$$

$$\tilde{\varepsilon} = ((U^{-1})' \otimes I_T)\varepsilon \quad (\text{B.6})$$

so that $\text{Var}(\tilde{\varepsilon}) = ((U^{-1})' \otimes I_T)(\Sigma \otimes I_T)(U^{-1} \otimes I_T) = (U^{-1})'\Sigma U^{-1} \otimes I_T = I_n \otimes I_T$. With a conditionally conjugate prior $\beta_{\pi_n} | \pi_n \sim \text{N}(\bar{\beta}_{\pi_n}, A_{\beta}^{-1})$, the associated posterior takes the form

$$\beta_{\pi_n} | \pi_n, \text{else} \sim \text{N}(\tilde{\beta}_{\pi_n}, (\tilde{X}_{\pi_n}'\tilde{X}_{\pi_n} + A_{\beta})^{-1}) \quad (\text{B.7})$$

where $\tilde{\beta}_{\pi_n} = (\tilde{X}_{\pi_n}'\tilde{X}_{\pi_n} + A_{\beta})^{-1}(\tilde{X}_{\pi_n}'\tilde{y}^* + A_{\beta}\bar{\beta}_{\pi_n})$.

C Product Descriptions

Jams, Jellies, and Spreads

	Brand	Module	Submodule
1	JIF	Peanut Butter	
2	SKIPPY	Peanut Butter	
3	SMUCKER'S	Peanut Butter	
4	CTL BR	Peanut Butter	
5	WELCH'S	Jelly	Grape
6	SMUCKER'S	Jelly	Grape
7	SMUCKER'S SQUEEZE	Jelly	Grape
8	CTL BR	Jelly	Grape
9	SMUCKER'S	Jelly	Strawberry
10	CTL BR	Jelly	Strawberry
11	SMUCKER'S	Jelly	Blackberry
12	SMUCKER'S	Jelly	Raspberry
13	SMUCKER'S	Preserves	Strawberry
14	CTL BR	Preserves	Strawberry
15	SMUCKER'S	Preserves	Raspberry
16	CTL BR	Preserves	Raspberry
17	SMUCKER'S	Preserves	Apricot
18	CTL BR	Preserves	Apricot

Juice Drinks

	Brand	Module
1	MOTT'S	Apple
2	SIMPLY APPLE	Apple
3	NESTLE JUICY JUICE	Apple
4	CTL BR	Apple
5	TROPICANA	Orange
6	SIMPLY ORANGE	Orange
7	MINUTE MAID	Orange
8	CTL BR	Orange
9	OCEAN SPRAY	Cranberry
10	NORTHLAND	Cranberry
11	CTL BR	Cranberry
12	WELCH'S	Grape
13	NESTLE JUICY JUICE	Grape
14	CTL BR	Grape
15	OCEAN SPRAY	Grapefruit
16	SIMPLY GRAPEFRUIT	Grapefruit
17	CTL BR	Grapefruit
18	MINUTE MAID	Fruit Drinks
19	SIMPLY LEMONADE	Fruit Drinks
20	COUNTRY TIME	Fruit Drinks
21	CTL BR	Fruit Drinks
22	MINUTE MAID	Fruit Drinks
23	HAWAIIAN PUNCH	Fruit Drinks
24	TAMPICO	Fruit Drinks
25	CTL BR	Fruit Drinks
26	CAMPBELL'S	Tomato
27	RED GOLD	Tomato
28	CTL BR	Tomato
29	V8	Other Vegetable
30	V8 SPLASH	Other Vegetable
31	V8 VFUSION	Other Vegetable
32	CTL BR	Other Vegetable

ONLINE APPENDIX

A LSP vs. DP Distributions

In this section, we compare the LSP distribution to the partition distribution induced by the Dirichlet process (DP). Since the LSP distribution is parameterized by a location partition ρ_n and a dispersion parameter $\tau > 0$, we fix $n = 100$ and set ρ_n to be a partition of $K = 5$ contiguous groups with 20 elements in each. We then consider a range of values of the scale parameter: $\tau \in \{0.1, 1, 10\}$. For the DP partitioning model, we again consider different values of the concentration parameter: $\alpha \in \{0.1, 1, 10\}$. The choice of G_0 is immaterial for the purposes of studying the induced partition distribution. To show how each model's parameters affect the resulting partition distribution, we generate 10,000 draws from each distribution and plot the associated pairwise similarity matrices in [Figure A.1](#).

