UNIVERSITY OF ALBERTA
FACULTY OF ARTS
Department of Economics

Working Paper No. 2025-01

Estimating Discrete Choice Demand Models with Sparse Market-Product Shocks

Zhentong Lu
Bank of Canada

Kenichi Shimizu
University of Alberta

January 2025

Estimating Discrete Choice Demand Models with Sparse Market-Product Shocks *

Zhentong Lu Kenichi Shimizu†

January 6, 2025

Abstract

We propose a new approach to estimating the random coefficient logit demand model for differentiated products when the vector of market-product level shocks is sparse. Assuming sparsity, we establish nonparametric identification of the distribution of random coefficients and demand shocks under mild conditions. Then we develop a Bayesian procedure, which exploits the sparsity structure using shrinkage priors, to conduct inference about the model parameters and counterfactual quantities. Comparing to the standard BLP (Berry, Levinsohn, & Pakes, 1995) method, our approach does not require demand inversion or instrumental variables (IVs), thus provides a compelling alternative when IVs are not available or their validity is questionable. Monte Carlo simulations validate our theoretical findings and demonstrate the effectiveness of our approach, while empirical applications reveal evidence of sparse demand shocks in well-known datasets.

Keywords: Demand Estimation, Sparsity, Bayesian Inference, Shrinkage Prior

JEL Codes: C1, C3, L00, D1

*We would like to thank the seminar participants of Bank of Canada, as well as the 58th Annual Meetings of the Canadian Economics Association, 2024 Econometric Society North American Summer Meeting, IAAE 2024 Annual Conference for helpful comments. The views expressed in this paper are the authors' and do not reflect those of the Bank of Canada's Governing Council.

†Lu: Financial Stability Department, Bank of Canada, Ottawa, K1A 0G9, Canada. Email: zlu@bankofcanada.ca. Shimizu: Department of Economics, University of Alberta, Edmonton, T6G 2H4, Canada. Email: kenichi.shimizu@ualberta.ca.

1 Introduction

Since Daniel L. McFadden's seminal work (McFadden, 2001), the discrete choice model has become a basic tool for understanding consumer demand for differentiated products in many empirical context. As an important advancement of the literature, the BLP framework (Berry, 1994; Berry, Levinsohn, & Pakes, 1995) allows researchers to estimate a flexible demand model that incorporates consumer preference heterogeneity and addresses price endogeneity, using aggregate (i.e., market-product level) data on price, quantity and other variables.

A prominent feature of the BLP framework is the inclusion of market-product-level demand shocks (a.k.a. unobserved market-product characteristics), which are observed by economic agents but not by econometricians. The dependence of price (or other endogenous variables) on these demand shocks provides a natural way to model the price endogeneity problem. However, this modeling strategy introduces a challenging estimation problem for two main reasons: (1) demand shocks enter the demand system nonlinearly, and (2) a dimensionality problem arises, as the number of parameters to estimate (including the demand shocks) exceeds the number of equations in the demand system. To address these challenges, BLP proposes inverting the demand system to recover the demand shocks, which are then interacted with a set of instrumental variables (IVs) to construct moment conditions for GMM estimation.

As emphasized by Berry and Haile (2014), the identification and estimation of the BLP model heavily depend on the availability of valid instrumental variables (IVs). In practice, finding suitable IVs for a specific empirical application is often challenging and requires consideration of data structure and availability, economic theory, institutional knowledge, and other contextual factors. Consequently, the literature has proposed and employed a wide range of IVs, including cost shifters (Berry, Levinsohn, & Pakes, 1999a; Goldberg & Verboven, 2001), BLP IVs (Berry, Levinsohn, & Pakes, 1995), Hausman IVs (Hausman, 1994; Nevo, 2001), optimal IVs (Berry, Levinsohn, & Pakes, 1999a; Reynaert & Verboven, 2014), differential IVs (Gandhi & Houde, 2019), and time-series or panel data-based IVs (Jin, Lu, Zhou, & Fang, 2021; Sweeting, 2013), among others. However, even with this variety

of alternatives, practitioners often face difficulties in selecting appropriate IVs, especially when estimation results are highly sensitive to the choice of IVs. This sensitivity may arise from the weak IV problem, a common concern in empirical research that can also emerge theoretically under specific model assumptions (Armstrong, 2016).

In this paper, we propose an alternative approach to estimating the BLP model that eliminates the need for instrumental variables (IVs) by leveraging a sparsity assumption on market-product-level demand shocks. Specifically, we assume that in some markets, the demand shocks for certain products take the same market-specific value. Under this assumption and regularity conditions, we demonstrate that the demand shocks (including their sparsity structure) can be identified as model parameters, alongside other parameters characterizing consumer preferences in the random coefficients logit demand model. The sparsity assumption reduces the number of unknown parameters, enabling identification directly from the constraints in the model without relying on IV-based restrictions.

Our identification strategy, based on the sparsity assumption, naturally leads to a likelihood-based inference framework, in contrast to the traditional BLP method, which relies on demand inversion and IV-based moment conditions. For inference, we develop a Bayesian shrinkage approach that incorporates the sparsity assumption to estimate model parameters, including the sparsity structure, as well as counterfactual quantities such as price elasticities.

To handle the potentially very high-dimensional space of sparsity patterns for the market-product demand shocks, we employ a type of shrinkage priors, a variable selection technique in Bayesian statistics. Shrinkage priors have their roots in high-dimensional statistics and machine learning literature, and are connected to penalized likelihood estimators such as LASSO (Tibshirani, 1996) in the sense that the posterior modes can be considered equivalent to these estimators (Casella, Ghosh, Gill, & Kyung, 2010). Shrinkage priors have been used successfully in linear econometric models such as VARs (e.g. Giannone, Lenza, & Primiceri, 2021); see e.g. Korobilis and Shimizu (2022) for a review of shrinkage priors and their applications to linear models in economics. However, their application to non-linear models has been limited, probably due to the lack of theoretical understanding on how introducing sparsity can aid in the identification of parameters in these models. We bridge this gap by

first studying how sparsity helps identification, which then motivates the use of shrinkage priors in our approach.

The proposed approach is both conceptually (returning to the likelihood framework for classic multinomial choice models) and computationally (offering a one-stop estimator for both preference parameters and the sparsity structure of demand shocks) straightforward. As such, it provides a compelling alternative to the BLP estimator, particularly when valid IVs are difficult to find or when researchers wish to evaluate the robustness of specific IV choices.

Monte Carlo simulation results show that, when the sparsity assumption holds in the data generating process (DGP), our approach performs similarly to the BLP estimator with strong IVs and outperforms the BLP estimator with potentially weak IVs. This supports our theoretical results on identification. Additionally, we examine cases where the demand shocks are not strictly sparse but exhibit approximate sparsity in the DGP, and find that our Bayesian shrinkage estimator still performs reasonably well in estimating the preference parameters, demonstrating robustness to mild misspecifications.

Depending on the empirical context, the sparsity assumption can have natural interpretations. For example, in supermarket scanner data applications where markets are defined by "store-week" pairs, the demand shocks may reflect unobserved promotion efforts at the store-week level for different products, usually identified by UPCs (Universal Product Code), after controlling for more aggregate fixed effects, such as brand, city, or quarter. In such cases, a store can only promote a selected subset of products in a given week due to limited shelf space (e.g., end-of-aisle displays), so the demand shocks for other products in the store-week will share the same value (either zero or the store-week level effect). Similarly, in Berry, Levinsohn, and Pakes (1995)'s automotive market application, the demand shocks largely capture unobservable advertising efforts, which can vary across brands and models in different markets. Some brands or models may engage in active advertising campaigns in a specific market, while others may choose to maintain a more "standard level" of marketing effort.

We explore these interpretations by applying our approach to two empirical applications. First, we apply it to a supermarket scanner dataset, focusing on the yogurt category. In

this case, we interpret the demand shocks as unobserved store-week-level promotion efforts. Our results demonstrate that the approach effectively captures the sparsity in promotion patterns across products, revealing important insights into consumer demand and store-level marketing strategies. Second, we revisit the automotive market data from Berry, Levinsohn, and Pakes (1995), where we treat advertising efforts as unobserved demand shocks. We explore how these advertising efforts vary across brands and models in different markets. Both applications demonstrate the flexibility of our approach in capturing sparsity in demand shocks, offering a robust alternative to traditional IV-based methods.

1.1 Related Literature

Our identification result is related to several findings in the literature on the identification of random coefficients in BLP and/or classic multinomial choice models, including, among others, Fox, il Kim, Ryan, and Bajari (2012), Fox and Gandhi (2016), Lu, Shi, and Tao (2023), and Dunker, Hoderlein, and Kaido (2023). While these results are developed under different assumptions and with distinct arguments, they are not directly applicable to our setting, where the ξ's are sparse and both the number of products and markets are growing.

Moon, Shum, and Weidner (2018) propose an approach to estimating the BLP model by modeling demand shocks as interactive fixed effects. While our paper shares a similar spirit of imposing structure on demand shocks, our identification and estimation strategies are fundamentally different. Their identification relies on additional exogenous variables and moment conditions, whereas ours depends solely on the sparsity condition described earlier. Moreover, their estimation strategy builds on least squares and minimum distance methods, while ours employs a Bayesian shrinkage approach. A related work by Gillen, Montero, Moon, and Shum (2019) introduces a LASSO-type estimator to select from a large number of control variables in the BLP model. In contrast, our focus is on addressing the challenge of high-dimensional demand shocks. Moreover, our estimation strategy differs significantly: their approach involves multiple steps of variable selection and requires post-selection inference, whereas ours is a Bayesian approach that delivers all results in a single pass.

Previous papers have proposed Bayesian estimation procedures of demand models for

aggregate data. The approach introduced by R. Jiang, Manchanda, and Rossi (2009) can be seen as a "Bayesian BLP," where the likelihood is constructed via demand inversion. Our approach differs in that, while they treat the market-product shocks as econometric residuals as in BLP, we treat them as parameters. As a result, our method does not require demand inversion and is more scalable with respect to the number of products and markets, unlike R. Jiang et al. (2009), which requires demand inversion at each MCMC iteration.

In contrast to this, other Bayesian approaches, such as those by Yang, Chen, and Allenby (2003) and Musalem, Bradlow, and Raju (2009), construct the likelihood by assigning an artificial set of consumer choices proportionally to the market shares. While this facilitates the estimation of the multinomial logit model for individual demand and the simulation of random coefficients, the sampling noise introduced at the data creation stage can potentially affect inference. Our approach avoids this issue by not requiring artificially assigned choices. Furthermore, as pointed out by Berry (2003), methods that assign artificial choices often lack a thorough discussion of identification and its relationship to prior restrictions. In contrast, our approach establishes a tight connection between identification under sparsity and estimation using shrinkage priors as a practical tool, effectively putting the identification argument into action.

Lastly, since our approach attempts to explore a large dimensional space of sparsity pattern of the market-product shocks, broadly speaking, this paper also contributes to the expanding literature on high-dimensional demand estimation: (Chiong & Shum, 2019: random projection for aggregate demand model with many products, Smith & Allenby, 2019: random partitions of products, Loaiza-Maya & Nibbering, 2022: high-dimensional probit models, Z. Jiang, Li, & Zhang, 2024: graphical lasso for flexible substitution patterns, Iaria & Wang, 2024: model of demand for bundles, Ershov, Laliberté, Marcoux, & Orr, 2024: estimation of complementarity with many products, Chib & Shimizu, 2024: scalable estimation of consideration set models).

The rest of the paper is organized as follows. Section 2 introduces a sparsity condition and establishes the identification of the model. Section 3 proposes a shrinkage-prior-based Bayesian estimation method. Section 4 investigates the performance of the proposed approach through Monte Carlo simulations. Section 5 applies the proposed approach to two

well-known real datasets and finds empirical evidence of sparsity in both. Section 6 concludes.

2 Model and Identification

2.1 Model

We consider a stylized random coefficient logit demand model for aggregate data, in the spirit of Berry, Levinsohn, and Pakes (1995). There are T markets, indexed by $t = 1, \ldots, T$, each consisting of $J_t + 1$ products, indexed by $j = 0, 1, \ldots, J_t$, and N_t consumers, indexed by $i = 1, \ldots, N_t$. The products indexed by $j > 0$ are "inside goods," and product 0 is the "outside option."

Each consumer i's utility from product j in market t is given by

$$u_{ijt} = X_{jt}^{\top}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

where $X_{jt} \in \mathbb{R}^{dx}$ is a vector of observed market-product characteristics, β_i represents consumer-specific taste parameters (i.e., random coefficients), which are i.i.d. across consumers and follow the distribution $f \in \mathcal{F}$, ξ_{jt} is the market-product level demand shock (a.k.a. unobserved characteristic), and ε_{ijt} is an i.i.d. idiosyncratic preference shock across i, j, and t, following the standard Gumbel distribution. To normalize the level of the random utility, the product characteristics and demand shock of the outside option, X_{0t}, ξ_{0t}, are set to zero.

As is typical in aggregate demand modeling, we allow certain variables in X_{jt}, such as price, to be endogenous. This endogeneity arises because these variables may depend on ξ_{jt}, which the firm can observe when setting prices or making other decisions, but econometricians cannot. We do not specify a detailed supply-side model for this dependence but will return to it later when discussing the identification assumptions.

In each market t, each consumer chooses the product that maximizes their utility, and

aggregating consumer choices gives the market share of each product j as follows:

$$\sigma_{jt}\left(\xi_t, f\right) = \int \frac{\exp\left(X_{jt}^{\top}\beta + \xi_{jt}\right)}{1 + \sum_{k=1}^{J_t} \exp\left(X_{kt}^{\top}\beta + \xi_{kt}\right)} f(\beta)\, d\beta, \tag{2}$$

where $\xi_t = (\xi_{1t}, \ldots, \xi_{J_t t})^{\top}$. Note that our setup encompasses the typical specification in most empirical applications, where some coefficients in X_{jt} are fixed (i.e., these random coefficients follow degenerate distributions). We will consider these special cases in the Monte Carlo simulations and empirical applications.

The observed market share of product j in market t is $s_{jt} = \frac{\sum_{i=1}^{N_t} y_{ijt}}{N_t}$, where y_{ijt} is an indicator that equals to 1 if i chooses product j in market t and 0 otherwise. By definition, the market share vector $(s_{0t}, ..., s_{J_t t}) \in \Delta^{J_t}$, where Δ^{J_t} is the standard J_t-simplex. The likelihood function of the observed choices is

$$L\left(f, \xi_1, ..., \xi_T\right) = \prod_{t=1}^{T} \prod_{i=1}^{N_t} \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, f\right)\right]^{y_{ijt}} = \prod_{t=1}^{T} \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, f\right)\right]^{q_{jt}}, \tag{3}$$

where $q_{jt} = \sum_{i=1}^{N_t} y_{ijt}$ is the total quantity of product j in market t. The aggregation across consumers (second equality) is due to the fact that the aggregate data do not contain individual-level attributes, e.g., demographics. When individual-level data are available, our approach can be modified easily to incorporate this information (such an extension is available upon request).

Without additional restrictions, estimating $(f, \xi_1, ..., \xi_T)$ based on the likelihood function (3) is infeasible, as there are only $\sum_{t=1}^{T} J_t$ linearly independent first-order conditions while the number of unknown parameters is $\sum_{t=1}^{T} J_t + \dim(f)$, where $\dim(f)$ denotes the dimensionality of f. In particular, these conditions effectively constitute the demand system

$$s_{jt} = \sigma_{jt}\left(\xi_t, f\right) \ \forall j, t, \tag{4}$$

and it is evident that the system is underidentified.

