# Data-Driven Nests*

Milena Almagro†, Kenneth H. Lai‡, Elena Manresa§

March 2026

**PRELIMINARY AND INCOMPLETE**

**Abstract**

Nested demand structures are used to capture flexible substitution patterns while preserving tractability. Assuming a known nesting structure can be problematic, as misspecification may lead to biased estimates. We propose a simple two-step estimation strategy that recovers both the nesting structure and structural parameters using market-level data on choices and product characteristics. In the first step, we use the grouped fixed-effects estimator of Bonhomme and Manresa (2015) to estimate a non-overlapping nesting structure. In the second step, we estimate preference parameters conditional on the estimated nests. We consider empirical settings with longitudinal observations of products across several markets and potential price endogeneity. We demonstrate the performance of our method in Monte Carlo simulations, where we recover the population nesting structure with at least 99% accuracy. We apply our approach to the U.S. beer market using NielsenIQ data, identifying four distinct beer groups and estimating a demand system that reveals rich substitution patterns. We benchmark our model against several alternative demand models—including a naively grouped nested logit and a mixed logit—and show that it more accurately captures substitution patterns in an out-of-sample validation exercise.

**JEL classification:** C23, C25, C26, C51, C52, D12

**Keywords:** Demand estimation, nested demand, nested logit, grouped-patterns of heterogeneity.

* We thank audiences at NYU, Cornell, Microsoft, Berkeley, UCL, Caltech, ML Workshop Groningen, MSU Advances in High-Dimensional Inference Workshop, Midwest Econometrics Workshop, IO Montreal Summer Conference, Exeter, EARIE 2025, Chicago Booth, Minnesota, FDIC, FGV, UCLA, and MIT for helpful comments. We are grateful to JF Houde for his valuable discussion. We also thank Rodrigo Adão, Giovanni Compiani, Chris Conlon, JP Dubé, Jeff Gortmaker, Jacob Leshno, Chris Rauh, and Francesca Molinari for their feedback and suggestions. Marek Bojko, Junbiao Chen, Juan Felipe Herrera, and Sriram Tolety provided outstanding research assistantship. Disclaimer: Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

† University of Chicago, Booth and NBER, email: milena.almagro@chicagobooth.edu

‡ Ludwig-Maximilians-Universität Munich, email: H.Lai1@lmu.de

§ Princeton University, email: em7118@princeton.edu

# 1 Introduction

Discrete choice models of random utility are widely used to model and quantify demand across a broad range of applications. A common approach specifies demand according to a nesting structure, which can be interpreted as consumers making sequential choices across different levels of nests (McFadden, 1978). The appeal of nested demand models lies in their ability to remain parsimonious—preserving closed-form expressions of demand—while accommodating flexible substitution patterns. A well-known limitation, however, is that the nesting structure must be specified prior to estimation, which is often not obvious to the researcher. Misspecification of product groups can lead to unreliable parameter estimates (Fosgerau et al., 2024a).

In this paper, we propose a method to estimate the nesting structure and the remaining parameters of nested demand models. Our method relies on market-level data with many markets and many products. The identification strategy exploits the log-linear structure of market shares implied by the Nested Logit (NL), under which products belonging to the same nest share a common model-implied demand curve with nest-specific slopes and a nest-by-market intercepts. We allow for endogenous prices and exogenous price shifters, and build on the framework of Choi and Okui (2024) to consistently estimate the nesting structure in the presence of endogenous covariates.

Our proposed method consists of two steps. The first step applies a k-means-type clustering algorithm in the spirit of Bonhomme and Manresa (2015) on demand curves implied by the NL model to recover the nesting structure among products. In the second step, using this estimated nested structure, we estimate preference parameters following Berry (1994). We establish consistency of the estimated nesting structure of our first-step estimator under regularity conditions, extending Bonhomme and Manresa (2015). Moreover, we show that the nesting structure is recovered at a faster-than-polynomial rate. This faster-than polynomial convergence rate implies that misclassification in the first step does not affect the asymptotic behavior of the second-step estimator and thus there is no need to adjust the standard errors of the utility parameter estimates in the second step. We also show how to estimate the number of groups by drawing on clustering heuristics from the machine learning literature. Specifically, we apply the Elbow method within a cross-validation exercise to determine the optimal number of partitions.