As suggested by the location-scale property of the LSP distribution, more mass is placed on the location partition ρ_n as τ tends towards zero (top left). However, as τ gets large, the probability mass gets spread more evenly across \mathcal{P}_n (top right). In contrast, the distribution induced by the DP has no way of being centered around a particular partition. Under this model, the probability that any two items are grouped together is therefore constant (bottom panel). The values of α only uniformly shift these probabilities for all items (i, j) .

By changing the location partition to contain only one group, we can also examine how accurately the LSP model can approximate the DP model. [Figure A.2](#) replicates the LSP pairwise similarity matrices from the top panel of [Figure A.1](#), except we now let ρ_n contain one group and divide the previous values of τ by 20. We find that the resulting pairwise similarity matrices closely match those of the DP model in the bottom panel of [Figure A.1](#).

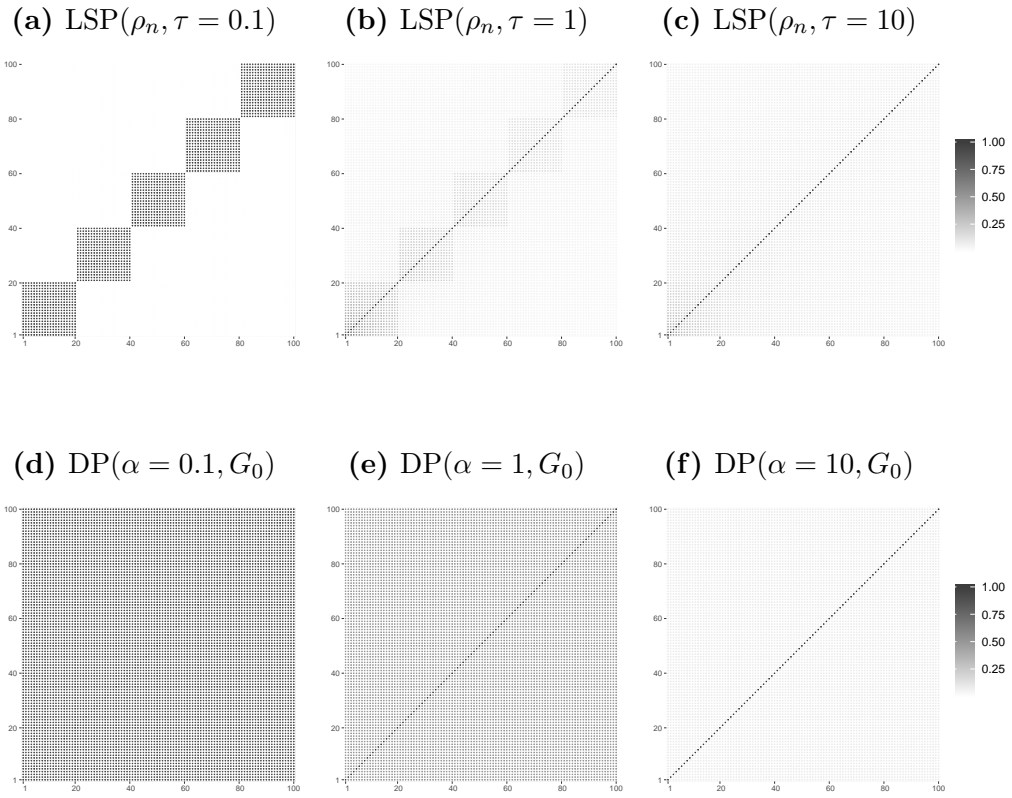


Figure A.1: Pairwise similarity matrices from various LSP and DP distributions. The location partition of the LSP distribution contains five contiguous groups.

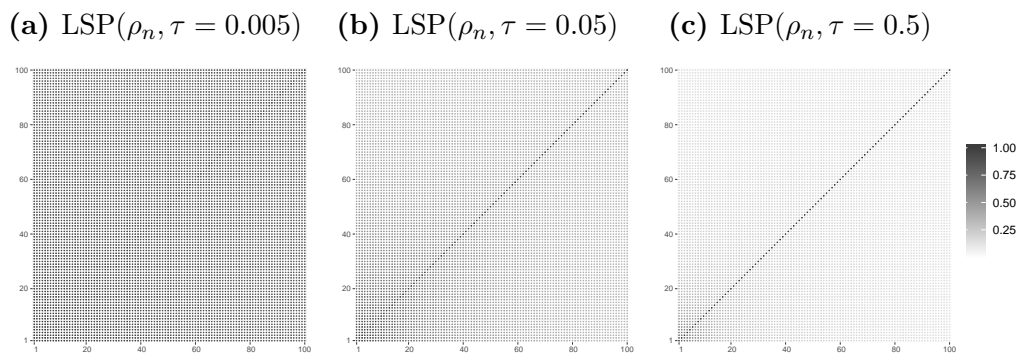


Figure A.2: Pairwise similarity matrices from an LSP distribution where ρ_n contains one group and $\tau \in \{0.005, 0.05, 0.5\}$.