In the following, we first review the standard BLP approach to addressing this dimensionality problem and then introduce our new strategy to resolve it.

2.2 The BLP Approach

The standard BLP approach to addressing the identification problem consists of two main components. First, for a given f, the demand system in (4) is inverted (invertibility is established in Berry (1994) and Berry, Gandhi, and Haile (2013)) to obtain:

$$\xi_{jt} = \sigma_{jt}^{-1}(s_t, f), \quad \forall j, t. \tag{5}$$

Next, the ξ_{jt}'s are treated as econometric residuals, satisfying the following conditional moment restrictions:

$$E\left[\xi_{jt} \mid Z_{jt}\right] = E\left[\sigma_{jt}^{-1}(s_t, f) \mid Z_{jt}\right] = 0, \quad \forall j, t, \tag{6}$$

where Z_{jt} is a vector of instrumental variables (IVs). If the IVs provide sufficient variation, then the distribution f is identified and can be estimated via GMM.

Since there are endogenous product characteristics, such as price, and market shares in the moment conditions, the IVs Z_{jt} must include exogenous variables that are excluded from X_{jt} (see Berry and Haile (2014)). A practical challenge in empirical applications is that it can be difficult to identify and/or construct valid IVs for implementing the BLP GMM estimator. While this issue has been extensively discussed in the literature, there is still significant debate about how best to address it. Notable works on the subject include, among others, Berry, Levinsohn, and Pakes (1995), Berry, Levinsohn, and Pakes (1999b), Reynaert and Verboven (2014), Armstrong (2016), and Gandhi and Houde (2019).

2.3 Sparsity Assumption on Demand Shocks

We propose an alternative approach to the identification problem based on the demand system (4). Instead of imposing conditional moment restrictions (or other distributional assumptions) on ξ_{jt}'s, like (6), we treat ξ_t's as parameters to be estimated and assume they exhibit a sparsity structure: for some market t, a sub-vector of ξ_t shares the same value, as formally stated in Assumption 1.

Assumption 1 (Sparsity)**.** *There exist a set of markets $\mathcal{S} \subset \{1, ..., T\}$ and a set of products $\mathcal{K}_t \subset \{1, ..., J_t\}$ for each $t \in \mathcal{S}$, such that $\xi_{jt} = \xi_{kt}$ for any $j, k \in \mathcal{K}_t$. Furthermore, $|\mathcal{S}| \to \infty$*

and $|\mathcal{K}_t| \to \infty$ *for each* $t \in \mathcal{S}$.

Assumption 1 basically says that for any market $t \in \mathcal{S}$, all the ξ_{jt}'s for $j \in \mathcal{K}_t$ have the same value, which is denoted as $\nu_t \in \mathbf{R}$. So the number of unknowns in ξ_t is reduced by $|\mathcal{K}_t| - 1$ (from J_t), which in turn implies that the number of unknowns in the demand system (4) decreases by $\sum_{t \in \mathcal{S}} (|\mathcal{K}_t| - 1)$. Intuitively, the reduction in the number of unknown parameters can circumvent the dimensionality problem and restore the identification of $(f, \xi_1, ..., \xi_T)$.

To ensure that there is a sufficient reduction in the number of parameters, Assumption 1 further requires that: 1) in a market with sparse ξ_t, the number of products with the same value of ξ_{jt} goes to infinity as $J_t \to \infty$; 2) the number of such markets (with sparse ξ_t's) goes to infinity as $T \to \infty$. Theses requirements are necessary for the nonparametric identification of $(f, \xi_1, ..., \xi_T)$, which can be relaxed if $\dim(f)$ is finite due to parametric restrictions, e.g., Gaussian distribution.

Assumption 1 has important implications for the canonical price (and/or other product characteristics) endogeneity problem. To see this, let us consider a concrete example.

Example 1. *Suppose one element of* X_{jt} *is price* P_{jt} *and it is determined by a linear function*

$$P_{jt} = Z_{jt}^{\top} \rho + \xi_{jt},$$

where Z_{jt} *is the vector of IVs and* ρ *is a vector of parameters. Also, for any market* t, ξ_{jt} *is generated as*

$$\xi_{jt} = \begin{cases} \nu_t, & \text{with prob } \varphi \\ \xi_{jt}^*, & \text{with prob } 1 - \varphi, \end{cases}$$

where ν_t *and* ξ_{jt}^* *are two mean-zero random variables that are independent of each other. So the probability* φ *captures the degree of sparsity in* ξ*'s.*

In this case, the price endogeneity problem can be measured by the conditional (given Z_t*) covariance of* P_{jt} *and* ξ_{jt}*. Suppressing the conditioning variables* Z_t *for notational simplicity, we have*

$$Cov(P_{jt}, \xi_{jt}) = Var(\xi_{jt}) = E\left(\xi_{jt}^2\right) = \varphi \, Var(\nu_t) + (1 - \varphi) \, Var\left(\xi_{jt}^*\right). \tag{7}$$

When φ is small, i.e., not much sparsity in ξ's, the second term in (7) dominates so price endogeneity is captured by the variance of ξ^*_{jt}. As φ gets larger, i.e., more sparsity in ξ's, the variance of ν_t becomes more important in determining the severity of the endogeneity problem. Thus, if the across market variation of ν_t is small comparing to the across market-product variation of ξ^*_{jt}, then more sparsity implies less endogeneity. Of course, the converse is also true. To sum up, the sparsity assumption has certain implications on the price endogeneity problem, but does not necessarily reduce its severity.

In general, the sparsity assumption is neither stronger or weaker than the conditional mean restriction (6). The assumption (6) does not restrict the form of price endogeneity; however, it requires valid IVs. On the contrary, the sparsity assumption imposes certain restrictions on the form of price endogeneity; however, it avoids the need for IVs completely.

One particularly interesting feature of the sparsity assumption is that it does not impose any restrictions on the ξ_{jt}'s in the non-sparse set. In particular, these ξ_{jt}'s can either be realizations of any continuous or discrete distributions. Also, they can arbitrarily depend on X_{jt} and Z_{jt} (these IVs are not valid in this case). On the contrary, typical statistical assumptions on ξ_{jt}'s, such as (6) or other distributional assumptions, imply much stronger restrictions on the non-sparse set.

Next, we shall explore how Assumption 1 can help identify the model. The next subsection will establish the main identification result that shows that both ξ_t's and f are identified by the demand system (4) under the sparsity assumption and some other conditions. Before diving into the formal result, it is instructive to illustrate its key idea via a nested-logit example.

Example 2 (Nested-Logit with Sparse ξ)**.** *Consider a nested-logit model (as in Berry (1994)) of consumer demand for 4 inside goods and an outside option. For convenience, we focus on a single market and omit the subscript t. The products are grouped into mutually exclusive nests, with the outside option 0 being the only member in its own nest. The utility function*

of consumer i can be written as

$$u_{ij} = \begin{cases} \beta X_j + \xi_j + \zeta_{ig(j)} + (1-\lambda)\,\epsilon_{ij}, & j = 1,2,3,4 \\ \zeta_{i0} + (1-\lambda)\,\epsilon_{i0}, & j = 0, \end{cases}$$

where $X_j \in \boldsymbol{R}$ is an observed product characteristic, $\zeta_{ig(j)}$ is a random coefficient following a specific distribution (Cardell, 1997), $g(j)$ labels the nest of product j, λ is the "nesting parameter" and ϵ_{ij} is the "logit error" following the standard Gumbel distribution.

Suppose we know the sparsity pattern is $\xi_2 = \xi_3 = \xi_4 = \nu$. Then the parameters to be identified are $\beta, \lambda, \xi_1, \nu$. Given the close-form inversion of nested-logit model, we obtain the following linear system

$$\begin{aligned} \xi_1 + \beta X_1 + \lambda \log\left(\bar{s}_{1|g(1)}\right) &= \log\left(\frac{s_1}{s_0}\right) \\ \nu + \beta X_2 + \lambda \log\left(\bar{s}_{2|g(2)}\right) &= \log\left(\frac{s_2}{s_0}\right) \\ \nu + \beta X_3 + \lambda \log\left(\bar{s}_{3|g(3)}\right) &= \log\left(\frac{s_3}{s_0}\right) \\ \nu + \beta X_4 + \lambda \log\left(\bar{s}_{4|g(4)}\right) &= \log\left(\frac{s_4}{s_0}\right), \end{aligned}$$

where s_j is the market share of product j and $\bar{s}_{j|g(j)}$ denotes the within-group share of product j in its nest $g(j)$.

There are 4 equations and 4 unknowns, so the parameters are determined by

$$\begin{pmatrix} \xi_1 \\ \nu \\ \beta \\ \lambda \end{pmatrix} = \begin{bmatrix} 1 & 0 & X_1 & \log\left(\bar{s}_{1|g(1)}\right) \\ 0 & 1 & X_2 & \log\left(\bar{s}_{2|g(2)}\right) \\ 0 & 1 & X_3 & \log\left(\bar{s}_{3|g(3)}\right) \\ 0 & 1 & X_4 & \log\left(\bar{s}_{4|g(4)}\right) \end{bmatrix}^{-1} \begin{bmatrix} \log\left(\frac{s_1}{s_0}\right) \\ \log\left(\frac{s_2}{s_0}\right) \\ \log\left(\frac{s_3}{s_0}\right) \\ \log\left(\frac{s_4}{s_0}\right) \end{bmatrix}. \tag{8}$$

We can see that the identification condition in this special case boils down to the invertibility of the matrix in (8). The invertibility requires that the vector X cannot be collinear with the indicator variables for the sparse set (the first two columns in the matrix), which

automatically holds when X is continuous.

This example highlights the key role of the sparsity assumption in identification: it reduces the number of unknown parameters from 6 (β, λ and all the ξ's) down to 4 so we have sufficient number of equations. Based on this insight, we can expect that, to identify a more complicated model with more parameters, we will need data on more products and/or markets, as well as a sufficient degree of sparsity. Also, the nested-logit model, which is a special case of the general random coefficient logit model (1), has a close-form inversion in ξ's, so we can derive an explicit solution for the parameters. However, for the general model that does not have close-form inversion, establishing identification requires additional technical conditions and arguments.

2.4 Non-parametric Identification with Sparse Demand Shocks

In this subsection, we establish the formal non-parametric identification result, Theorem 1 under the sparsity assumption. Given Assumption 1, without loss of generality, suppose $\mathcal{K}_t = \{K_t + 1, ..., J_t\}$, i.e., the ξ_{jt}'s for the last $J_t - K_t$ products takes the same value ν_t. Then for each market t, let

$$\Xi_t = \left\{\xi_t \in \mathbf{R}^{J_t} : \xi_{jt} = \nu_t \text{ for any } j \in \mathcal{K}_t\right\}$$

denote the space of ξ_t's restricted by Assumption 1.

For a given $f \in \mathcal{F}$ and any market t, there are $K_t + 1$ unknowns, $(\xi_{1t}, ..., \xi_{K_t t})$ and ν_t, and J_t equations in the demand system (4). So intuitively we may only need the first $K_t + 1$ equations, i.e,

$$s_{jt} = \sigma_{jt}\left(\xi_t, f\right), \; j = 1, ..., K_t + 1, \tag{9}$$

to solve for the vector $(\xi_{1t}, ..., \xi_{K_t t}, \nu_t)$. Lemma 1 confirms that this is the case so the demand system (9), and hence (4), is invertible in $\xi_t \in \Xi_t$ for any $f \in \mathcal{F}$, which is a slight modification of the invertibility result in Berry (1994).

Lemma 1. *Suppose $s_{jt} > 0$ for any j, t. Then for any $f \in \mathcal{F}$ and any t, there is a unique $\xi_t \in \Xi_t$ that satisfies the demand system $s_{jt} = \sigma_{jt}\left(\xi_t, f\right)$, $j = 1, ..., J_t$. More-*

over, for any t, $\xi_{jt} = \tilde{\sigma}_{jt}^{-1}(\tilde{s}_t, f)$ *for any* $j = 1, ..., K_t + 1$, *where* $\tilde{s}_t = (s_{1t}, ..., s_{K_t+1,t})$ *and* $\left[\tilde{\sigma}_{1t}^{-1}(\tilde{s}_t, f), ..., \tilde{\sigma}_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right]$ *denotes the solution of the system (9).*

Proof. See Appendix A.1. □

Lemma 1 gives us the inversion of the subsystem (9), i.e., $\xi_{jt} = \tilde{\sigma}_{jt}^{-1}(\tilde{s}_t, f)$, $j = 1, ..., K_t+1$. We can substitute these inverse demand functions into the last $J_t - K_t - 1$ equations of (4) to obtain

$$s_{jt} = \sigma_{jt}\left(\tilde{\sigma}_t^{-1}(\tilde{s}_t, f), f\right), \ j = K_t + 2, ..., J_t, \forall t, \tag{10}$$

where $\tilde{\sigma}_t^{-1}(\tilde{s}_t, f) = \left[\sigma_{1t}^{-1}(\tilde{s}_t, f), ..., \sigma_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right]^\top$. Note that the only unknown object in (10) is f, so the identification problem becomes whether f is uniquely determined by (10). To establish identification, we need to introduce additional assumptions.

Assumption 2. *The random vector* X_{jt} *is independent across* j, t, *and has continuous and full support in* $\mathbf{R}^{d_X}$.

Assumption 2 rules out the cases where X_{jt} has discrete or bounded support. This is not surprising if we want to identify a continuous f nonparametrically. Similar continuous support assumptions are also imposed in related studies, see among others, the Assumption 2 of Fox et al. (2012) and the Assumption 7 of Lu et al. (2023). The continuous support assumption can be relaxed when a sub-vector of the coefficients on X_{jt} are fixed and/or f is parameterized by a finite number of parameters, as commonly estimated models in practice.

Also, Assumption 2 requires that X_{jt} has full support, so that we can establish identification based on the "identification at infinity" argument, as in Lewbel (2000), Khan and Tamer (2009), among others.

Assumption 3. *For any* $f \in \mathcal{F}$, $\sup_{x \in \mathcal{B}} \int \exp\left(x^\top \beta\right) f(\beta) \, d\beta < \infty$ *for any bounded open* $\mathbf{R}^{d_X}$*-ball* $\mathcal{B}$.

Assumption 3 is a regularity condition restricting the tail of f to be exponential or subexponential, which is satisfied by Gaussian distributions, for example. This assumption is necessary for our identification argument based on the uniqueness of Laplace transform; it can be viewed as an alternative restriction on the shape of f to the bounded support assumption imposed in Fox, il Kim, Ryan, and Bajari (2012).

As we shall show in the proof of Theorem 1, Assumption 1, 2, and 3 imply that f can be nonparametrically identified by the subsystem (10). Combining this result with Lemma 1, we can conclude that both ξ_{jt}'s and f are identified by the demand system (4), as stated in Theorem 1.