We consider several extensions to the baseline model. First, we extend the baseline to a setting without exogenous price shifters, where prices are exogenous conditional on product fixed effects; in this case, endogeneity is handled by demeaning products within each market. Second, we show how the algorithm generalizes to recover multiple layers of nesting. Third, we incorporate consumer heterogeneity in a setting where both consumer characteristics and individual choices are observed. Finally, we show that our method extends to models where idiosyncratic shocks follow a Type II Extreme Value (Fréchet) distribution leading to nested CES demand, which is widely used in trade and spatial economics applications.

We conduct a Monte Carlo exercise to assess the ability of our method to recover both the true nesting structure and the underlying preference parameters. The experiment considers products observed across multiple markets, with one exogenous characteristic and one endogenous variable — prices. We find that with as few as ten markets, our method correctly assigns approximately 99% of products to their true groups. This match rate increases monotonically with the number of markets and products, reaching 100% in larger samples. We also validate our heuristic approach to selecting the number of groups.

We apply our estimator to demand for beer across U.S. states using store-level panel data from NielsenIQ. We first use the method with a simplified indirect utility, with continuous covariates only — prices, units, and liquid content. We recover four well-differentiated groups, which we label *Regular Lager*, *Light Beer*, *Special/Craft Beer*, and *Boutique Craft*. These groups exhibit substantial heterogeneity in product characteristics, including prices, beer type, packaging volume and size, and the likelihood of being a domestic product or a top-ten national brand. Notably, despite incorporating no spatial information in the estimation of the nesting structure, we find strong geographic correlation in product popularity: *Light Beer* is disproportionately consumed in the South, while *Boutique Craft* is concentrated in the Northeast.

We also perform an out-of-sample validation exercise to assess the substitution patterns implied by the groups recovered in the first step. The exercise exploits the Bud Light boycott as an exogenous negative demand shock. Using a difference-in-differences design, we define the treatment group as all *Light Beer* products other than Bud Light — i.e., products belonging to the same estimated nest — and the control group as products from all remaining groups. Following the boycott, *Light Beer* products experienced significantly larger sales

increases in 2023 relative to controls. These reduced-form results suggest that consumers perceive *Light Beer* products as closer substitutes to Bud Light than products from other groups, that we take as suggestive evidence that our estimated nesting structure provides reasonable substitution patterns. We confirm the robustness of these findings through several exercises, including group-specific treatment effect estimates and a placebo test using randomly assigned groups.

When we move to the second step estimation of structural parameters, we find an average elasticity of -3.561, similar to previous work studying beer ranging from -6.3 to -2.4 (Pinkse and Slade, 2004; Slade, 2004; Romeo, 2016; Miller and Weinberg, 2017). Moreover, we also find that all the nesting parameters are consistent with utility maximization and lie between 0 and 1 without the need to impose any additional restrictions in the structural estimation. We estimate a high degree of heterogeneity in substitution patterns across different groups, ranging from $\sigma = 0.042$ for *Boutique Craft* to $\sigma = 0.807$ for *Light Beer*. In terms of own-price elasticities, we find that *Light Beer* is the most inelastic group and *Boutique Craft* the most elastic. On the other hand, we also find that *Boutique Craft* has the largest within-group substitution effects following a price shock.

We also perform several robustness checks. First, we instrument for prices using Hausman instruments as an alternative to our baseline approach, finding that the optimal number of groups remains four and that the average characteristics within each group are similar to those in the baseline. Second, we examine the grouping structure for three and five groups, uncovering a broadly hierarchical pattern: with three groups, *Regular Lager* and *Special/Craft Beer* merge into a single group, while with five groups, *Regular Lager* splits into two. Third, when augmenting the covariate set with categorical information on beer type — one of the variables provided by NielsenIQ — yields marginal improvements in in-sample predictive power.

We finally compare the results of our baseline model to two common approaches to demand estimation. First, we estimate second-step parameters using the "naive" classification that leverages the categorical variable "Beer Type" provided by Nielsen as the nesting structure. Second, we also estimate a random coefficient model—the most flexible model of random utility. When doing so, we find that the Data-Driven NL outperforms these two alternative approaches in the out-of-sample validation test in the ability to predict substitution patterns. When looking at the results of the structural estimation, we find that some

of the parameters of the Naive NL are not consistent with utility maximization and that the substitution patterns implied by the mixed logit resembled those of a simple logit model.