B Simulation Studies

B.1 Prior Sensitivity Analysis

Data are generated from the separable demand model in Section 2. Let $n = 20$ be the number of products and $T = 100$ the number of time periods. The matrix of log prices is generated from a $\text{Unif}(0, 0.5)$ distribution, the log expenditure variables are generated from a $\text{Unif}(0, 1)$ distribution, and the expenditure shares are generated from a $\text{Dirichlet}(1, \dots, 1)$ distribution. The relatively few time periods and little variation in prices are chosen to show the impact of the prior on elasticity and partition parameters. For ease of interpretation, we let π_n be a partition of $K = 4$ contiguous groups with each group having 5 products.

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

Then conditional on π_n , we generate the elements of $\beta_{\pi_n} = (\eta, \theta)$: $\eta_{ii} \sim \text{Unif}(-3, 0)$, $\eta_{ij} | \pi_n \sim \text{Unif}(0, 3)$, and $\theta_{kl} | \pi_n \sim \text{Unif}(-10, 10)$. We also let ψ_{ij} contain an intercept only, fix $\gamma_i = 1$ for all products, and set $\Sigma = 0.1 + I_n$. A heat map of the induced $n \times n$ price elasticity matrix is shown in Figure B.1.

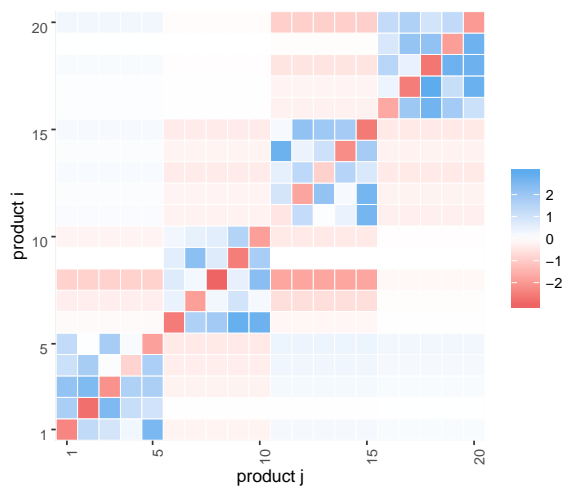


Figure B.1: A heat map of the true price elasticity matrix used to generate the data.

As discussed in Section 4.3, the choice of hyperparameters in the joint prior of β_{π_n} and π_n can have significant bearing on the posterior. We highlight these effects by considering three prior specifications. The first is an informative LSP prior consistent with what is used in our empirical analysis: $\eta|\pi_n \sim N(0, 10I)$, $\theta|\pi_n \sim N(0, 100I)$, and $\pi_n \sim \text{LSP}(\rho_n, 1/(n \log n))$ where ρ_n contains only one group. The second is more diffuse: $\eta|\pi_n \sim N(0, 100I)$, $\theta|\pi_n \sim N(0, 100I)$, and $\pi_n \sim \text{LSP}(\rho_n, n/(n \log n))$ where again ρ_n contains only one group. The third specifies a DP prior on π_n with the concentration parameter equal to $\alpha = n$.

The model is estimated using the MCMC routine outlined in Section 4.2. We demonstrate the flexibility of our approach and treat γ as an estimated parameter with a diffuse normal prior. The Markov chain is run for $R = 50,000$ iterations and the step size of the random-walk proposals for π_n are chosen to be $v = 1/(n \log n)$.

We first consider the output from the model with an informative prior. The trace plot of the partition parameter is shown in Figure B.2. 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 dashed 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 the true partition.