Theorem 1. *If the conditions of Lemma 1, Assumptions 1, 2 and 3 hold, then $\xi_t \in \Xi_t$ for all $t = 1, ..., T$ and $f \in \mathcal{F}$ are identified by the demand system (4).*

Proof. See Appendix A.2. □

Remark 1. *Comparing with the canonical identification result for BLP in Berry and Haile (2014), which relies on demand inversion and IVs, Theorem 1 exploits the sparsity structure in ξ's but does not require IVs. Note that demand inversion is still used in the proof of Theorem 1: we employ the inversion of the subsystem (9) to establish the identification of ξ's for a given f. However, in our context, the inversion is only used as a theoretical device in the proof; as we shall see later, our estimation procedure does not require explicitly computing the demand inversion. This stands in contrast with the BLP estimation strategy, where demand inversion is computed repeatably in the estimation procedure.*

3 Bayesian Shrinkage Approach to Estimation

The identification result established by Theorem 1 is conditional on the sparsity structure defined by Ξ_t's. If the sparsity structure were known, then we could simply implement the MLE (using (3)) with restrictions on ξ_t's defined by Ξ_t's. However, in practice, we typically do not observe the sparsity structure ex-ante, just as the situation in the classical high-dimensional regression context with many predictors. In this section, we shall borrow insights from the high-dimensional Bayesian statistics literature to design an inference procedure that can both uncover the latent sparsity structure and deliver estimates of the parameters of interests.

The latent sparsity structure is a high-dimensional object comprised of the following components:

(A) $\mathcal{S} \subset \{1, \ldots, T\}$, the set of "sparse" markets exhibiting sparsity in ξ's;

(B) $\mathcal{K}_t \subset \{1, \ldots, J_t\}$ for each $t \in \mathcal{S}$, the set of "sparse" products such that $\xi_{jt} = \nu_t$ for any $j \in \mathcal{K}_t$.

For each market $t \in \mathcal{S}$, there are essentially 2^{J_t} possible configurations of $\mathcal{K}_t$. Moreover, there are 2^T possible sets of sparse markets $\mathcal{S}$. For example, in the automobile market application of Berry, Levinsohn, and Pakes (1995), there are 20 markets and on average 110 products in each market, indicating a very large dimensionality of the parameter space that we need to explore when estimating the model.

The problem of finding such sparsity structure is akin to the variable selection problem in high-dimensional linear regression with many predictors for which frequentist penalized likelihood methods such as LASSO (Tibshirani, 1996) and Bayesian shrinkage prior methods are widely used. In this paper, we propose a stochastic search method based on shrinkage priors for the ease of uncertainty quantification (Casella et al., 2010, Womack, León-Novelo, & Casella, 2014, Porwal & Raftery, 2022). For recent applications of shrinkage priors in econometrics, primarily on linear models, see e.g., Giannone, Lenza, and Primiceri (2021), Koop and Korobilis (2023), and Smith and Griffin (2023). Our novel identification result in Theorem 1 motivates the use of shrinkage priors in the non-linear model considered in this paper.

Conceptually, ex-ante deviations of ξ_{jt} are penalized or "shrunk" towards zero via a prior distribution that is centered around zero. In particular, we employ a type of spike-and-slab priors (Mitchell & Beauchamp, 1988, George & McCulloch, 1993, George & McCulloch, 1997, Ishwaran & Rao, 2005, Narisetty & He, 2014, Ročková & George, 2018) to shrink ξ_{jt} towards zero. With the spike-and-slab priors, one can easily obtain probabilistic statement about sparsity (i.e. $\xi_{jt} = \nu_t$) in contrast to the penalized likelihood approaches which are based on constrained optimization problems.

3.1 The likelihood

We introduce our estimation procedure with a commonly used parametric specification of the distribution of random coefficients (f), for the ease of exposition. A non-parametric extension using sieve approximation, as in Lu, Shi, and Tao (2023) and Wang (2023), is

possible, given the general non-parametric identification result of Theorem 1, however, it is beyond the scope of the current paper so we leave it for future research.

Specifically, let the random coefficients follow an independent normal distribution:

$$\beta \sim N_{d_X}(\bar{\beta}, \Sigma),$$

where $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_{d_X}^2)$. Again we assume independence for the ease of exposition, it is straightforward to include non-zero off diagonal elements in Σ[1]. Note that this formulation nests the common case where only some of the d_X covariates are assigned random coefficients. For example, with $d_X = 3$ and if only the first variable has random coefficients, we have $\Sigma = \text{diag}(\sigma_1^2, 0, 0)$.

For convenience, we re-parametrize the non-negative elements in Σ as in R. Jiang, Manchanda, and Rossi (2009). First, we decompose $\Sigma = RR'$, where $R = \text{diag}(\sigma_1, \ldots, \sigma_{d_X})$. Then let $r = (r_1, \ldots, r_{d_X})'$ be the log standard deviations of the random coefficients, i.e. $r_k = \log(\sigma_k)$. Consequently, $R = \text{diag}(e^{r_1}, \ldots, e^{r_{d_X}})$. Next, let $\xi_{jt} = \bar{\xi}_t + \eta_{jt}$, where $\bar{\xi}_t$ is a "common shock" to market t and η_{jt} is the market-product (j,t)- specific deviation from $\bar{\xi}_t$. Note that a sparse vector ξ_t (as in Theorem 1) is a special case of this formulation: letting $\bar{\xi}_t = \nu_t$, so $\eta_{jt} = 0$ if $j \in \mathcal{K}_t$ and $\eta_{jt} = \xi_{jt} - \nu_t$ otherwise.

Given the above parameterization, the utility function (1) can be re-written as

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt},$$

where $\delta_{jt} = X_{jt}^\top \bar{\beta} + \xi_{jt}$ and $\mu_{ijt} = X_{jt}^\top R v_i$. The d_X-dimensional vector v_i is i.i.d. and follows the product of d_X independent standard normal distributions. Consequently, the predicted market share is

$$\sigma_{jt}\left(\xi_t, \bar{\beta}, r\right) = \int \frac{\exp\left(\delta_{jt} + \mu_{ijt}\right)}{1 + \sum_{k=1}^{J_t} \exp\left(\delta_{kt} + \mu_{ikt}\right)} \phi\left(v_i | 0, I\right) dv_i, \tag{11}$$

where $\phi\left(\cdot | 0, I\right)$ is the d_X-dimensional standard normal density. We approximate the integral

[1] Berry, Levinsohn, and Pakes (1995) also imposes independence.

based on R_0 i.i.d. draws of v_i from the normal distribution. The likelihood is defined as

$$p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T | q\right) = \prod_{t=1}^{T} \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, \bar{\beta}, r\right)\right]^{q_{jt}}, \tag{12}$$

where $\eta_t = (\eta_{1t}, \ldots, \eta_{J_t t})^\top$, $q = \{q_1, \ldots, q_T\}$ and $q_t = (q_{1t}, \ldots, q_{J_t t})^\top$.

Note that we could define $\bar{\xi}_t$'s as part of $\bar{\beta}$ by modifying the covariates appropriately, but especially when T is large, it is computationally more efficient to separately update the market specific intercepts from the slopes. We therefore treat them separately in what follows.

3.2 Prior

The sparsity assumption (Assumption 1) is a key identifying restriction we exploit to estimate the model. Our Bayesian approach naturally incorporates the sparsity restriction via a spike-and-slab prior on η_{jt}'s. A practical advantage of this approach is that it provides a one-step inference procedure for both model parameters and the latent sparsity structure of ξ's.

The (continuous) spike-and-slab prior (George & McCulloch, 1993, George & McCulloch, 1997) is a popular method for stochastic search variable selection. Other types of shrinkage priors can be also used in our framework, but the unique feature of the spike-and-slab type priors is the ease of interpretation of the estimated sparsity structure as we will see shortly.

Specifically, we define the prior on the unobserved market-product shocks as

$$\eta_{jt} \sim (1 - \gamma_{jt}) N(0, \tau_0^2) + \gamma_{jt} N(0, \tau_1^2), \tag{13}$$

independently over j and t, where $0 < \tau_0^2 \ll \tau_1^2$ are the prior variances in the two component mixture of normals with τ_0^2 taking a very small value and τ_1^2 a large value[2]. The binary indicator γ_{jt} equals to 0 if the η_{jt} belongs to the "spike" component and 1 if it is in the "slab" component. Intuitively, if $\gamma_{jt} = 0$, η_{jt} is shrunk toward zero (i.e. ξ_{jt} is shrunk toward

[2]Note that the original spike-and-slab prior has the dirac-delta function in place of $N(0, \tau_0^2)$ (Mitchell & Beauchamp, 1988). While the formulation (13) is an approximation to the original spike-and-slab prior, the mixture of two normals formulation has become popular due to its computational simplicity. Note that, by choosing τ_0^2 small enough, the spike component can made arbitrarily close to the dirac-delta function.

$\bar{\xi}_t$). On the other hand, when $\gamma_{jt} = 1$, it is unrestricted (i.e. ξ_{jt} deviates from $\bar{\xi}_t$). To be more precise, the posterior density on η_{jt} is proportional to the likelihood times its prior. When it belongs to the spike component, the prior dominates, and virtually η_{jt} is shrunk to zero in the posterior distribution. The posterior mean of γ_{jt} will inform us the ex-post uncertainty of whether η_{jt} is zero or not, i.e. $(jt) \in \mathcal{K}_t$ or not. Note that for each market t, the J_t-dimensional vector $\gamma_t = (\gamma_{1t}, \ldots, \gamma_{J_t t})'$ summarizes the 2^{J_t} possible sparsity patterns.

A priori, the binary variable γ_{jt}'s are i.i.d. and follow a Bernoulli distribution with the prior inclusion probability ϕ_t that is specific to market t so to allow for different degrees of sparsity across markets $t = 1, \ldots, T$:

$$\gamma_{jt} \overset{iid}{\sim} \text{Bernoulli}(\phi_t), \ j = 1, \ldots, J_t.$$

We follow the convention and specify a beta prior on ϕ_t:

$$\phi_t \overset{iid}{\sim} \text{Beta}(\underline{a}_\phi, \underline{b}_\phi), \ t = 1, \ldots, T.$$

Markets (t) in the sparse set $\mathcal{S}$ are associated with small values of ϕ_t, and market-product pairs (jt) with $\gamma_{jt} = 0$ are in the sparse product set $\mathcal{K}_t$. We set $(\underline{a}_\phi, \underline{b}_\phi) = (1, 1)$ so that ex-ante, all the market-specific prior inclusion probability ϕ_t has mean of 0.5. In other words, the prior probability that each market-product shock η_{jt} is sparse is 50%.

For the remaining parameters, we employ standard priors independently

$$\begin{aligned}
\bar{\beta} &\sim N_{d_X}(\underline{\mu}_\beta, \underline{V}_\beta), \\
\bar{\xi}_t &\overset{ind}{\sim} N(\underline{\mu}_{\xi_t}, \underline{V}_{\xi_t}), \quad t = 1, \ldots, T, \\
r_k &\overset{ind}{\sim} N(0, \underline{V}_{r,k}), \quad k = 1, \ldots, d_X,
\end{aligned}$$

a d_X-dimensional normal prior for the slope vector $\bar{\beta}$, an normal prior for the market-specific product shock $\bar{\xi}_t$ independently over markets, and a normal prior for the log standard deviation of the random coefficients independently over covariates. We let $(\underline{\mu}_\beta, \underline{V}_\beta) = (0, 10 \cdot I_{d_X})$ and $(\underline{\mu}_{\xi_t}, \underline{V}_{\xi_t}) = (0, 10)$ for all markets t to give sufficiently uninformative priors on the fixed

slope parameter $\bar{\beta}$ and the market specific shocks $\bar{\xi}_t$. For the log standard deviations, we let $\underline{V}_{r,k} = 0.5$ for all k to give sufficiently uninformative prior on Σ.

The hyperparameters (τ_0^2, τ_1^2) in the spike-and-slab prior (13) are chosen by the researcher. In the literature of shrinkage priors, it is known that computational problems can arise when the ratio τ_1^2/τ_0^2 is too large. They can be avoided when $\tau_1^2/\tau_0^2 \leq 10,000$ (George & McCulloch, 1993, George & McCulloch, 1997). We recommend to fix them as $(\tau_0^2, \tau_1^2) = (10^{-3}, 1)$ and use these values in our simulation studies and empirical applications below. The prior variance in the slab component $\tau_1^2 = 1$ is a reasonably large value. For example, in a canned-tuna category data, R. Jiang, Manchanda, and Rossi (2009) found the posterior mean of the (uniform) variance of the market-product shocks to be around 0.33; and in a facial tissue application, Musalem, Bradlow, and Raju (2009) found the corresponding value to be around 0.72. A semi-automated approach would be to fit the model under $\eta_{jt} \sim N(0, \tau^2)$ i.i.d. for all (jt) with an uninformative prior on τ^2, and let e.g., $\tau_0^2 = 10^{-2}\hat{\tau}^2$ and $\tau_1^2 = 10\hat{\tau}^2$ where $\hat{\tau}^2$ is the posterior mean. In our experience, the default option works better.

3.3 Posterior inference

We have the slope parameters $\bar{\beta}$, the log standard deviations for the random coefficients $r = (r_1, \ldots, r_{d_X})'$, the market-specific intercepts $\bar{\xi} = \{\bar{\xi}_1, \ldots, \bar{\xi}_T\}$, the market-product specific deviations $\eta = \{\eta_1, \ldots, \eta_T\}$, where $\eta_t = (\eta_{1t}, \ldots, \eta_{J_t t})'$, the binary indicator variables $\Gamma = \{\gamma_1, \ldots, \gamma_T\}$, where $\gamma_t = (\gamma_{1t}, \ldots, \gamma_{J_t t})'$, and the inclusion probabilities $\phi = (\phi_1, \ldots, \phi_T)'$. The data contains the quantity demanded $q = \{q_1, \ldots, q_T\}$, where $q_t = \{q_{1t}, \ldots, q_{J_t t}\}$ and the market-level covariates $X = \{X_1, \ldots, X_T\}$. Then, from Bayes theorem, suppressing the dependency on the covariates, the posterior density of interest is defined as

$$p\left(\bar{\beta}, r, \bar{\xi}, \eta, \Gamma, \phi | q\right) \propto p\left(q | \bar{\beta}, r, \bar{\xi}, \eta\right) \cdot p\left(\bar{\beta}, r, \bar{\xi}, \eta, \Gamma, \phi\right). \tag{14}$$

The first term on the right hand side is the likelihood function (12), and the second term in (14) gives the prior on the parameters and factors as $p\left(\bar{\beta}, r, \bar{\xi}, \eta, \Gamma, \phi\right) = p(\bar{\beta})p(r)p(\bar{\xi})p(\eta, \Gamma, \phi)$. The first three terms are the prior on $\bar{\beta}$, $\bar{\xi}$, and r, respectively. The last term defines the

prior on η_{jt}'s and

$$p\left(\eta, \Gamma, \phi\right) = p\left(\eta|\Gamma, \phi\right) p(\Gamma|\phi)\pi(\phi) = \prod_{t=1}^{T}\prod_{j=1}^{J_t}\left\{\phi(\eta_{jt}|0,\tau_0^2)^{1-\gamma_{jt}}\phi(\eta_{jt}|0,\tau_1^2)^{\gamma_{jt}}(1-\phi_t)^{1-\gamma_{jt}}\phi_t^{\gamma_{jt}}\right\}\pi(\phi_t), \tag{15}$$

where $\pi(\phi_t)$ is the prior on ϕ_t.

The model is estimated via Markov chain Monte Carlo (MCMC). We obtain a posterior sample $\{\bar{\beta}^{(g)}, r^{(g)}, \bar{\xi}^{(g)}, \eta^{(g)}, \Gamma^{(g)}, \phi^{(g)}\}_{g=1}^{G}$, where G is the total number of MCMC draws (after discarding an appropriate burn-in draws). Using the posterior sample, one can easily conduct inference on any functions of the model parameters such as elasticity.