**Literature review.** This paper contributes to the literature on discrete choice demand models by proposing a data-driven approach to estimate the nesting structure of the Nested Logit (NL) model. Introduced by McFadden (1978), the NL model relaxes the IIA restriction of the standard logit by allowing correlated unobserved utility within groups of alternatives. It is best understood as an intermediate case between the standard logit and the Random Coefficients (RC) model of Berry et al. (1995): it retains computational tractability while allowing for within-group preference correlation, and can be interpreted as a particular RC specification in which a random coefficient loads on group dummies (Cardell, 1997).[1] While RC models offer greater flexibility, they require numerical integration and can face computational challenges in large-scale applications (Dubé et al., 2012; Knittel and Metaxoglou, 2014). Grigolon and Verboven (2014) document this trade-off empirically using European car data, finding that both the NL and RC models are rejected in favor of a hybrid RC-NL specification, yet yield similar conclusions for merger simulations—and that RC models on continuous characteristics alone cannot fully capture discrete sources of market segmentation. Thanks to their tractability, nested demand models have been widely applied in industrial organization (Verboven, 1996; Parry and Small, 2005; Einav, 2007; Grennan, 2013; Ciliberto and Williams, 2014; Berry et al., 2016; Björnerstedt and Verboven, 2016; Allcott et al., 2024), international trade (Goldberg, 1995; Broda and Weinstein, 2006; Verhoogen, 2008; Atkeson and Burstein, 2008; Khandelwal, 2010; Fajgelbaum et al., 2011), urban and spatial economics (Davis et al., 2019; Couture et al., 2024; Bordeu, 2024; Couture et al., 2025), and labor economics (Azar et al., 2022; Grigsby, 2022). Crucially, however, all of these applications *assume the nesting structure is known and specified a priori*. In many empirical contexts the appropriate grouping of products is far from obvious. Our paper departs from this assumption by treating the nesting structure as an object to be estimated from the data rather than being imposed by the researcher.

This paper addresses this limitation by estimating the nesting structure from the data. Along these lines, Hortacsu et al. (2023) is the closest paper to ours. They propose a method that exploits non-negativity and sparsity constraints to estimate the nesting structure of an

[1] More generally, any discrete choice model can be approximated by an RC model (McFadden and Train, 2000).

Inverse Generalized Nested Logit model (Fosgerau et al., 2024b) using aggregate data on market shares, prices, and product characteristics. Their approach allows for overlapping nests, which can generate flexible substitution patterns, but it is designed for settings with a small number of products and an asymptotic sequence in the number of markets. In contrast, we study a panel data setting with many products and focus on recovering non-overlapping nests as the product space grows large. A related contribution is Kovach and Tserenjigmid (2022), who provide a revealed-preference characterization of nested logit and show how the nesting structure can be recovered from individual-level stochastic choice data. Our paper instead focuses on estimating the nesting structure in empirical demand models using market-level data, where researchers observe shares, prices, and product characteristics and the number of products can be large.

Second, this paper contributes to the literature on grouped fixed-effect estimators (Hahn and Moon, 2010; Bonhomme and Manresa, 2015). Specifically, we adapt the framework proposed by Bonhomme and Manresa (2015) that allows for group-specific time patterns and individual group membership in panel data settings. Grouped patterns of heterogeneity arise naturally as a structural feature of the nested logit model, and we re-interpret the problem of finding unknown groups/nests as uncovering the structure of unobserved heterogeneity. We extend the theoretical analysis of Bonhomme and Manresa (2015) to encompass the nested logit model, which has a non-standard feature in that group-level variables in the regression are stochastic, although after normalization they converge to a well-defined limit as the number of products grows.

The two-step approach, where in the first step the groups are recovered by k-means, and in the second step the parameters of the utility are recovered mimics Bonhomme et al. (2022). By contrast, the groups in this paper do not intend to approximate the underlying heterogeneity but rather recover the nests. We provide conditions such that the groups and their associated grouped fixed effects can be estimated consistently as the number of markets and the number of products tend to infinity. Finally, we adapt the arguments in Choi and Okui (2024) to be able to accommodate endogenous covariates which are prevalent in many economic settings of industrial organization and trade.

# 2 The Data-Driven Nested-Logit Model

In this section, we begin by reviewing the nested logit model McFadden (1978), departing from the original only by relaxing the assumption that nests are known. The following subsection discusses empirical specifications of indirect utility — including models with endogenous prices — commonly used in the literature.