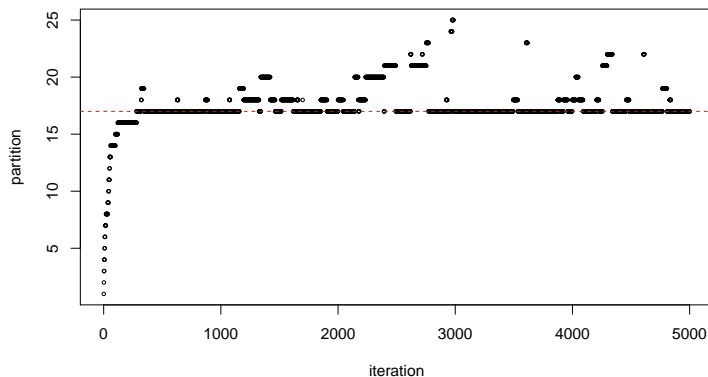


Figure B.2: The trace plot for the partition parameter π_n . The dashed horizontal line represents the true partition used to generate the data.

Figure B.3 plots the posterior distributions over the number of groups K corresponding to both diffuse and informative priors. We find that diffuse priors tend to favor models with many groups, as the posterior mass is concentrated concentrated to the right of $K = 4$ (shown by the dashed line). While this effect can be countered through the choice of a more informative prior on π_n , the effect would also diminish as the number of observations or amount of price variation increases. Going forward, we focus our analysis on the model with the informative LSP prior.

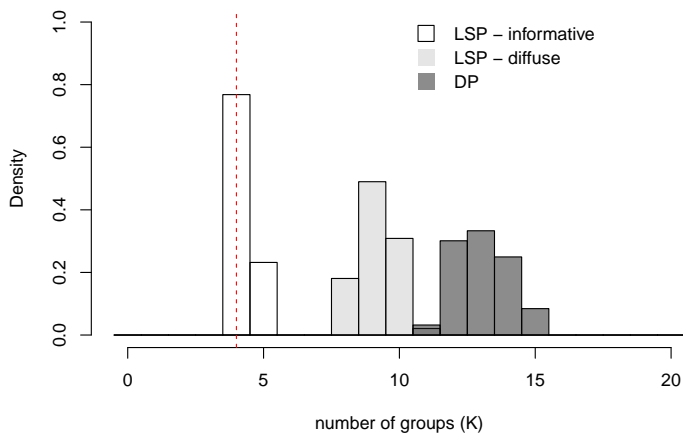
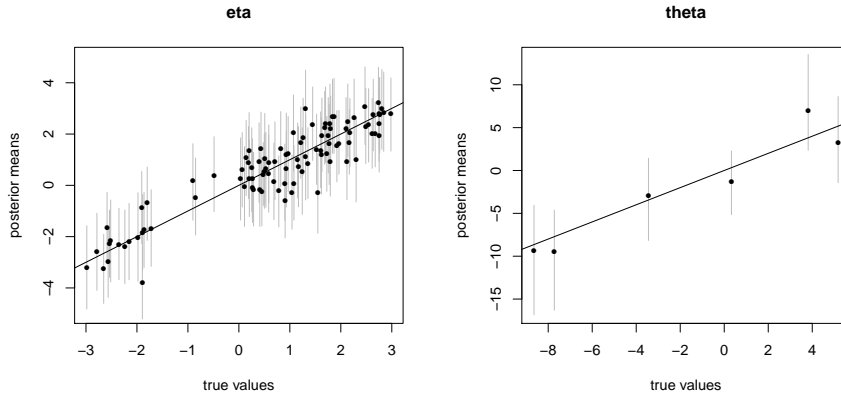


Figure B.3: The posterior distributions of K are plotted for models with diffuse and informative priors. The dashed vertical line represents the true number of groups ($K = 4$).

Figure B.4 plots the posterior estimates of $\beta_{\pi_n} = (\eta, \theta)$ for a model where π_n is estimated and the model where π_n is fixed at its true value. This allows us to both assess the ability of our model to recover parameters, as well as check for any losses in efficiency that arise when π_n is estimated. We find that the associated 95% credible intervals always cover the true parameter values, and there is minimal difference in the length of the credible intervals between the two models. Figure B.5 plots the 95% posterior credible intervals for γ and ψ for the model when π_n is estimated. All credible intervals cover the true values. While Σ is also well recovered, we do not report the estimates for the sake of brevity.