Roughly speaking, our MCMC algorithm for sampling from the joint posterior distribution iterates between two sets of conditional distributions. The first set of conditionals is used for updating $(\bar{\beta}, r, \bar{\xi}, \eta)$, the utility parameters common across markets and products $(\bar{\beta}, r)$ as well as the market-specific intercepts and the market-product specific shocks $(\bar{\xi}, \eta)$. The second set of conditionals is for the parameters related to the latent sparsity structure (Γ, ϕ). The two sets of conditionals are:

$$\begin{aligned}
&\bar{\beta},\ r,\ \bar{\xi},\ \eta \,\big|\, \underline{\mu}_{\beta},\ \underline{V}_{\beta},\ \{\underline{V}_{r,k}\},\ \{\underline{\mu}_{\xi_t}, \underline{V}_{\xi_t}\},\ \Gamma,\ \tau_0^2,\ \tau_1^2,\ X,\ q \\
&\Gamma,\ \phi \,\big|\, \eta,\ \tau_0^2,\ \tau_1^2,\ \underline{a}_{\phi},\ \underline{b}_{\phi}.
\end{aligned}$$

The first step can be implemented using Metropolis-Hasting algorithm whose efficient sampling is made possible by exploiting the existence of gradients and Hessian matrices of the log-likelihood function with respect to the relevant parameters. The second set of conditionals can be implemented based on the standard conjugacy. The computational details of the algorithm can be found in Appendix B.

Remark 2. *Our proposed inference procedure has several appealing features. First of all, it is conceptually simple as it is based on the standard likelihood (12) (i.e., McFadden's classic framework) coupled with shrinkage priors on ξ's; in particular, it does not rely on the BLP machinery of demand inversion or IVs as additional identification restrictions.*

Moreover, since no demand inversion is needed, our method has two practical advantages over alternative approaches that are based on the inversion. First, our method can

accommodate zeros in market shares data, which is an important empirical problem in many applications, offering an alternative to existing approaches such as Gandhi, Lu, and Shi (2023). Second, our method is computationally more scalable than alternative Bayesian procedures like R. Jiang, Manchanda, and Rossi (2009), Hortaçsu, Natan, Parsley, Schwieg, and Williams (2023), where the inversion needs to be computed in each MCMC iteration; this advantage becomes more prominent as T and/or J_t's become large.

Finally, our method can conveniently deliver inference results (e.g., credible intervals) using posterior draws, for model parameters, including the sparsity structure of ξ's, and counterfactual quantities such as price elasticities. The computation of price elasticities is described in Appendix C and demonstrated in an empirical application in Section 5.

4 Monte Carlo Simulations

In this section, we examine the performance of our proposed approach via a series of Monte Carlo experiments, and compare with the standard BLP estimator with alternative IV choices.

4.1 Simulation Design

We generate data from the following random coefficient logit model, where the utility of consumer i for product j in market t is specified as

$$u_{ijt} = \beta_{pi} p_{jt} + \beta_w^* w_{jt} + \xi_{jt}^* + \varepsilon_{ijt},$$

where $\beta_{pi} \sim N(\beta_p^*, \sigma^{*2})$ is the random coefficient on the endogenous variable price p_{jt}, β_w^* is a fixed coefficient on the exogenous product characteristic w_{jt}, ε_{ijt} is i.i.d. across i, j, t following the standard Gumbel distribution.

The exogenous product characteristic w_{jt} is i.i.d. across j, t and generated from $U(1, 2)$, the uniform distribution with support $(1, 2)$. The endogenous variable price is generated as

$$p_{jt} = \alpha_{jt}^* + 0.3 w_{jt} + u_{jt},$$

where u_{jt} can be interpreted as a "cost shock" that is i.i.d. across j, t and drawn from a $N(0, .7^2)$. The unobserved market-product characteristics are generated as

$$\xi^*_{jt} = \bar{\xi}^*_t + \eta^*_{jt},$$

where $\bar{\xi}^*_t$ is fixed at -1 for all t.

The key parameters of interest are $\beta^*_p = -1$, $\beta^*_w = 0.5$, and $\sigma^* = 1.5$. The specification of α^*_{jt} and η^*_{jt} varies by the following four DGP designs: sparse ξ with exogenous p (DGP1), sparse ξ with endogenous p (DGP2), non-sparse ξ with exogenous p (DGP3), non-sparse ξ with endogenous p (DGP4).

In DGP1, for each t, the first 40% of the elements in the vector $\eta^*_t = (\eta^*_{1t}, \ldots, \eta^*_{Jt})^\top$ are non-zero: the odd components are set to 1 while the even ones are set to -1. The remaining 60% of the components are set to zero. The α^*_{jt} in the price equation is set to 0 for each (j, t), so the vector $\alpha^*_t = (\alpha^*_{1t}, \ldots, \alpha^*_{Jt})^\top$ is independent of η^*_t.

In DGP2, we introduce price endogeneity by letting α^*_{jt} depend on ξ^*_{jt}. In particular, we set η^*_{jt}'s the same as in DGP1 and let $\alpha^*_{jt} = 0.3$ if $\eta^*_{jt} = 1$, $\alpha^*_{jt} = -0.3$ if $\eta^*_{jt} = -1$, and $\alpha^*_{jt} = 0$ otherwise. This implies a positive correlation between price p_{jt} and the unobserved characteristics ξ^*_{jt}.

In DGP3 and DGP4, we consider a non-sparse structure of η^*_{jt}. In particular, they are i.i.d. draws from the normal distribution with zero mean and standard deviation 1/3, i.e., $\eta^*_{jt} \sim N(0, (1/3)^2)$. The distribution has a large mass around zero, which can be regarded as approximately sparse. The purpose of this design is to examine how our approach works when the sparsity assumption is mildly violated. The α^*_{jt} in DGP3 is the same as DGP 1. For DGP4, price is endogenous and positively correlated with ξ: we let $\alpha^*_{jt} = 0.3$ if $\eta^*_{jt} \geq 1/3$, $\alpha^*_{jt} = -0.3$ if $\eta^*_{jt} \leq -1/3$, and $\alpha^*_{jt} = 0$ otherwise.

Given the specification of the utility function, the market shares and quantities are simulated based on (11), using $N_t = 1000$ consumer draws from the distribution of the random coefficient on price. The number of products is the same across the markets i.e. $J_t = J$ $\forall t$. We consider different numbers of markets and products: $T \in \{25, 100\}$ and $J \in \{5, 15\}$. We simulate 50 data sets $\{(q^{(r)}, X^{(r)})\}_{r=1}^{50}$ for each case, and implement the following three

estimation strategies.

- The BLP estimator that uses $(1, w_{jt}, w_{jt}^2, u_{jt}, u_{jt}^2)$ as instruments, labeled as "BLP (with cost IV)", where u_{jt} is the exogenous cost shock in the price equation which is typically unobservable in real data. This estimator uses a set of valid IVs and provides a benchmark for comparing the other two approaches.

- The BLP estimator that uses $(1, w_{jt}, w_{jt}^2, w_{jt}^3, w_{jt}^4)$ as IVs, labeled as "BLP (without cost IV)". This set of IVs is a natural choice in practice when the only observed exogenous variable is w_{jt} and the cost shock is unavailable to the researcher. We also tried other IVs, including the BLP type of IVs, e.g., the sum of other products' w's, and they perform similarly or worse than our current choice. Note that this choice undermines the IV rank condition because the moments of w_{jt} tend to be highly correlated with each other. We use this case to illustrate the identification problem caused by poor choices of IVs, e.g., weak IVs, which may happen in practice.

- Our proposed Bayesian shrinkage approach with a spike-and-slab prior ("Shrinkage"). We use the priors described earlier and $R_0 = 200$ i.i.d. draws from the d_X-dimensional independent normal distribution for approximating the choice probabilities.

We report the estimation results of $\bar{\beta}$, σ, and ξ from the repeated study. For the Bayesian shrinkage approach, we use the posterior mean as the point estimator to make it comparable with the BLP estimator. For the BLP estimator, we estimate ξ by solving for the mean utility $\hat{\delta}_{jt}$'s at the estimated $\hat{\sigma}$ and define $\hat{\xi}_{jt} = \hat{\delta}_{jt} - x_{jt}^\top \hat{\beta}$.

4.2 Results

Table 1 reports bias and standard deviation of the estimators under DGP1 and DGP2. As expected, in general, BLP (with cost IV) and Shrinkage outperform BLP (without cost IV). In particular, BLP (without cost IV) has large biases and standard deviations in many cases, highlighting the potentially severe identification and estimation issues caused by weak or invalid IVs.

In both exogenous (panel (a)) and endogenous (panel (b)) cases in Table 1, the shrinkage approach clearly outperforms the BLP (without cost IV) in terms of bias and standard deviation; in many cases, it achieves similar or even better performance to the benchmark estimator BLP (with cost IV), especially in terms of estimating σ and ξ's. This result supports our identification strategy that exploits the sparsity of ξ instead of relying on IVs; also it shows that the Bayesian shrinkage inference procedure works well in the current setting.

To further confirm that our inference procedure works as expected, we examine the estimated sparsity pattern of ξ. The nice feature of the spike-and-slab prior is that it allows us to compute the posterior probability that η_{jt} is nonzero (i.e. ξ_{jt} deviates from the market-specific common shock $\bar{\xi}_t$), which is equivalent to the event $\gamma_{jt} = 1$. The last column of Table 1 reports the probability that $\gamma_{jt} = 1$ when the true value η^*_{jt} is indeed nonzero (first row) and the probability of the same event when η^*_{jt} is zero (second row). Overall, our procedure can uncover the sparsity structure in ξ reasonably well, giving a higher probability for the market-product pair (j, t) when the true value of η^*_{jt} is nonzero and a lower probability otherwise.

In DGP3 and DGP4, the market-product shocks are not sparse but approximately so. We consider such cases to examine the robustness of our approach when the sparsity assumption is mildly violated. Table 2 shows the results in the same format as Table 1. We can see that even with non-sparse ξ's in the data generating process, the proposed method out-performs BLP (without cost IV) and is comparable to BLP (with cost IV) in most cases.

In summary, the simulation studies indicate that the proposed approach effectively uncovers the latent sparsity structure in the unobserved market-product shocks ξ's when they are sparse. It also provides reliable estimates for other structural parameters under both sparse and non-sparse ξ's. Furthermore, our approach often matches the performance of the BLP estimator with strong but impractical IVs and outperforms the BLP estimator when poor IVs are used, making it a compelling alternative when good IVs are unavailable or their validity is uncertain.

Table 1: Simulation results of DGP1 and DGP2

			BLP (with cost IV)					BLP (without cost IV)					Shrinkage					
J	T		Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Prob.
5	25	Bias	0.07	0.07	-0.01	-0.39	0.17	0.06	0.64	-0.12	-0.53	0.56	0.05	0.06	-0.03	-0.13	0.05	1.00
		SD	0.10	0.13	0.05	0.60	0.66	0.35	1.01	0.24	1.38	1.04	0.10	0.09	0.06	0.18	0.62	0.20
5	100	Bias	0.02	-0.00	0.01	-0.10	0.15	-0.22	-2.08	0.55	0.03	1.58	0.07	0.08	-0.05	-0.15	0.06	1.00
		SD	0.08	0.12	0.04	0.37	0.66	0.97	7.13	1.80	2.27	2.25	0.08	0.11	0.07	0.16	0.62	0.23
15	25	Bias	-0.04	-0.07	0.00	0.13	0.15	-0.20	-0.47	0.05	0.62	0.67	-0.02	-0.01	0.01	0.01	0.03	0.99
		SD	0.07	0.07	0.03	0.26	0.68	0.62	1.61	0.35	2.37	1.18	0.01	0.01	0.01	0.01	0.60	0.09
15	100	Bias	-0.00	-0.01	-0.00	-0.00	0.13	-0.08	-0.97	0.27	0.23	0.84	-0.01	-0.01	0.01	0.01	0.03	0.99
		SD	0.04	0.05	0.02	0.13	0.66	0.30	2.39	0.66	0.98	1.33	0.00	0.00	0.00	0.00	0.60	0.09

(a) DGP1/sparse exogeneous case

			BLP (with cost IV)					BLP (without cost IV)					Shrinkage					
J	T		Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Prob.
5	25	Bias	0.04	0.02	0.01	-0.22	0.17	0.10	0.63	-0.13	-0.65	0.48	0.05	0.09	-0.02	-0.13	0.05	1.00
		SD	0.12	0.14	0.06	0.55	0.67	0.28	0.57	0.15	1.13	0.84	0.05	0.09	0.02	0.16	0.61	0.20
5	100	Bias	0.03	-0.00	0.01	-0.16	0.17	0.10	0.53	-0.13	-0.61	0.72	0.05	0.10	-0.04	-0.11	0.05	1.00
		SD	0.11	0.12	0.04	0.56	0.67	0.18	1.48	0.40	0.97	1.12	0.06	0.10	0.04	0.11	0.61	0.20
15	25	Bias	-0.04	-0.06	0.00	0.14	0.16	-0.31	-0.18	-0.07	1.05	0.98	-0.01	0.01	-0.00	0.01	0.03	1.00
		SD	0.08	0.07	0.03	0.27	0.68	0.67	1.85	0.52	2.55	1.51	0.01	0.01	0.01	0.02	0.60	0.10
15	100	Bias	-0.00	-0.01	-0.00	0.00	0.13	0.00	-0.18	0.05	0.03	0.71	-0.01	0.02	-0.00	-0.01	0.03	0.99
		SD	0.04	0.05	0.02	0.14	0.66	0.17	1.37	0.43	0.66	1.13	0.00	0.01	0.00	0.01	0.60	0.10

(b) DGP2/sparse endogeneous case

Note: The bias/SD of ξ are the averages of (absolute value of) bias/SD of ξ_{jt}. The Prob. column shows the posterior probabilities that $\gamma_{jt} = 1$ when $\eta^*_{jt} \neq 0$ (1st row) and that $\gamma_{jt} = 1$ when $\eta^*_{jt} = 0$ (2nd row), both averaged over j and t. The prior probability that $\gamma_{jt} = 1$ is 0.5. Int=the intercept term $\bar{\xi}$.