We consider a model with $J$ different products and $M$ independent markets representing different geographies or time periods.[2] Agent $i$ in market $m$, receives utility from product $j$ given by:

$$U_{ijm} = \delta_{jm} + \epsilon_{ijm} \quad j \in \{1, ..., J\}, m \in \{1, ..., M\},$$

where $\delta_{jm}$ is the average utility for choice $j$ in market $m$ and $\epsilon_{ijm}$ is an idiosyncratic taste shock. The probability that agent $i$ chooses product $j$ is given by the probability that product $j$ maximizes its indirect utility:

$$\begin{aligned}\mathbb{P}_{jm} &= \mathbb{P}(U_{ijm} > U_{ij'm} \, \forall j' \neq j) \\ &= \mathbb{P}(\delta_{jm} - \delta_{j'm} > \epsilon_{ij'm} - \epsilon_{ijm} \, \forall j' \neq j).\end{aligned}$$

We assume that $(\epsilon_{ijm})_{j=1}^{J}$ is i.i.d. across $i$ and $m$, and follows a nested-logit distribution. That is, for a partition $\mathcal{K} \equiv \{k(j)\}_{j=1}^{J}$, the taste shocks $(\epsilon_{1m}, ..., \epsilon_{Jm})$ have the following cumulative distribution function:[3]

$$\exp\Big( - \sum_{k=1}^{K} ( \sum_{\{j|k(j)=k\}} e^{-\frac{\epsilon_{jm}}{\sigma^k}} )^{\sigma^k} \Big),$$

where $\sigma^k \in (0, 1]$ captures differences in degree of substitutability across products among the different nests. With this generalized extreme value (GEV) distribution, the probability

[2] While we present our model with a balanced panel for simplicity, the method extends to unbalanced panels in which products appear in different subsets of markets, as long as the number of product-market observations grows to infinity.

[3] In Section 5.4, we extend our model to the case with multiple layers of sequential nests.

of choosing product $j$ in group $k(j)$ is given by:

$$\mathbb{P}_{jm} = \frac{e^{\frac{\delta_{jm}}{\sigma^{k(j)}}} \left( \sum_{\{j'|k(j')=k\}} e^{\frac{\delta_{j'm}}{\sigma^{k(j)}}} \right)^{\sigma^{k(j)}-1}}{\sum_{l=1}^{K} \left( \sum_{\{j'|k(j')=l\}} e^{\frac{\delta_{j'm}}{\sigma^{l}}} \right)^{\sigma^{l}}}. \tag{1}$$

To understand substitution patterns, it is useful to present the expressions of price elasticities:

$$\mathcal{E}^{j}_{j',m} = \begin{cases} \frac{1}{\sigma^{k(j)}} \left(1 - \sigma^{k(j)} \mathbb{P}_{jm} - (1 - \sigma^{k(j)}) \mathbb{P}_{j|k(j),m}\right) \frac{\partial \delta_{jm}}{\partial p_{jm}} p_{jm} & \text{if } j' = j, \\ - \mathbb{P}_{j'm} \frac{\partial \delta_{j'm}}{\partial p_{j'm}} p_{j'm} & \text{if } k(j') \neq k(j), \\ \left( (\sigma^{k(j)} - 1) \frac{\mathbb{P}_{j'|k(j),m}}{\sigma^{k(j)}} - \mathbb{P}_{j'm} \right) \frac{\partial \delta_{j'm}}{\partial p_{j'm}} p_{j'm} & \text{if } k(j') = k(j). \end{cases} \tag{2}$$

From the expressions above, it is easy to see that when $\sigma^k \to 0$, products become more elastic and there is more substitution within nests. By contrast, when $\sigma^k = 1$, we are back to the standard logit model. It immediately follows that $\mathcal{E}^{j}_{j',m} \geq \mathcal{E}^{j}_{j'',m}$ if with $k(j) = k(j')$ and $k(j) \neq k(j'')$: products within the same nest are closer substitutes than products across nests.[4]

Empirical applications of the NL model assume that the nesting structure, $\{k(j)\}_{j=1}^{J}$, is known to the researcher, yet this assumption is debatable. Nest misspecification biases preference estimates and attenuates $\sigma^k$ through measurement error in conditional choice probabilities (Fosgerau et al., 2019). In practice, the absence of prior knowledge about the population nesting structure often leads researchers to try different candidate groupings, many of which yield structural parameters that violate utility maximization, with $\sigma^k \notin (0, 1]$ for some $k$. Researchers therefore frequently search across alternative nesting structures until they find a specification that delivers both plausible nests and admissible parameter estimates. Importantly, to the best of our knowledge, there is no statistical justification for this iterative search procedure. We depart from this approach by treating the nesting structure as unknown and proposing an estimator that consistently recovers both the groups, $\{k(j)\}_{j=1}^{J}$,

[4] For a full derivation of the substitution patterns see Appendix A.1.

and the structural preference parameters.