(a) Estimated π_n



(b) Fixed π_n

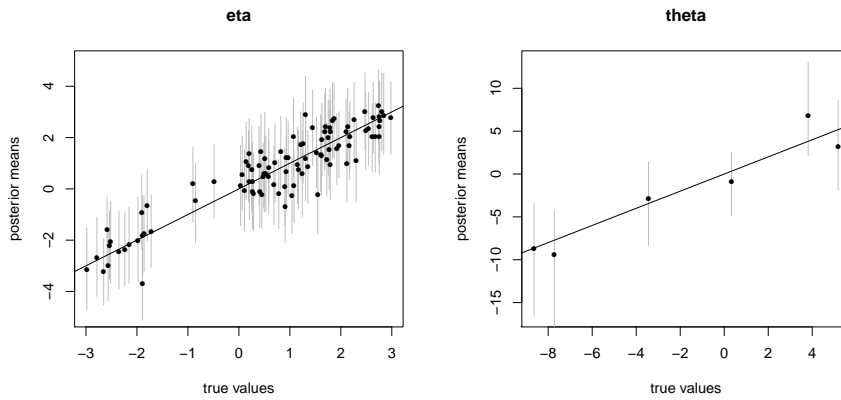


Figure B.4: Conditional posterior means and 95% credible intervals are plotted against the true values of η and θ . The top and bottom panels corresponds to models where π_n is estimated and fixed, respectively. Dots indicate posterior means.

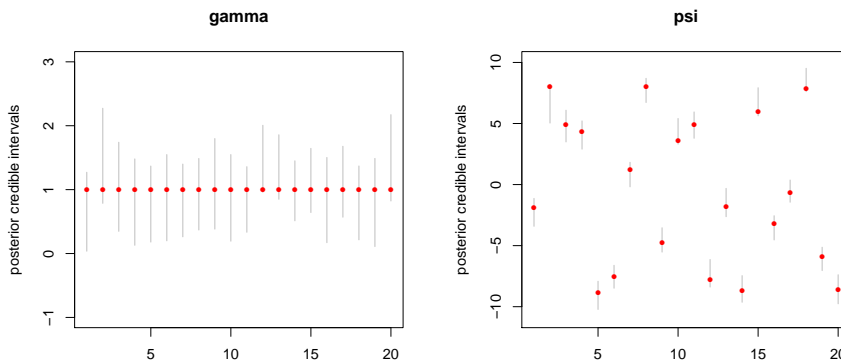


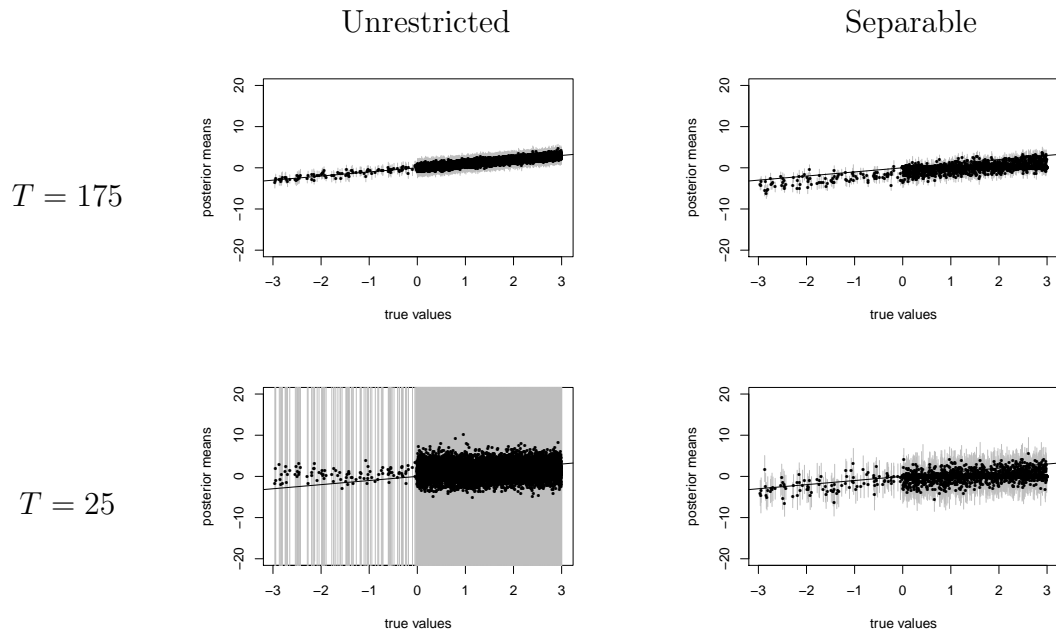
Figure B.5: 95% Posterior credible intervals are plotted for γ and ψ . Dots indicate true parameter values.