Table 2: Simulation results of DGP3 and DGP4

			BLP (with cost IV)					BLP (without cost IV)					Shrinkage				
J	T		Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ
5	25	Bias	0.05	0.01	-0.00	-0.21	0.63	-0.16	-0.08	-0.05	0.46	0.93	0.04	0.20	-0.07	-0.28	0.61
		SD	0.06	0.20	0.05	0.38	0.58	0.56	1.19	0.24	2.33	1.03	0.08	0.14	0.13	0.18	0.55
5	100	Bias	0.01	-0.05	0.01	0.06	0.6	-0.04	0.23	-0.06	0.17	0.81	0.13	0.17	-0.10	-0.31	0.57
		SD	0.07	0.10	0.03	0.33	0.58	0.32	0.60	0.20	1.57	0.91	0.13	0.08	0.07	0.29	0.54
15	25	Bias	0.01	0.00	-0.00	-0.07	0.60	-0.10	-0.36	0.07	0.29	0.83	-0.29	0.04	0.00	0.03	0.65
		SD	0.09	0.08	0.02	0.24	0.57	0.50	1.40	0.30	1.77	0.95	0.08	0.10	0.03	0.12	0.50
15	100	Bias	0.01	-0.00	0.00	-0.02	0.59	0.04	-0.19	0.07	-0.15	0.74	-0.26	0.03	-0.01	-0.04	0.62
		SD	0.03	0.06	0.02	0.14	0.58	0.14	0.73	0.23	0.57	0.79	0.04	0.04	0.04	0.06	0.50

(c) DGP3/non-sparse exogeneous case

			BLP (with cost IV)					BLP (without cost IV)					Shrinkage				
J	T		Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ	Int	β_p	β_w	σ	ξ
5	25	Bias	0.01	-0.01	-0.01	-0.07	0.4	-0.06	0.51	-0.16	0.14	0.63	0.01	0.13	-0.06	-0.11	0.35
		SD	0.08	0.11	0.03	0.29	0.39	0.25	0.8	0.23	0.88	0.74	0.08	0.06	0.07	0.17	0.31
5	100	Bias	0.03	0.01	-0.01	-0.08	0.39	0.06	0.27	-0.07	-0.3	0.46	0.02	0.12	-0.04	-0.11	0.34
		SD	0.05	0.08	0.02	0.22	0.39	0.17	0.36	0.11	0.81	0.50	0.06	0.06	0.03	0.11	0.30
15	25	Bias	-0.01	-0.05	0.01	0.03	0.39	0.05	0.32	-0.09	-0.22	0.52	-0.02	0.09	-0.02	0.02	0.32
		SD	0.04	0.07	0.01	0.14	0.39	0.17	0.75	0.19	0.75	0.57	0.03	0.05	0.03	0.09	0.25
15	100	Bias	0.01	0.01	0.00	-0.06	0.37	0.10	-0.08	0.05	-0.43	0.84	-0.01	0.10	-0.03	-0.00	0.31
		SD	0.03	0.04	0.01	0.11	0.38	0.14	1.85	0.52	0.62	0.99	0.01	0.03	0.02	0.03	0.24

(d) DGP4/non-sparse endogeneous case

Note: The bias/SD of ξ are the averages of (absolute value of) bias/SD of ξ_{jt}. Int=the intercept term $\bar{\xi}$.

5 Empirical Applications

In this section, we begin by applying the proposed method to analyze consumer demand and store promotion strategies in the yogurt market using the IRI dataset.[3] This application emphasizes the ability of our method to uncover the sparse patterns in store promotion activities, where only a few products receive special promotions in a given store during a specific week due to space constraints. In the second application, we revisit the U.S. auto market dataset from Berry, Levinsohn, and Pakes (1995) to assess the performance of our method in a well-documented market setting. In both applications, we find evidence of sparsity in the market-product level demand shocks, highlighting the relevance of our approach for capturing such latent structure and exploiting it for identification.

5.1 Consumer Demand and Store Promotion in Yogurt Market

5.1.1 Data

This analysis focuses on the yogurt category, using data from 95 stores located in New York market (defined by the IRI dataset) for a single week, the week of June 25 - July 1, 2012. This sample selection allows us to make the sample size manageable while retaining sufficient variation in product characteristics, prices, and promotional activities. Specifically, the data provide detailed UPC level information, including weekly price, quantity, product characteristics, and marketing mix variables, for each store in the sample.

We aggregate the UPCs into "products," which are defined by a combination of brand, size category (size 1-size 4 defined using three thresholds: 0.9, 1.3, and 1.9 pints), and product characteristics — such as flavored or not, low fat, Greek, organic, etc. — as shown in Table 4. Product prices are calculated as quantity-weighted averages, while quantities are obtained through simple summation across UPCs within each product. The marketing mix indicator variables, Display and Feature, are marked as active (equal to 1) if any UPC within the product is active. This aggregation decreases the number of observations while preserving the essential variation in product attributes.

A market is defined by a store, with the consumers' choice set comprising all products

[3]See Bronnenberg, Kruger, and Mela (2008) for a description of the IRI marketing dataset.

available in that store. The market share of a product is calculated as the quantity sold divided by the population size within the local area surrounding the store, as provided by the IRI dataset. In total, there are 5,927 unique market-product pairs in the data.

Table 4 presents summary statistics for several randomly selected products. The "No. of Market" column indicates the markets where each product is available, highlighting substantial variations in consumers' choice sets. The "Market Share (%)" and "Price" columns, as well as the marketing mix columns, report the averages across different markets for each product. One pattern that stands out is the greater variation in market shares across markets compared to prices, as reflected in the mean-to-standard-deviation ratio.

Table 4: Summary Statistics for Several Randomly Chosen Products

Product	No. of	Market	Price	Product Characteristics						Marketing Mix	
No.	Markets	Share (%)		Brand	Size (pt.)	Flavor	Low Fat	No Fat	Greek	Organic	Display	Feature
1	10	0.015	2.459	ALPINA	0.4	0	1	0	0	0	0	0
		(0.009)	(0.627)									
2	63	2.575	1.348	AXELROD	0.375	1	1	0	0	0	0.090	0.787
		(2.024)	(0.393)								(0.222)	(0.407)
3	20	0.026	1.482	AXELROD	2	0	0	1	0	0	0.044	0
		(0.027)	(0.099)								(0.211)	
4	90	1.370	3.051	CHOBANI	0.375	0	1	0	1	0	0.219	0.555
		(1.362)	(0.533)								(0.416)	(0.500)
5	52	0.015	3.643	CHOBANI	1	1	1	0	1	0	0	0.028
		(0.019)	(0.411)									(0.166)
6	58	0.053	3.033	CHOBANI	2	1	1	0	1	0	0.002	0.011
		(0.055)	(0.306)								(0.028)	(0.104)
7	51	0.069	2.211	YOPLAIT	0.375	1	1	0	0	0	0.042	0.144
		(0.071)	(0.318)								(0.202)	(0.354)
8	85	0.264	2.120	YOPLAIT	0.375	1	0	1	0	0	0.030	0.104
		(0.263)	(0.263)								(0.157)	(0.307)

Note: This table presents summary statistics for a selection of randomly chosen products in the yogurt category. The first number in each cell represents the mean, while the second number in parentheses indicates the standard deviation across markets. If the standard deviation is zero, it is omitted.

5.1.2 Model Specification

With the above data, we consider the following discrete choice demand model, where the utility function is specified as:

$$u_{ijt} = X_{jt}^{\top}\beta_i + \xi_{jt} + \varepsilon_{ijt},$$

where X_{jt} is a 24-dimensional vector of product characteristics, including price, dummy variables for 16 brands, 4 product sizes, and indicators for whether the product is flavored, non-fat, low-fat, Greek, or organic (i.e., $d_X = 24$). We introduce random coefficients on price and the organic indicator. Specifically, $\beta_i \sim N_{24}(\bar{\beta}, \Sigma)$, where $\Sigma = \text{diag}(\sigma_1^2, 0, \ldots, 0, \sigma_{d_X}^2)$.

The market-product demand shocks, which capture promotion efforts, are modeled as:

$$\xi_{jt} = \bar{\xi}_t + \eta_{jt},$$

where $\bar{\xi}_t$ represents a store-level demand shock, potentially reflecting overall store-level promotions, and η_{jt} denotes the product-specific deviation, driven by promotional efforts that could originate from the manufacturer or the store itself.

The product-specific promotion is naturally sparse due to the space constraints of stores, as a store can only promote a limited number of products in a given week. In the data, we observe variables such as display and feature, which partially capture these promotional efforts (and notably, these variables already exhibit a sparse pattern). However, they are noisy measures of the actual promotion effort, meaning some promotional activities are not recorded. Using our approach, we aim to directly estimate the underlying promotion efforts and, ex-post, evaluate how well the observed display and feature variables explain the estimated promotion.

Recall that there are 5,927 market-product pairs in the data, implying 5,927 independent first order conditions if one were to use the MLE to estimate the model defined by (12). However, the number of parameters to estimate is $5927(\eta_{jt}) + 95(\bar{\xi}_t) + 24(\bar{\beta}) + 2(\Sigma)$, making the model underidentified. As our theoretical result shows, introducing sparsity on η_{jt}'s can restore identification, and we implement our Bayesian shrinkage approach to estimate the

model under this sparsity assumption.

We use the priors defined in Section 3 and $R_0 = 200$ i.i.d. draws from the standard normal distribution for approximating the integrals in the choice probabilities. The MCMC procedure consists of 10,000 draws, with the first 3,000 discarded for burn-in, leaving 7,000 draws for estimation and inference.

For comparison, we implement the standard BLP GMM estimator using the same model specification. The instrumental variables (IVs) are constructed by interacting lagged prices (along with other product characteristics) with market dummies.[4] Additionally, we include results from simple logit specifications estimated using both OLS and IV methods for comparison.

5.1.3 Estimation Results

The estimation results for the preference parameters (β and Σ) are presented in Table 5. The estimated slope coefficients for price and the organic indicator in the proposed approach have reasonable signs and magnitudes, aligning closely with the results from the BLP approach. Both approaches also provide significant evidence of dispersion in the random coefficients for these variables. Additionally, well-known brands, such as Chobani, Fage Total, and Stonyfield Organic Oikos, exhibit relatively larger brand fixed effects in the consumer utility function. Finally, the random coefficient model (either Bayesian or BLP estimates) implies a more elastic demand compared to the model without random coefficients, as indicated by the last two rows of the table. This difference is primarily driven by the dispersion in the random coefficients on price, which captures heterogeneity in consumer sensitivity to price changes.

Overall, our Bayesian approach produces similar results to the BLP approach in this case. We emphasize that our Bayesian shrinkage approach does not rely on IVs, and the agreement between the two approaches here validates the BLP results that rely on IVs. However, such agreement is not guaranteed in general; in cases where the two approaches diverge, it becomes essential to assess which underlying assumption - sparsity or the validity

[4] We construct the BLP GMM estimator based on $E[\eta_{jt} | Z_{jt}] = 0$, where Z_{jt} is the vector of chosen IVs, and estimate $\bar{\xi}_t$'s as market fixed effects.

of IVs - is more plausible in the specific context.

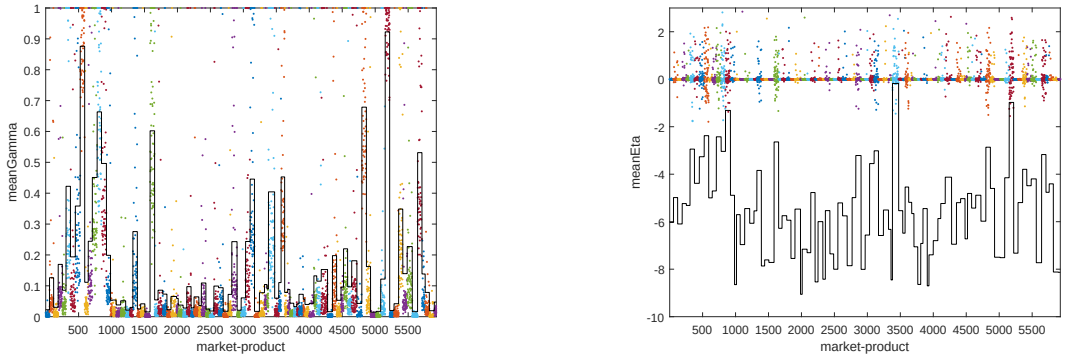
Table 5: Estimation results for the yogurt application: preference parameters

	Random Coefficient Logit								Simple Logit			
	Bayesian Shrinkage				BLP				IV		OLS	
	Mean of RC	CI	S.D. of RC	CI	Mean of RC	CI	S.D. of RC	CI	Mean	CI	Mean	CI
Price	-1.21	(-1.31 , -1.08)	0.42	(0.37 , 0.48)	-1.00	(-1.39 , -0.60)	0.35	(0.35 , 0.35)	-0.43	(-0.48 , -0.38)	-0.56	(-0.60 , -0.52)
AXELROD	0.13	(-0.02 , 0.29)			0.11	(-0.14 , 0.37)			0.29	(0.09 , 0.49)	0.18	(-0.04 , 0.40)
CABOT	0.62	(0.21 , 0.99)			0.57	(0.37 , 0.77)			0.63	(0.43 , 0.82)	0.54	(0.31 , 0.77)
CHOBANI	2.03	(1.92 , 2.16)			2.02	(1.87 , 2.17)			1.94	(1.79 , 2.09)	1.91	(1.77 , 2.06)
DANNON ALL NATURAL	-0.40	(-0.56 , -0.25)			0.19	(-0.04 , 0.42)			0.36	(0.22 , 0.50)	0.27	(0.13 , 0.41)
DANNON LIGHT N FIT	1.12	(0.93 , 1.30)			1.21	(1.21 , 1.21)			1.35	(1.15 , 1.56)	1.27	(1.08 , 1.47)
DANNON OIKOS	0.86	(0.67 , 1.08)			1.18	(1.18 , 1.18)			1.08	(0.93 , 1.23)	1.06	(0.89 , 1.24)
FAGE TOTAL	1.99	(1.75 , 2.21)			2.02	(1.94 , 2.11)			1.91	(1.77 , 2.06)	1.97	(1.82 , 2.12)
LA YOGURT	-0.13	(-0.25 , -0.01)			0.41	(0.21 , 0.62)			0.64	(0.42 , 0.86)	0.47	(0.30 , 0.64)
PRIVATE LABEL	-0.20	(-0.33 , -0.10)			0.30	(0.17 , 0.44)			0.57	(0.44 , 0.70)	0.45	(0.33 , 0.57)
STONYFIELD ORGANIC	1.04	(0.77 , 1.46)			1.17	(0.88 , 1.46)			1.32	(1.12 , 1.51)	1.27	(1.06 , 1.49)
STONYFIELD ORGANIC OIKOS	2.23	(1.90 , 2.66)			1.68	(1.68 , 1.68)			2.27	(2.01 , 2.53)	2.45	(2.17 , 2.73)
VOSKOS	0.46	(0.11 , 0.78)			0.70	(0.49 , 0.91)			0.64	(0.43 , 0.85)	0.61	(0.37 , 0.85)
YOPLAIT	-1.40	(-1.67 , -1.14)			-0.97	(-1.16 , -0.78)			-0.99	(-1.14 , -0.83)	-1.05	(-1.24 , -0.86)
YOPLAIT LIGHT	0.11	(-0.03 , 0.26)			0.33	(0.14 , 0.52)			0.42	(0.23 , 0.62)	0.32	(0.14 , 0.50)
YOPLAIT ORIGINAL	0.47	(0.32 , 0.61)			0.72	(0.65 , 0.80)			0.81	(0.63 , 0.99)	0.76	(0.54 , 0.98)
Size 2	-2.27	(-2.49 , -2.06)			-2.07	(-2.20 , -1.93)			-2.02	(-2.16 , -1.89)	-2.13	(-2.27 , -2.00)
Size 3	-3.04	(-3.95 , -2.30)			-2.45	(-2.69 , -2.21)			-2.45	(-2.67 , -2.24)	-2.52	(-2.79 , -2.25)
Size 4	-2.55	(-2.69 , -2.40)			-2.15	(-2.22 , -2.08)			-2.04	(-2.12 , -1.95)	-2.13	(-2.21 , -2.04)
Flavored	0.92	(0.81 , 1.02)			0.60	(0.49 , 0.71)			0.62	(0.55 , 0.70)	0.62	(0.54 , 0.69)
Nonfat	0.46	(0.33 , 0.64)			0.56	(0.46 , 0.66)			0.56	(0.45 , 0.66)	0.56	(0.45 , 0.67)
Lowfat	0.31	(0.16 , 0.50)			0.52	(0.42 , 0.62)			0.53	(0.43 , 0.63)	0.51	(0.40 , 0.62)
Greek	0.03	(-0.17 , 0.26)			-0.25	(-0.25 , -0.25)			-0.45	(-0.58 , -0.32)	-0.28	(-0.40 , -0.15)
Organic	-0.94	(-1.36 , -0.66)	0.25	(0.08 , 0.56)	-1.08	(-5.39 , 3.23)	0.65	(0.65 , 0.65)	-0.98	(-1.15 , -0.81)	-0.92	(-1.11 , -0.74)
Market FE	Omitted											
	Own Price Elasticity											
Mean	-2.00				-1.62				-1.22		-1.61	
S.D.	0.56				0.53				0.59		0.77	

Note: The table reports estimated preference parameters with the 95% credible/confidence intervals, as well as the averages of means and standard deviations of own-price elasticities. When the left end of the confidence interval for a SD of RC is negative, we replace it with 0 to respect the non-negative constraint on the parameter.