## 2.1 Empirical Model

We now proceed to describe the different components of our model.

### 2.1.1 Utility

We assume that average utility component $\delta_{jm}$ is *linear* and given by

$$\delta_{jm} = \beta_p \log p_{jm} + \beta_x \mathbf{x}_{jm} + \xi_{jm}, \tag{3}$$

where $p_{jm}$ denotes prices, $\mathbf{x}_{jm}$ is a vector of exogenous product characteristics that may include higher-order terms and interactions, and $\xi_{jm}$ represents unobserved demand shocks.

### 2.1.2 Prices

We assume that $\mathbf{x}_{jm}$ is orthogonal to $\xi_{jm}$, while allowing prices to be potentially correlated with $\xi_{jm}$. In such case, we assume the existence of an exogenous price shifter $z_{jm}$:

$$\mathbb{E}[\xi_{jm}|\mathbf{x}_{jm}, z_{jm}] = 0.$$

However, imposing these moment conditions alone is not sufficient, since groups are also unknown in our setting. See (Choi and Okui, 2024). We therefore proceed by projecting prices on exogenous covariates and estimating the demand nesting structure jointly via 2SLS.

We start by presenting the structural pricing equation, which takes the general form:

$$p_{jm} = m_i(z_{jm}, \mathbf{x}_{jm}, \eta_{jm}), \tag{4}$$

where $m(\cdot)$ is an unknown function that may be nonlinear in its arguments and $\eta_{jm}$ is an unobserved price component potentially correlated with $\xi_{jm}$.

We consider a linear projection of prices on exogenous covariates and grouping structure $k'$:

$$\log p_{jm} = \gamma_x^{k'(j)\prime} \mathbf{x}_{jm} + \gamma_z^{k'(j)\prime} z_{jm} + \kappa_m^{k'(j)} + \nu_{jm}, \tag{5}$$

where $\nu_{jm}$ is a projection error, orthogonal to $\mathbf{x}_{jm}$, $\mathbf{z}_{jm}$, and $\kappa_m^{k'(j)}$ by construction.

The inclusion of $k'(j)$ in (5) is motivated by two reasons.[5] First, failure to account for a pricing grouping structure distinct from demand, $k(j) \neq k'(j)$ for some $j$, has direct consequences for the second stage: the residuals entering the demand classification step will retain latent group-specific variation different from $k(j)$, rendering the demand group estimates inconsistent.

Second, pricing grouping structures arise naturally and are a feature of many standard pricing models. For example, let's consider Nash-Bertrand competition with single-product firms, where the first-order condition gives:

$$\log p_{jm} = \log mc_{jm} + \mu_{jm}, \qquad \mu_{jm} = -\log\left(1 - \frac{\sigma^{k(j)}}{\beta_p\left(1 - \sigma^{k(j)}\,\mathbb{P}_{jm} - (1-\sigma^{k(j)})\,\mathbb{P}_{j|k(j),m}\right)}\right), \tag{6}$$

where $mc_{jm}$ is marginal cost and $\mu_{jm}$ is the markup under nested logit demand. The markup depends nonlinearly on the within-nest share $\mathbb{P}_{j|k(j),m}$, and therefore inherits the nest structure $k(j)$: products in the same nest face systematically different competitive conditions than products in other nests. Moreover, under our asymptotic experiment when $J \to \infty$, the markup $\mu_{jm}$ converges to a group-specific fixed effect $\kappa^{k(j)} \equiv \log(1 - \frac{\sigma^{k(j)}}{\beta_p})$ as $\mathbb{P}_{jm}, \mathbb{P}_{j|k(j),m} \to 0$. Beyond markups, marginal cost may also have its own heterogeneity, such as $mc_{jm} = \gamma_x^{mc,c(j)\prime} x_{jm} + \gamma_z^{mc,c(j)\prime} z_{jm} + \kappa_m^{mc,c(j)} + \omega_{jm}$. In both cases, (6) inherits a grouping structure in prices. The group-varying coefficients $\gamma_x^{k'(j)}$ and $\gamma_z^{k'(j)}$ and the group-market fixed effects $\kappa_m^{k'(j)}$ in (5) are designed to absorb this structure. Importantly, $k'(j)$ need not coincide with the demand-side grouping $k(j)$: products may present heterogeneity differently on the supply side than on the demand side, and we treat the two as distinct unknown objects to be estimated separately. For details of how other common pricing models map to equation (5), see Appendix A.2.