B.2 High-Dimensional Example

This section examines the performance of our separable demand model with a fixed partition in high-dimensional settings. We consider data simulated from either an unrestricted log-linear model or a separable model. Let $n = 100$ denote the number of products. We explore both regular and high-dimensional settings in which T is either 75% greater than n (i.e., $T = 175$) or 75% less than n (i.e., $T = 25$), respectively. For the separable models, we let π_n be a partition with $K = 10$ contiguous groups where each group contains 10 products. Prices and expenditure are generated from a $\text{Unif}(0,1)$ distribution and the shares are generated from a $\text{Dirichlet}(1, \dots, 1)$ distribution. Unrestricted price elasticities are generated as $\beta_{ii} \sim \text{Unif}(-3, 0)$ and $\beta_{ij} \sim \text{Unif}(0, 3)$, the separability parameters are generated as $\theta_{k\ell} \sim \text{Unif}(0, 10)$, and the error covariance matrix is set to be $\Sigma = 0.1 + I_n$. Proper and relatively diffuse conjugate priors are used on all model parameters.

The results are summarized as follows: [Figure B.6](#) plots the posterior means and credible intervals for the n^2 elasticity parameters in each model; [Table B.1](#) reports in-sample and predictive fit statistics as well as the computational time associated with each model. We find that when T is large, the model that matches the data generating process performs the best. However, when T is small, the unrestricted models exhibit a tremendous amount of posterior uncertainty which render them practically useless. For example, the unrestricted model produces extremely wide 95% credible intervals that, on average, range from -35 to 35. This is also true for the unrestricted model's predictive fit, as the standard deviations of the predicted RMSE are more than 150 times larger than in the large T case. The separable model also exhibits more uncertainty when $n > T$, but to a much lesser degree. This suggests that separability can be a useful restriction for high-dimensional regression models, especially when the partitioning of goods can be reasonably specified a priori.

Data: unrestricted



Data: separable

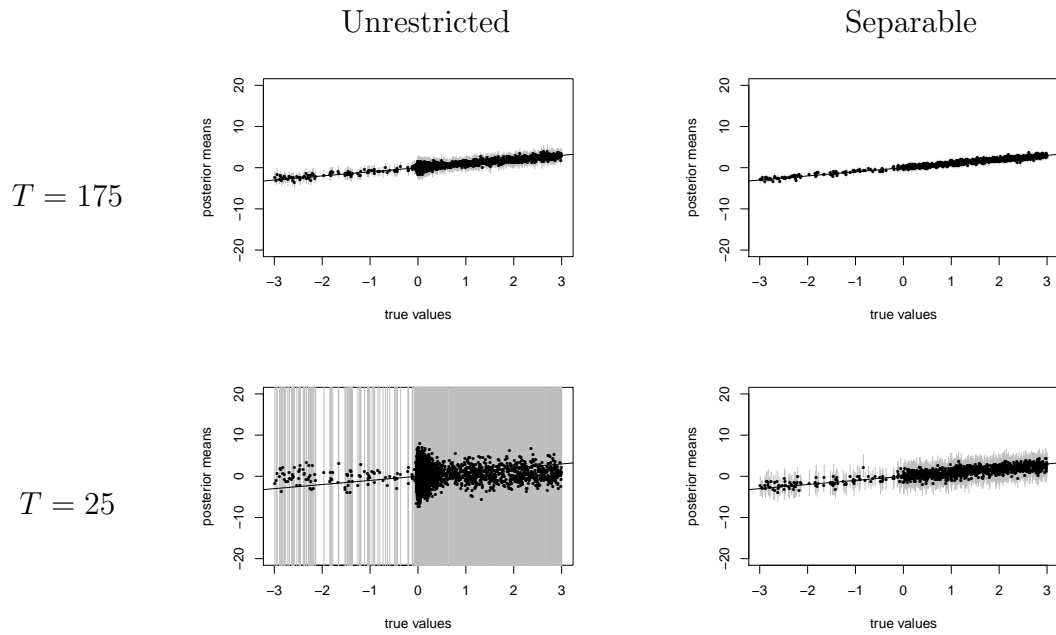


Figure B.6: Posterior means and credible intervals of price elasticities for the unrestricted and separable models. The top two panels correspond to data generated from an unrestricted model, while the bottom two correspond to data generated from a separable model.