We now turn to discuss the latent sparsity structure of the market-product shocks η_{jt} uncovered by our procedure, as summarized in Figure 1. The solid lines in Figure 1a represent the posterior means of ϕ_t's, which indicate the degree of sparsity in each market. The average posterior mean is notably low, at 0.16, compared to the prior mean of 0.5. These results suggest that most markets in this dataset exhibit sparsity, meaning many markets belong to the sparse market set $\mathcal{S}$. Only a few markets exhibit dense η_{jt}'s, as reflected by posterior means of ϕ_t exceeding 0.7.

Figure 1: Estimated sparsity structure in the yogurt data.



(a) Posterior means of γ_{jt} (b) Posterior means of η_{jt}

(a): for each market t, the colored dots refer to posterior means of γ_{jt}'s and the solid line is the posterior mean of ϕ_t. The prior mean of ϕ_t is 0.5. (b): the colored dots represent posterior means of η_{jt}'s and the solid line is the posterior mean of $\bar{\xi}_t$. ξ_{jt} is the vertical sum. 5927 market-product pairs. 95 markets (stores).

The colored dots in Figure 1a show the posterior means of γ_{jt}'s. A smaller γ_{jt} means that market-product pair (jt) is more likely in the sparse set $\mathcal{K}_t$. Different colors of these dots correspond to different markets. Among the 5927 market-product pairs in the sample, there are only 658 pairs with posterior mean of γ_{jt} greater than 0.5, meaning that the majority of ξ_{jt}'s are shrunk towards the market-specific values $\bar{\xi}_t$ with high probability.

Figure 1b illustrates the posterior means of the shocks η_{jt} as colored dots and the market-specific intercepts $\bar{\xi}_t$ as solid lines. By definition, the posterior mean of ξ_{jt} is obtained by vertically summing η_{jt} and $\bar{\xi}_t$. As expected, most η_{jt}'s are effectively zero, confirming a high degree of sparsity in the data. An interesting pattern emerges: the distribution of η_{jt} is right-skewed, with more (jt) pairs exhibiting positive η_{jt}'s than negative ones. These positive η_{jt}'s likely capture store-product level promotion efforts, as discussed earlier.

The estimated market-product demand shocks reflect promotional efforts. To evaluate how much of these efforts are explained by the observed marketing mix variables, namely display and feature indicators, we regress the estimated η_{jt}'s on these variables. Table 6 presents the regression results for both the shrinkage and BLP approaches. The slopes on the marketing mix variables are positive and significant in both cases, suggesting that η_{jt}

effectively captures store-product-level promotion activities. However, based on the R^2, the marketing mix indicator variables explain only 2.5% of the estimated promotion efforts, highlighting the presence of potentially substantial unobserved store-level marketing activities. This finding underscores the noisiness of the marketing mix variables in capturing store promotions and emphasizes the importance of incorporating the market-product demand shocks ξ_{jt}'s into the model.

Table 6: Regression of the estimated η_{jt}'s on marketing mix variables.

	Shrinkage		BLP	
Intercept	0.046	(0.036, 0.0554)	-0.059	(-0.089, -0.030)
Display	0.287	(0.220, 0.355)	0.722	(0.507, 0.937)
Feature	0.101	(0.072, 0.130)	0.393	(0.300, 0.487)
Adjusted R^2	0.0248		0.0236	

Estimated slopes with 95% confidence intervals in parentheses.

Finally, we examine product-level price elasticities, which are a key output of demand estimation. As mentioned earlier, once MCMC draws are obtained, computing point and interval estimates of elasticity is straightforward, representing a key advantage of the proposed approach. Table 7 presents posterior means and 95% credible intervals of own-price elasticity for selected products in a market, sorted by price. For comparison, the table also includes elasticity estimates based on the BLP approach and the simple logit model (IV and OLS).

One notable observation is that when random coefficients (particularly on price) are incorporated, as in the shrinkage and BLP approaches, a U-shaped relationship between price and elasticity emerges, consistent with findings in the literature, such as Berry et al. (1995). In contrast, the simple logit model shows a monotonic increase in elasticity with price. Furthermore, the magnitudes of the elasticities estimated by our procedure are reasonable, and all products exhibit elastic demand (i.e., greater than 1).

Table 7: Price elasticity of some products in a market

		Own Price Elasticities				
Product	Price	Bayesian Shrinkage		BLP	Logit-IV	Logit-OLS
PRIVATE LABEL Nonfat Size 4	1.495	-1.41	(-1.54 , -1.30)	-1.19	-0.64	-0.84
AXELROD Lowfat Size 4	1.645	-1.50	(-1.65 , -1.39)	-1.27	-0.70	-0.92
DANNON ACTIVIA Flavored Lowfat Size 3	2.660	-1.97	(-2.21 , -1.83)	-1.66	-1.13	-1.49
BROWN COW Flavored Greek Size 1	3.973	-2.11	(-2.49 , -1.82)	-1.66	-1.70	-2.23
CHOBANI Flavored Lowfat Greek Size 1	4.452	-2.06	(-2.48 , -1.68)	-1.52	-1.90	-2.50
CHOBANI Flavored Nonfat Greek Size 1	4.470	-2.05	(-2.48 , -1.67)	-1.52	-1.91	-2.51
FAGE TOTAL Nonfat Greek Size 2	4.536	-2.05	(-2.48 , -1.65)	-1.49	-1.94	-2.54
DANNON GREEK Nonfat Greek Size 1	5.403	-1.85	(-2.38 , -1.32)	-1.08	-2.31	-3.03
THE GREEK GODS Nonfat Greek Size 1	5.840	-1.72	(-2.30 , -1.13)	-0.82	-2.50	-3.27
FAGE TOTAL Nonfat Greek Size 1	6.579	-1.48	(-2.14 , -0.80)	-0.38	-2.81	-3.69

The 95% credible intervals for the shrinkage approach shown in parentheses.

5.2 Revisit the BLP Auto Data

We revisit the classic BLP application to the U.S. automobile market. The BLP auto dataset contains product-level prices, quantities, and characteristics for major car models in the U.S. market for each year from 1971 to 1990. Following Berry et al. (1995), we define each year as a market, resulting in 20 markets ($T = 20$) and an average of approximately 110 products per market. A detailed description of the dataset is provided in Berry, Levinsohn, and Pakes (1995).

This additional application is valuable for several reasons. First, the industry context differs sharply: the automobile market involves durable goods and large, infrequent purchases, whereas the yogurt market represents low-cost grocery items and frequent consumption. Second, the scope of the data is distinct: the auto dataset captures national-level demand for automobiles, while the yogurt application focuses on highly disaggregated store-level activities. By applying our method to these two contrasting settings, we demonstrate its flexibility in uncovering sparse demand shocks in distinct scenarios with diverse market structures and datasets.

We consider a similar model structure as in the yogurt application. The product characteristics (X_{jt}) include price, horsepower/weight (log), weight (log), size (log), dollar/mile (log), and indicators for air conditioning, power steering, automatic transmission, and for-

ward drive ($d_X = 9$). Random coefficients on X_{jt} follow $\beta_i \sim N_{d_X}(\bar{\beta}, \Sigma)$, where $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_{d_X}^2)$. The market-product shocks (ξ_{jt}) are decomposed into market fixed effects ($\bar{\xi}_t$) and product-specific deviations (η_{jt}).

This specification differs from Berry et al. (1995) in two key ways: (1) it excludes a supply-side model, and (2) it includes market fixed effects to account for market-level heterogeneity. Our goal is not to replicate their results but to demonstrate how our approach can uncover sparsity in this classic dataset.

Compared to the yogurt application, the auto dataset differs in several aspects. While the yogurt data feature more markets and fewer products per market, the auto dataset includes fewer markets (20 years) but an average of 110 products per market, totaling 2,217 market-product pairs. As in the yogurt case, the number of parameters exceeds the number of independent first-order conditions, making sparsity assumptions on η_{jt} essential to restore identification. We use our shrinkage approach for estimation, following the same MCMC procedure. Again, for comparison, we include the standard BLP GMM estimator, with IVs constructed by interacting the BLP IVs with market dummies, as well as simple logit specifications (OLS and IV).

Table 8: Estimation results for the automobile application: preference parameters

	Random Coefficient Logit				Simple Logit	
	Bayesian Shrinkage		BLP		IV	OLS
	Mean of RC	S.D. of RC	Mean of RC	S.D. of RC		
Price	-0.29	0.12	-0.42	0.15	-0.10	-0.08
	(-0.41, -0.2)	(0.09, 0.17)	(-0.48, -0.36)	(0.13, 0.17)	(-0.11, -0.09)	(-0.09, -0.07)
HP/Weight (log)	-0.46	1.02	0.02	1.22	0.65	0.50
	(-1.4, 0.19)	(0.64, 1.65)	(-0.46, 0.51)	(0.77, 1.68)	(0.44, 0.85)	(0.24, 0.77)
Weight (log)	-0.56	0.57	0.38	0.09	-0.67	-1.45
	(-1.64, 0.42)	(0.23, 1.04)	(-0.41, 1.17)	(0, 1.29)	(-1.27, -0.08)	(-2.17, -0.74)
Size (log)	3.65	0.88	3.67	0.17	5.09	5.65
	(2.41, 4.73)	(0.3, 1.61)	(2.80, 4.44)	(0, 2.63)	(4.44, 5.73)	(4.83, 6.48)
Dollar/Mile (log)	-2.99	1.6	-1.36	0.69	-1.39	-1.14
	(-4.01, -2.12)	(0.99, 2.27)	(-1.86, -0.85)	(0.12, 1.25)	(-1.67, -1.12)	(-1.53, -0.75)
AC	0.39	0.39	0.66	0.38	0.25	.05
	(0.07, 0.71)	(0.2, 0.82)	(0.45, 0.87)	(0.02, 0.74)	(0.13, 0.37)	(-0.10, 0.20)
Power Steering	0.05	0.6	0.08	0.08	-0.19	-.28
	(-0.21, 0.3)	(0.34, 0.95)	(-0.07, 0.24)	(0, 0.53)	(-0.29, -0.09)	(-0.43, -0.14)
Automatic	0.12	0.42	0.30	0.03	0.28	.27
	(-0.1, 0.36)	(0.19, 0.76)	(0.14, 0.45)	(0, 0.50)	(0.18, 0.39)	(0.14, 0.41)
FWD	0.08	0.36	0.03	0.72	0.10	.15
	(-0.15, 0.29)	(0.17, 0.79)	(-0.11 0.16)	(0.48, 0.96)	(0.01, 0.18)	(0.03, 0.27)
Market FE	Omitted					
	Own Price Elasticity					
Mean	-1.52		-2.38		-1.06	-1.18
S.D.	0.49		0.96		0.78	0.86

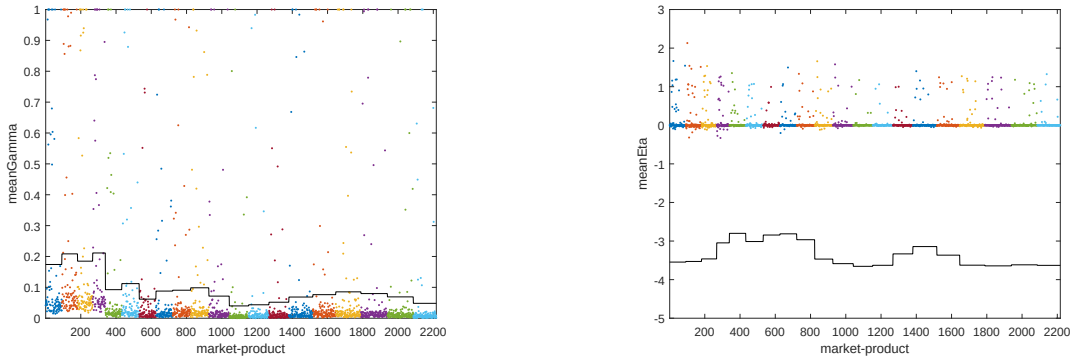
Note: The table reports estimated preference parameters with the 95% credible/confidence intervals, as well as means and standard deviations of own-price elasticities.

When the left end of the confidence interval for a SD of RC is negative, we replace it with 0 to respect the non-negative constraint on the parameter.

The estimation results for the preference parameters (β and Σ) are presented in Table 8. For the mean random coefficients, our Bayesian shrinkage approach yields estimates with reasonable signs and magnitudes, closely aligning with those of the standard BLP

estimates. Regarding the standard deviations (SDs) of random coefficients, the Bayesian shrinkage approach indicates considerable dispersion for all random coefficients, suggesting rich heterogeneity in consumers' tastes across all product characteristics. In contrast, several SDs from the BLP estimates, including those for weight, size, power steering, and automatic transmission, are virtually zero. These near-zero estimates may be attributed to the weak IV problem, as highlighted by Reynaert and Verboven (2014). Furthermore, while the BLP estimator is sensitive to the choice of IVs - based on our experiments with the data, though specific results are not reported here - our Bayesian shrinkage approach is immune to this issue, making it a particularly advantageous tool in practice.

Figure 2: Estimated sparsity structure in the automobile data.



(a) Posterior means of γ_{jt} (b) Posterior means of η_{jt}

(a): for each market t, the colored dots refer to posterior means of γ_{jt}'s and the solid line is the posterior mean of ϕ_t. The prior mean of ϕ_t is 0.5. (b): the colored dots represent posterior means of η_{jt}'s and the solid line is the posterior mean of $\bar{\xi}_t$. ξ_{jt} is the vertical sum. 2217 market-product pairs. 20 markets (years 1971-1990).

Now, we turn to the latent sparsity structure of the market-product shocks η_{jt} identified by our procedure, as summarized in Figure 2. This figure serves as the counterpart to Figure 1 from the yogurt application. Overall, we find stronger evidence of sparsity in this dataset compared to the yogurt data. The solid lines in Figure 2a show the posterior means of ϕ_t's, which indicate sparse markets. The average posterior mean is notably small, at 0.098, with particularly low values (less than 0.05) observed in the 1982, 1983, and 1990 markets. Among the 2,217 market-product pairs in this dataset, only 133 have a posterior mean of

γ_{jt} greater than 0.5, represented by the colored dots. Additionally, Figure 2b reveals that the distribution of η_{jt} is right-skewed, a pattern consistent with the yogurt application.

6 Conclusion

In this paper, we have proposed a new approach to estimating the random coefficient logit demand model with sparse market-product level demand shocks. Our approach eliminates the need for instrumental variables (IVs), which are required in the standard BLP GMM method. We show that, under certain regularity conditions, the demand shocks and their sparsity structure can be identified along with other model parameters. We also propose a Bayesian shrinkage estimation procedure that offers a scalable and flexible alternative to existing methods.