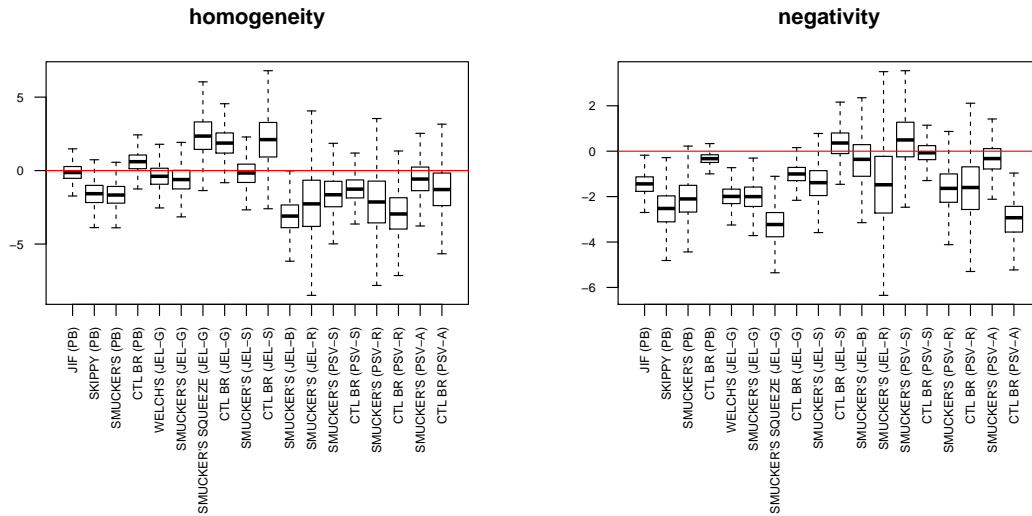
Table B.1

Data	Model	In-Sample RMSE		Predicted RMSE		Time (in sec)*
		Mean	SD	Mean	SD	
<i>Unrestricted</i>						
$T = 175$	Unrestricted	1.355	0.017	2.400	0.059	0.93
	Separable	71.502	1.242	70.000	1.236	30.39
$T = 25$	Unrestricted	6.123	1.294	49.949	10.76	0.75
	Separable	72.027	1.263	71.928	1.254	15.09
<i>Separable</i>						
$T = 175$	Unrestricted	1.358	0.015	2.433	0.053	0.90
	Separable	1.061	0.002	1.131	0.007	30.15
$T = 25$	Unrestricted	6.578	1.740	54.064	15.937	0.75
	Separable	1.318	0.032	1.830	0.043	15.15

*Time is measured as average computation time in seconds per 100 MCMC iterations.

C Integrability

(a) Jams, Jellies, and Spreads



(b) Juice Drinks

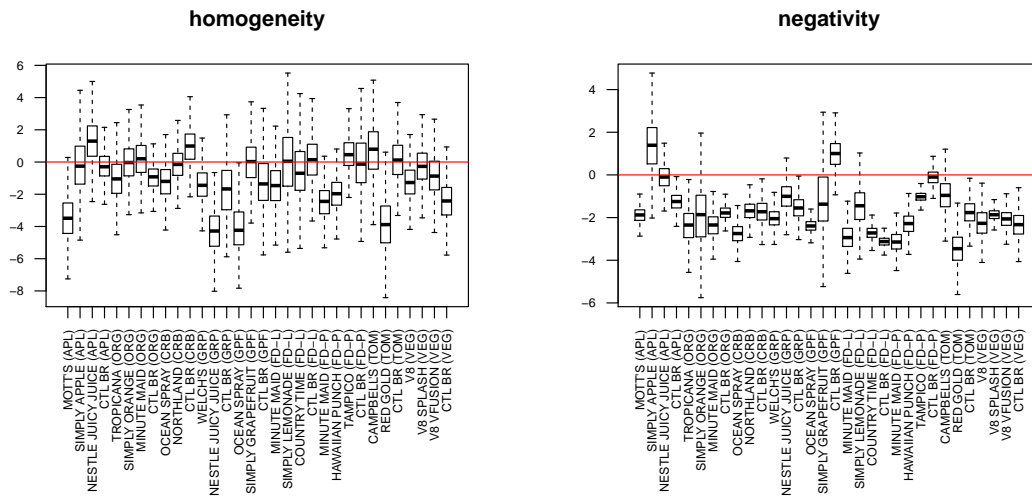


Figure C.1: The distribution of the homogeneity and negativity integrability conditions are plotted for each data set. Homogeneity holds whenever the boxplot intersects the horizontal line. Negativity holds whenever the boxplot intersects or falls below the horizontal line.