We demonstrate the applicability of our approach through two empirical applications. First, in the context of supermarket scanner data, we interpret the demand shocks as unobserved promotion efforts at the store-week level, capturing the sparsity in promotional activities across products. Second, we revisit the automotive market, where we model unobserved advertising efforts as demand shocks and show how our method identifies the underlying structure of these shocks across brands and models. In both cases, we find strong evidence of sparsity in demand shocks, supporting the relevance of the sparsity assumption in real-world data.

References

Armstrong, T. B. (2016). Large market asymptotics for differentiated product demand estimators with economic models of supply. Econometrica, 84(5), 1961–1980.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. The RAND Journal of Economics, 242–262.

Berry, S. (2003). Comment on bayesian analysis of simultaneous demand and supply. Quantitative Marketing and Economics, 1(3), 285–291.

Berry, S., Gandhi, A., & Haile, P. (2013). Connected substitutes and invertibility of demand. Econometrica, 81(5), 2087-2111. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA10135

Berry, S., & Haile, P. A. (2014). Identification in differentiated products markets using market level data. Econometrica, 82(5), 1749-1797. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9027

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, 841–890.

Berry, S., Levinsohn, J., & Pakes, A. (1999a). Voluntary export restraints on automobiles: Evaluating a trade policy. American Economic Review, 89(3), 400–431.

Berry, S., Levinsohn, J., & Pakes, A. (1999b, June). Voluntary export restraints on automobiles: Evaluating a trade policy. American Economic Review, 89(3), 400-430. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.89.3.400

Bronnenberg, B. J., Kruger, M. W., & Mela, C. F. (2008). The iri marketing data set. Marketing Science, 27(4), 745–748. doi: 10.1287/mksc.1080.0450

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. Econometric Theory, 13(2), 185–213.

Casella, G., Ghosh, M., Gill, J., & Kyung, M. (2010). Penalized regression, standard errors, and bayesian lassos. Bayesian Analysis.

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. The American Statistician, 49(4), 327–335.

Chib, S., & Shimizu, K. (2024). Scalable estimation of multinomial response models with random consideration sets. https://arxiv.org/pdf/2308.12470.

Chiong, K. X., & Shum, M. (2019). Random projection estimation of discrete-choice models with large choice sets. Management Science, 65(1), 256–271.

Dunker, F., Hoderlein, S., & Kaido, H. (2023). Nonparametric identification of random coefficients in aggregate demand models for differentiated products. The Econometrics Journal, 26(2), 279–306.

Ershov, D., Laliberté, J.-W., Marcoux, M., & Orr, S. (2024). Estimating complementarity with large choice sets: An application to mergers. RAND J. of Economics (accepted).

Fox, J. T., & Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. The RAND Journal of Economics, 47(1), 118–139.

Fox, J. T., il Kim, K., Ryan, S. P., & Bajari, P. (2012). The random coefficients logit model is identified. Journal of Econometrics, 166(2), 204–212.

Gandhi, A., & Houde, J.-F. (2019). Measuring substitution patterns in differentiated-products industries. NBER Working paper(w26375).

Gandhi, A., Lu, Z., & Shi, X. (2023). Estimating demand for differentiated products with zeroes in market share data. Quantitative Economics, 14(2), 381–418.

George, E. I., & McCulloch, R. E. (1993). Variable selection via gibbs sampling. Journal of the American Statistical Association, 88(423), 881–889.

George, E. I., & McCulloch, R. E. (1997). Approaches for bayesian variable selection. Statistica sinica, 339–373.

Giannone, D., Lenza, M., & Primiceri, G. E. (2021). Economic predictions with big data: The illusion of sparsity. Econometrica, 89(5), 2409–2437.

Gillen, B. J., Montero, S., Moon, H. R., & Shum, M. (2019). Blp-2lasso for aggregate discrete choice models with rich covariates. The Econometrics Journal, 22(3), 262–281.

Goldberg, P. K., & Verboven, F. (2001). The evolution of price dispersion in the european car market. The Review of Economic Studies, 68(4), 811–848.

Hausman, J. A. (1994). Valuation of new goods under perfect and imperfect competition. National Bureau of Economic Research Cambridge, Mass., USA.

Hortaçsu, A., Natan, O. R., Parsley, H., Schwieg, T., & Williams, K. R. (2023). Demand estimation with infrequent purchases and small market sizes. Quantitative Economics, 14(4), 1251–1294.

Iaria, A., & Wang, A. (2024). An empirical model of quantity discounts with large choice sets. Available at SSRN 3946475.

Ishwaran, H., & Rao, J. S. (2005). Spike and slab variable selection: frequentist and bayesian strategies.

Jiang, R., Manchanda, P., & Rossi, P. (2009). Bayesian analysis of random coefficient logit models using aggregate data. Journal of Econometrics, 149(2), 136–148.

Jiang, Z., Li, J., & Zhang, D. (2024). A high-dimensional choice model for online retailing. Management Science.

Jin, G. Z., Lu, Z., Zhou, X., & Fang, L. (2021). Flagship entry in online marketplaces (Tech. Rep.). National Bureau of Economic Research.

Khan, S., & Tamer, E. (2009). Inference on endogenously censored regression models using conditional moment inequalities. Journal of Econometrics, 152(2), 104–119.

Koop, G., & Korobilis, D. (2023). Bayesian dynamic variable selection in high dimensions. International Economic Review, 64(3), 1047–1074.

Korobilis, D., & Shimizu, K. (2022). Bayesian approaches to shrinkage and sparse estimation. Foundations and Trends® in Econometrics, 11(4), 230–354.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. Journal of econometrics, 97(1), 145–177.

Loaiza-Maya, R., & Nibbering, D. (2022). Scalable bayesian estimation in the multinomial probit model. Journal of Business & Economic Statistics, 40(4), 1678–1690.

Lu, Z., Shi, X., & Tao, J. (2023). Semi-nonparametric estimation of random coefficients logit model for aggregate demand. Journal of Econometrics.

McFadden, D. (2001). Economic choices. American economic review, 91(3), 351–378.

Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. Journal of the american statistical association, 83(404), 1023–1032.

Moon, H. R., Shum, M., & Weidner, M. (2018). Estimation of random coefficients logit demand models with interactive fixed effects. Journal of Econometrics, 206(2), 613–

644.

Musalem, A., Bradlow, E. T., & Raju, J. S. (2009). Bayesian estimation of random-coefficients choice models using aggregate data. Journal of Applied Econometrics, 24(3), 490–516.

Narisetty, N. N., & He, X. (2014). Bayesian variable selection with shrinking and diffusing priors.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. Econometrica, 69(2), 307–342.

Porwal, A., & Raftery, A. E. (2022). Comparing methods for statistical inference with model uncertainty. Proceedings of the National Academy of Sciences, 119(16), e2120737119.

Reynaert, M., & Verboven, F. (2014). Improving the performance of random coefficients demand models: The role of optimal instruments. Journal of Econometrics, 179(1), 83–98.

Ročková, V., & George, E. I. (2018). The spike-and-slab lasso. Journal of the American Statistical Association, 113(521), 431–444.

Smith, A. N., & Allenby, G. M. (2019). Demand models with random partitions. Journal of the American Statistical Association.

Smith, A. N., & Griffin, J. E. (2023). Shrinkage priors for high-dimensional demand estimation. Quantitative Marketing and Economics, 21(1), 95–146.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. Econometrica, 81(5), 1763–1803.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267–288.

Wang, A. (2023). Sieve blp: A semi-nonparametric model of demand for differentiated products. Journal of Econometrics, 235(2), 325-351. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304407622000860

Widder, D. V. (1941). Laplace transform (pms-6). Princeton: Princeton University Press. Retrieved from https://doi.org/10.1515/9781400876457

Womack, A. J., León-Novelo, L., & Casella, G. (2014). Inference from intrinsic bayes'

procedures under model selection and uncertainty. Journal of the American Statistical Association, 109(507), 1040–1053.

Yang, S., Chen, Y., & Allenby, G. M. (2003). Bayesian analysis of simultaneous demand and supply. Quantitative marketing and economics, 1, 251–275.

A Mathematical Proofs

A.1 Proof of Lemma 1

Proof. For any t, consider the demand functions for the first $K_t + 1$ products

$$\sigma_{jt}(\xi_t, f) = \int \frac{\exp\left(X_{jt}^\top \beta + \xi_{jt}\right)}{1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \xi_{kt}\right) + \sum_{k=K_t+1}^{J_t} \exp\left(X_{kt}^\top \beta + \nu_t\right)} f(\beta)\, d\beta, j = 1, ..., K_t+1. \tag{16}$$

Fixing any $f \in \mathcal{F}$, we will show that the $(K_t + 1)$-dimensional system

$$s_{jt} = \sigma_{jt}(\xi_t, f), j = 1, ..., K_t + 1 \tag{17}$$

uniquely determines $(\xi_{1t}, ..., \xi_{K_t t}, \nu_t)$.

Following the argument in the Appendix of Berry (1994), we just need to show the Jacobian matrix of the system (17) has a dominant diagonal, i.e.,

$$\frac{\partial \sigma_{jt}(\xi_t, f)}{\partial \xi_{jt}} > \sum_{m \neq j} \left| \frac{\partial \sigma_{jt}(\xi_t, f)}{\partial \xi_{mt}} \right|, \forall j = 1, ..., K_t + 1, \tag{18}$$

where $\xi_{K_t+1,t} = \nu_t$.

Observe that

$$\sum_{m=1}^{K_t+1} \frac{\partial \sigma_{jt}(\xi_t, f)}{\partial \xi_{mt}} = \int \dot{\sigma}_{jt}(\beta, \xi_t, f) \left[1 - \sum_{m=1}^{K_t+1} \dot{\sigma}_{mt}(\beta, \xi_t, f)\right] f(\beta)\, d\beta > 0,$$

where

$$\dot{\sigma}_{jt}(\beta, \xi_t, f) \equiv \frac{\exp\left(X_{jt}^\top \beta + \xi_{jt}\right)}{1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \xi_{kt}\right) + \sum_{k=K_t+1}^{J_t} \exp\left(X_{kt}^\top \beta + \nu_t\right)}.$$

Also, it is straightforward to verify that $\frac{\partial \sigma_{jt}(\xi_t, f)}{\partial \xi_{jt}} > 0$ for all j and $\frac{\partial \sigma_{jt}(\xi_t, f)}{\partial \xi_{kt}} < 0$ for any $k \neq j$. These inequalities imply that the dominant diagonal condition (18) holds. □

A.2 Proof of Theorem 1

Proof. Note that for any market t and product $j \geq K_t + 1$, the demand function can be written as

$$\begin{aligned}\sigma_{jt}(\xi_t, f) &= \int \frac{\exp\left(X_{jt}^\top \beta + \nu_t\right)}{1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \xi_{kt}\right) + \sum_{k=K_t+1}^{J_t} \exp\left(X_{kt}^\top \beta + \nu_t\right)} f(\beta)\, d\beta \\ &= \int \frac{\exp\left(X_{jt}^\top \beta\right)}{\left[1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \xi_{kt}\right)\right] \exp(-\nu_t) + \sum_{k=K_t+1}^{J_t} \exp\left(X_{kt}^\top \beta\right)} f(\beta)\, d\beta.\end{aligned}$$

Substitute ξ_{jt} with the inverse demand function $\tilde{\sigma}_{jt}^{-1}(\tilde{s}_t, f)$, we can obtain

$$\sigma_{jt}\left(\tilde{\sigma}_t^{-1}(\tilde{s}_t, f), f\right) = \int \frac{\exp\left(X_{jt}^\top \beta\right)}{\exp\left[H_t(\beta, f)\right]} f(\beta)\, d\beta, \tag{19}$$

where

$$H_t(\beta, f) \equiv \log\left\{\left[1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \tilde{\sigma}_{kt}^{-1}(\tilde{s}_t, f)\right)\right] \exp\left(-\tilde{\sigma}_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right) + \sum_{k=K_t+1}^{J_t} \exp\left(X_{kt}^\top \beta\right)\right\}.$$

Given market t, note that the right-hand-side of (19) varies by j only through X_{jt}, so we can define a market share function $\bar{\sigma}_t(\cdot)$ that does not have subscript j, i.e.,

$$\bar{\sigma}_t(X_{jt}, f) \equiv \sigma_{jt}\left(\tilde{\sigma}_t^{-1}(\tilde{s}_t, f), f\right).$$

Now consider the Laplace transform of the function $\frac{f(\beta)}{\exp[H_t(\beta, f)]}$,

$$\bar{\sigma}_t(x, f) = \int \frac{\exp\left(x^\top \beta\right) f(\beta)}{\exp\left[H_t(\beta, f)\right]} d\beta,\ \forall x \in \mathcal{B}, \tag{20}$$

where $\mathcal{B}$ is some bounded open $\mathbf{R}^{d_X}$-ball and Assumption 3 ensures the transform is well-defined. By the uniqueness of the inverse Laplace transform (see, e.g., the Theorem 6b of

Widder (1941)), if $\bar{\sigma}_t(x, f) = \bar{\sigma}_t(x, f^0)$ for all $x \in \mathcal{B}$, then

$$\frac{f(\beta)}{\exp[H_t(\beta, f)]} = \frac{f^0(\beta)}{\exp[H_t(\beta, f^0)]} \tag{21}$$

for all β.[5]

Next, we will show that the only f that satisfies (21) for all markets is $f = f^0$, where f^0 denotes the true value of f. For any $f \neq f^0$, there exists some $b_1 \neq 0$ and $b_2 \neq 0$ such that $f(b_1) > f^0(b_1)$ and $f(b_2) < f^0(b_2)$. Observe that for any β

$$\begin{aligned}
&\exp[H_t(\beta, f)] - \exp\left[H_t\left(\beta, f^0\right)\right] \\
=&\left[1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \tilde{\sigma}_{kt}^{-1}(\tilde{s}_t, f)\right)\right] \exp\left(-\tilde{\sigma}_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right) \\
-&\left[1 + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta + \tilde{\sigma}_{kt}^{-1}\left(\tilde{s}_t, f^0\right)\right)\right] \exp\left(-\tilde{\sigma}_{K_t+1,t}^{-1}\left(\tilde{s}_t, f^0\right)\right) \\
=&\Delta_{K_t+1,t}\left(f, f^0\right) + \sum_{k=1}^{K_t} \exp\left(X_{kt}^\top \beta\right) \tilde{\Delta}_{k,t}\left(f, f^0\right),
\end{aligned} \tag{22}$$

where

$$\Delta_{K_t+1,t}\left(f, f^0\right) = \exp\left(-\tilde{\sigma}_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right) - \exp\left(-\tilde{\sigma}_{K_t+1,t}^{-1}\left(\tilde{s}_t, f^0\right)\right),$$

$$\tilde{\Delta}_{k,t}\left(f, f^0\right) = \exp\left[\tilde{\sigma}_{kt}^{-1}(\tilde{s}_t, f) - \tilde{\sigma}_{K_t+1,t}^{-1}(\tilde{s}_t, f)\right] - \exp\left[\tilde{\sigma}_{kt}^{-1}\left(\tilde{s}_t, f^0\right) - \tilde{\sigma}_{K_t+1,t}^{-1}\left(\tilde{s}_t, f^0\right)\right].$$

Note that the terms $\Delta_{K_t+1,t}(f, f^0)$ and $\tilde{\Delta}_{k,t}(f, f^0)$ (for any k) do not depend on β. For any given f, let us examine the sign of (22) in the following three cases. First, if

$$\min\left\{\Delta_{K_t+1,t}\left(f, f^0\right), \min_{k \in \{1,\ldots,K_t\}}\left[\tilde{\Delta}_{k,t}\left(f, f^0\right)\right]\right\} > 0,$$

then (22) is positive for any β and thus (21) does not hold at b_2. Second, if

$$\max\left\{\Delta_{K_t+1,t}\left(f, f^0\right) < 0, \max_{k \in \{1,\ldots,K_t\}}\left[\tilde{\Delta}_{k,t}\left(f, f^0\right)\right]\right\} < 0,$$

[5] A related identification result is the Lemma 1 of Lu, Shi, and Tao (2023).

then (22) is negative for any β and thus (21) does not hold at b_1. Third, if

$$\min\left\{\Delta_{K_t+1,t}\left(f,f^0\right),\min_{k\in\{1,\ldots,K_t\}}\left[\tilde{\Delta}_{k,t}\left(f,f^0\right)\right]\right\}<0$$

and

$$\max\left\{\Delta_{K_t+1,t}\left(f,f^0\right)<0,\max_{k\in\{1,\ldots,K_t\}}\left[\tilde{\Delta}_{k,t}\left(f,f^0\right)\right]\right\}>0,$$

then (22) can be positive or negative depending on the vector $\left(X_{1t}^{\top}\beta,...,X_{K_t,t}^{\top}\beta\right)$. Assumption 2 implies that for any finite β, the random vector $\left(X_{1t}^{\top}\beta,...,X_{K_t,t}^{\top}\beta\right)$ has full support in $\mathbf{R}^{K_t}$. Thus (22) can be positive (or negative) with positive probability for any β; it follows that (21) does not hold at b_2 (or b_1) with positive probability.

Hence, for any $f\neq f^0$, there exists some $x\in\mathcal{B}$ such that $\bar{\sigma}_t\left(x,f\right)\neq\bar{\sigma}_t\left(x,f^0\right)$ with a positive probability. Due to the continuity of both sides of the inequality in x, the inequality holds for all x in a subset of $\mathcal{B}$ with a positive probability. Thus f^0 is identified. Furthermore, given f^0, Lemma 1 implies that $\left\{\xi_{jt}^0\right\}_{j,t}$ (ξ_{jt}^0 denotes the true value) are identified, which concludes the proof.

□

B Computation Details of the MCMC Procedure

In this section, we describe how the model is estimated in the proposed approach. Recall that we have the d_X-dimensional slope parameter $\bar{\beta}$, the log standard deviations for the random coefficients $r=(r_1,\ldots,r_{d_X})'$, the market-specific intercepts $\bar{\xi}=\{\bar{\xi}_1,\ldots,\bar{\xi}_T\}$, the market-product specific deviations $\eta=\{\eta_1,\ldots,\eta_T\}$, where $\eta_t=(\eta_{1t},\ldots,\eta_{J_tt})'$, the binary indicator variables $\Gamma=\{\gamma_1,\ldots,\gamma_T\}$, where $\gamma_t=(\gamma_{1t},\ldots,\gamma_{J_tt})'$, and the inclusion probabilities $\phi=(\phi_1,\ldots,\phi_T)'$. The data contains the quantity demanded $q=\{q_1,\ldots,q_T\}$, where $q_t=\{q_{1t},\ldots,q_{J_tt}\}$ and the market-level covariates $X=\{X_1,\ldots,X_T\}$. We obtain a posterior sample $\{\bar{\beta}^{(g)},r^{(g)},\bar{\xi}^{(g)},\eta^{(g)},\Gamma^{(g)},\phi^{(g)}\}_{g=1}^{G}$, where G is the total number of MCMC draws (after discarding an appropriate burn-in draws). The MCMC itertaes the following steps. The

order of the updates is arbitrary.

$$
\begin{aligned}
&\text{Draw } \bar{\beta}^{(g+1)} \text{ given } \bar{\beta}^{(g)}, \underline{\mu}_{\beta},\ \underline{V}_{\beta},\ r^{(g)},\ \bar{\xi}^{(g)},\ \eta^{(g)},\ q,\ X \\
&\text{Draw } r^{(g+1)} \text{ given } r^{(g)}, \{\underline{V}_{r,k}\},\ \bar{\beta}^{(g+1)},\ \bar{\xi}^{(g)},\ \eta^{(g)},\ q,\ X \\
&\text{Draw } \bar{\xi}^{(g+1)} \text{ given } \bar{\xi}^{(g)}, \{\underline{\mu}_{\xi_t}, \underline{V}_{\xi_t}\},\ \bar{\beta}^{(g+1)},\ r^{(g+1)},\ \eta^{(g)},\ q,\ X \\
&\text{Draw } \eta^{(g+1)} \text{ given } \eta^{(g)}, \Gamma,\ \tau_0^2,\ \tau_1^2,\ \bar{\beta}^{(g+1)},\ r^{(g+1)},\ \bar{\xi}^{(g+1)},\ q,\ X \\
&\text{Draw } \Gamma^{(g+1)} \text{ given } \eta^{(g+1)},\ \tau_0^2,\ \tau_1^2, \\
&\text{Draw } \phi^{(g+1)} \text{ given } \Gamma^{(g+1)},\ \underline{a}_{\phi},\ \underline{b}_{\phi}.
\end{aligned}
$$

Below, we illustrate how to conduct the updates above i.e. how to sample from the conditional posterior distributions of the parameters. Each of the conditional distributions below are defined given all other parameters, the hyperparameters, and the data, which are denoted by $\bullet$.

B.1 Sampling $\bar{\beta}, \bar{\xi}, \eta, r$

The conditional posterior for $(\bar{\beta}, \bar{\xi}, \eta, r)$ is

$$
\pi(\bar{\beta}, \bar{\xi}, \eta, r|\bullet) \propto p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T, r|q\right) \cdot \pi(\bar{\beta}) \cdot \pi(\bar{\xi}) \cdot \pi(\eta) \cdot \pi(r),
$$

where

$$
p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T, r|q\right) = \prod_{t=1}^{T} \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, \bar{\beta}, r\right)\right]^{q_{jt}}
$$

is the likelihood function and $\pi(\cdot)$ is the prior. The conditional posterior does not belong to a known class of distributions, so we employ a Metropolis–Hastings (M-H) algorithm to sample these parameters. One could update $\bar{\beta}, \bar{\xi}, \eta$, and r in one block, but the dimensionality of the parameter vector to be sampled is typically large (e.g. $d_X + T + TJ + d_X > 2,000$ in the auto market application), and the sampling might be inefficient. To increase the computational speed and avoid calculation of cross-derivatives, the parameter subvectors $\bar{\beta}, \bar{\xi}, \eta$, and r are set to be independent in the proposal. Furthermore, note that conditional on $\bar{\beta}$ and r, updating of $\{\bar{\xi}_t\}$ and $\{\eta_{jt}\}$ can be done independently over market t.

Specifically, for each $\theta \in \{\bar{\beta}, \eta\}$, we use a tailored Metropolis–Hastings (TMH) algorithm to sample θ from its conditional posterior (Chib & Greenberg, 1995). First, the mode of the conditional posterior, $\hat{\theta}$ is obtained

$$\hat{\theta} = \arg\max_{\theta} \log\left[L(\theta|\bullet)\pi(\theta)\right],$$

where $L(\theta|\bullet)$ is the likelihood function relevant to θ and $\pi(\theta)$ is the prior. The maximization is performed by a Newton's method. At iteration g, let $\theta^{(g)}$ be the value of θ. A candidate value is drawn as

$$\tilde{\theta} \sim N_{\dim(\theta)}\left(\hat{\theta}^{(g)}, \kappa_{\theta}^{2}\hat{V}_{\theta}\right),$$

where

$$\hat{V}_{\theta}^{-1} = -\frac{\partial^2}{\partial\theta\partial\theta'}\log L(\theta|\bullet)\pi(\theta)\Big|_{\theta=\hat{\theta}}.$$

This candidate is accepted with probability

$$\min\left\{\frac{\pi(\tilde{\theta}|\bullet)\phi(\theta^{(g)}|\hat{\theta}, \hat{V}_{\theta})}{\pi(\theta^{(g)}|\bullet)\phi(\tilde{\theta}|\hat{\theta}, \hat{V}_{\theta})}, 1\right\},$$

where $\phi(\quad)$ denotes the density of normal distribution. We fix κ_{θ} at 2.38 $\dim(\theta)^{-0.5}$, and if necessary, we tune it based on draws from a short chain that was run for the purpose of calibrating in order to achieve acceptance rate between 0.3 and 0.5. The likelihood is known to be concave with respect to each $\theta \in \{\bar{\beta}, \eta\}$ under the Gumbel error distribution, so the convergence to $\hat{\theta}$ is fast and only requires a few iterations in many cases. The gradients and Hessians of the log-likelihood with respect to $\bar{\xi}$ and r are also available and so TMH can be used, but the random-walk MH works efficiently for updating these parameters based on our experience, which we describe below.

B.1.1 Sampling $\bar{\beta}$

The conditional posterior for $\bar{\beta}$ is

$$\pi(\bar{\beta}|\bullet) \propto p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T|q\right) \cdot \pi(\bar{\beta}),$$

where

$$p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T | q\right) = \prod_{t=1}^{T} \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, \bar{\beta}, r\right)\right]^{q_{jt}}$$

is the likelihood function and $\pi(\cdot)$ is the prior.

B.1.2 Sampling η_{jt}

The conditional posterior for $\eta_t = (\eta_{1t}, \ldots, \eta_{J_t t})'$ is independent over t and

$$\pi(\eta_t|\bullet) \propto p\left(\bar{\beta}, r, \bar{\xi}_t, \eta_t | q_t\right) \cdot \prod_{j=1}^{J_t} \phi(\eta_{jt}|0, \gamma_{jt}\tau_1^2 + (1-\gamma_{jt})\tau_0^2),$$

for $t = 1, \ldots, T$, where

$$p\left(\bar{\beta}, r, \bar{\xi}_t, \eta_t | q_t\right) = \prod_{j=0}^{J_t} \left[\sigma_{jt}\left(\xi_t, \bar{\beta}, r\right)\right]^{q_{jt}},$$

is the market t's likelihood contribution.

B.1.3 Sampling $\bar{\xi}_t$

The conditional posterior for $\bar{\xi}_t$ is independent over t and

$$\pi(\bar{\xi}_t|\bullet) \propto p\left(\bar{\beta}, r, \bar{\xi}_t, \eta_t | q_t\right) \cdot \pi(\bar{\xi}_t),$$

for $t = 1, \ldots, T$. We employ a random walk Metropolis-Hastings (RWMH) algorithm to sample from the conditional posterior. At iteration g, let $\bar{\xi}_t^{(g)}$ be the value of $\bar{\xi}_t$. A candidate value is drawn as

$$\tilde{\bar{\xi}}_t \sim N\left(\bar{\xi}_t^{(g)}, \kappa_\xi^2 \underline{S}_\xi\right),$$

where $\underline{S}_\xi$ is a fixed scalar and κ_ξ is a scaling constant. We let $\underline{S}_\xi = 1$. We run an initial MCMC for the purpose of calibrating κ_ξ to achieve acceptance rate between 0.3 and 0.5. This candidate is accepted with probability

$$\min\left\{\frac{\pi(\tilde{\bar{\xi}}_t|\bullet)}{\pi(\bar{\xi}_t^{(g)}|\bullet)}, 1\right\}.$$

B.1.4 Sampling r

The conditional posterior distribution of r is

$$\pi(r|\bullet) \propto p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T|q\right) \cdot \pi(r),$$

where $p\left(\bar{\beta}, r, \bar{\xi}_1, ..., \bar{\xi}_T, \eta_1, \ldots, \eta_T|q\right)$ is the likelihood function and $\pi(\cdot)$ is the prior. We employ a random walk Metropolis-Hastings (RWMH) algorithm to sample from the conditional posterior. At iteration g, let $r^{(g)}$ be the value of r. A candidate value is drawn as

$$\tilde{r} \sim N_K\left(r^{(g)}, \kappa_r \underline{S}_r\right),$$

where $\underline{S}_r$ is a $K \times K$ scale matrix and κ_r is a scaling constant. We run an initial MCMC for the purpose of calibration. We first tune κ_r to achieve acceptance rate between 0.3 and 0.5, and set $\underline{S}_r$ as the estimated covariance matrix of the draws. This candidate is accepted with probability

$$\min\left\{\frac{\pi(\tilde{r}|\bullet)}{\pi(r^{(g)}|\bullet)}, 1\right\}.$$

B.2 Sampling γ_{jt}

We can derive the conditional posterior distribution of the binary indicator γ_{jt}, which is the Bernoulli distribution with the following success probability:

$$\Pr(\gamma_{jt} = 1|\bullet) = \frac{\phi_t \cdot \phi(\eta_{jt}|0, \tau_1^2)}{(1-\phi_t) \cdot \phi(\eta_{jt}|0, \tau_0^2) + \phi_t \cdot \phi(\eta_{jt}|0, \tau_1^2)}, \tag{23}$$

independently for $j = 1, \ldots, J_t$ and $t = 1, \ldots, T$.

B.3 Sampling ϕ

Under the conjugate prior i.e. $\phi_t \sim \text{Beta}(\underline{a}_\phi, \underline{b}_\phi)$, the posterior conditional distribution is available in the closed form:

$$\phi_t|\bullet \sim \text{Beta}\left(\underline{a}_\phi + \sum_{j=1}^{J_t} \gamma_{jt}, \underline{b}_\phi + \sum_{j=1}^{J_t}(1-\gamma_{jt})\right), \tag{24}$$

independently for $t = 1, \ldots, T$.

C Computing Price Elasticities

Price elasticities are a key output of demand estimation and provide a description of the substitution patterns among competing products implied by the estimated model. Importantly, elasticities are functions of the model parameters, and therefore, the posterior draws can be conveniently used for their uncertainty quantification. For example, after implementing the proposed MCMC, one can easily construct credible intervals for elasticities based on the following formulas.

The demand elasticity of product j with respect to the price change in product m in market t is

$$E_{jm,t}(\xi_t, \bar{\beta}, r) = \frac{\%\Delta\sigma_{jt}}{\%\Delta p_{mt}} = \frac{p_{mt}}{\sigma_{jt}(\xi_t, \bar{\beta}, r)} \cdot \frac{\partial \sigma_{jt}(\xi_t, \bar{\beta}, r)}{\partial p_{mt}}, \tag{25}$$

where p_{jt} is the observed price of product j in market t, β_{price} is the slope on price, and σ_{jt} is the model-predicted market share. The last term can be written as

$$\frac{\partial \sigma_{jt}(\xi_t, \bar{\beta}, r)}{\partial p_{mt}} = \int \beta_{price,i} \cdot \frac{\partial \sigma_{ijt}}{\partial \delta_{mt}} f(\beta_i) d\beta_i = \int \left(\bar{\beta}_{price} + \sigma_{price} v_{i,price}\right) \cdot \frac{\partial \sigma_{ijt}}{\partial \delta_{mt}} \phi(v_i|0, I) dv_i,$$

where $\bar{\beta}_{price}$ is the slope on price, σ_{price} is the standard deviation on the random coefficients on price, and $v_{i,price}$ is the element in the vector v_i corresponding to price. The partial derivatives are given as

$$\frac{\partial \sigma_{ijt}}{\partial \delta_{mt}} = \begin{cases} \sigma_{ijt} \cdot (1 - \sigma_{ijt}) \text{ if } j = m \\ -\sigma_{ijt} \cdot \sigma_{imt} \text{ if } j \neq m. \end{cases}$$