References

- Bajari, P., Nekipelov, D., Ryan, S. P., and Yang, M. (2015). Demand Estimation with Machine Learning and Model Combination . *American Economic Review: Papers and Proceedings*, 105(5), 481-485.
- Barten, A. (1964). Consumer Demand Functions under Conditions of Almost Additive Preferences. *Econometrica*, 32(1-2), 1-38.
- Bell, E. T. (1934). Exponential Numbers. *The American Mathematical Monthly*, 41(7), 411-419.
- Berry, S. T., Levinsohn, J., and Pakes, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, 63(4), 841–890.
- Blackorby, C., Primont, D., and Russell, R. R. (1977). On Testing Separability Restrictions with Flexible Functional Forms. *Journal of Econometrics*, 5(2), 195-209.
- Blackwell, D., and MacQueen, J. B. (1973). Ferguson Distributions Via Pólya 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.
- Chintagunta, P. K., and Nair, H. S. (2011). Structural Workshop Paper—Discrete-Choice Models of Consumer Demand in Marketing. , 30(6), 977-996.
- Deaton, A., and Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge University Press.
- DellaVigna, S., and Gentzkow, M. (2017). Uniform Pricing in US Retail Chains. *NBER Working Paper No. 23996*.
- Driscoll, P., McGuirk, A., and Alwang, J. (1992). Testing Hypotheses of Functional Structure: Some Rules for Determining Flexibility of Restricted Production Models. *American Journal of Agricultural Economics*, 74(1), 100-108.
- Dubé, J. P., Hitsch, G. J., Rossi, P. E., and Vitorino, M. A. (2008). Category Pricing with State-Dependent Utility. *Marketing Science*, 27(3), 417-429.

- Eddelbuettel, D., and François, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, 40(8), 1-18.
- 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*, 35(1), 182-197.
- 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.
- Hitsch, G. J., Hortagsu, A., and Lin, X. (2017). Prices and Promotions in US Retail Markets: Evidence from Big Data. *Booth Working Paper Series 17-18*.
- Jacobs, B. J. D., Donkers, B., and Fok, D. (2016). Model-Based Purchase Predictions for Large Assortments. *Marketing Science*, 35(3), 389-404.
- Jorgenson, D. W., and Lau, L. J. (1975). The Structure of Consumer Preferences. *Annals of Economic and Social Measurement*, 4(1), 49-101.
- LaFrance, J. T. (1986). The Structure of Constant Elasticity Demand Models. *American Journal of Agricultural Economics*, 68(3), 543-552.
- Leontief, W. (1947). Introduction to a Theory of the Internal Structure of Functional Relationships. *Econometrica*, 15(4), 361-373.
- Lindley, D. (1957). A Statistical Paradox. *Biometrika*, 44(1-2), 187-192.
- McFadden, D., and Train, K. (2000). Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics*, 15(5), 447-470.
- Mehta, N. (2007). Investigating Consumers' Purchase Incidence and Brand Choice Decisions across Multiple Product Categories: A Theoretical and Empirical Analysis. *Marketing Science*, 26(2), 196-217.

- Mehta, N. (2015). A Flexible Yet Globally Regular Multigood Demand System. *Marketing Science*, 34(6), 843-863.
- Montgomery, A. L. (1997). Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data. *Marketing Science*, 16(4), 315-337.
- 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., Jara, A., and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*. Springer Series in Statistics.
- Pollak, R. A. (1971). Conditional Demand Functions and the Implications of Separable Utility. *Southern Economic Journal*, 37(4), 423-433.
- 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. (2017). *bayesm: Bayesian Inference for Marketing/Micro-Econometrics*. R package version 3.1-0.1. <https://CRAN.R-project.org/package=bayesm>.
- Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley and Sons Ltd.
- Rubinfield, D. L. (2000). Market Definition with Differentiated Products: The Post/Nabisco Cereal Merger. *Antitrust Law Journal*, 68, 163–185.
- Smith, A. N., and Allenby, G. M. (2018). *Demand Models with Random Partitions*. Working paper.
- Sono, M. (1961). The Effect of Price Changes on the Demand and Supply of Separable Goods. *International Economic Review*, 2(3), 239-271.
- Thomassen, Ø., Smith, H., Seiler, S., and Schiraldi, P. (2017). Multi-Category Competition and Market Power: A Model of Supermarket Pricing. *American Economic Review*, 107(8), 2308-2351.