[5] In models without a price grouping structure, one could instead estimate the price projection nonparametrically in $(z_{jm}, x_{jm})$, which imposes no functional form on $m(\cdot)$.

### 2.1.3 Toward a regression equation.

Taking logs in equation (1) we obtain:

$$\begin{aligned}\log \mathbb{P}_{jm} &= \frac{\delta_{jm}}{\sigma^{k(j)}} + (\sigma^{k(j)} - 1)\log\Big(\sum_{\{j|k(j)=k\}} e^{\frac{\delta_{j'm}}{\sigma^{k(j)}}}\Big) - \log\Big(\sum_{l=1}^{K}\Big(\sum_{\{j'|k(j')=l\}} e^{\frac{\delta_{j'm}}{\sigma^{l}}}\Big)^{\sigma^l}\Big) \\ &= \frac{\delta_{jm}}{\sigma^{k(j)}} + \lambda_m^{k(j)}, \end{aligned} \tag{7}$$

where

$$\lambda_m^k \equiv (\sigma^{k(j)} - 1)\log\Big(\sum_{\{j|k(j)=k\}} e^{\frac{\delta_{j'm}}{\sigma^{k(j)}}}\Big) - \log\Big(\sum_{l=1}^{K}\Big(\sum_{\{j'|k(j')=l\}} e^{\frac{\delta_{j'm}}{\sigma^{l}}}\Big)^{\sigma^l}\Big). \tag{8}$$

Substituting the expression for $\delta_{jm}$ in equation (7):

$$\log \mathbb{P}_{jm} = \beta_p^{k(j)} \log p_{jm} + \beta_x^{k(j)} \mathbf{x}_{jm} + \lambda_m^{k(j)} + \tilde{\xi}_{jm}, \tag{9}$$

where $\beta_p^{k(j)} \equiv \frac{\beta_p}{\sigma^{k(j)}}$, $\beta_x^{k(j)} \equiv \frac{\beta_x}{\sigma^{k(j)}}$, and $\tilde{\xi}_{jm} \equiv \frac{\xi_{jm}}{\sigma^{k(j)}}$. Observe that $\beta^{k(j)} = (\beta_p^{k(j)}, \beta_x^{k(j)})$ in (9) is common across all products in the same nest and market. Therefore, $\beta^{k(j)}$ and $\lambda_m^{k(j)}$ are group-specific slope and intercepts, where group membership $k(j)$ is unknown. Equation (9) is our main estimation equation for the estimation of the nesting structure.

## 2.2 Intuition on identification.

From the previous equation, the nesting structure implies that log choice probabilities exhibit group-specific slopes, $\beta_p^{k(j)}$ and $\beta_x^{k(j)}$, as well as group-specific intercepts $\lambda_m^{k(j)}$. These coefficients are common across products within the same group but differ across groups. For expositional purposes, assume there is no endogeneity in prices. Averaging across markets and using that the idiosyncratic component $\xi_{jm}$ averages out as the number of markets grows large yields the approximate relationship

$$\log P_j \approx \beta_x^{k(j)} \bar{x}_j + \beta_p^{k(j)} p_j + \lambda^{k(j)}. \tag{10}$$

This expression implies that products belonging to the same nest share a common demand curve. Consequently, the relationship between prices and log market shares can be described by $K$ linear relationships, one for each nest. When nests differ in their substi-

tution parameters $\sigma^k$, the slopes $(\beta_x^k, \beta_p^k)$ differ across groups and the data align along lines with different slopes. Even when slopes coincide, groups may still be separated through differences in intercepts $\lambda^k$, as differences in prices and characteristics generates differences in $\lambda^k$, which can be seen from equation (8).[6] Thus, we can leverage the differences in slopes and intercepts across nests to identify a partition of products into nests. In our setting, clusters correspond to *demand curves*, so the relevant objects are lines rather than points. Products are assigned to the group whose demand curve best fits their observed relationship between prices and market shares.

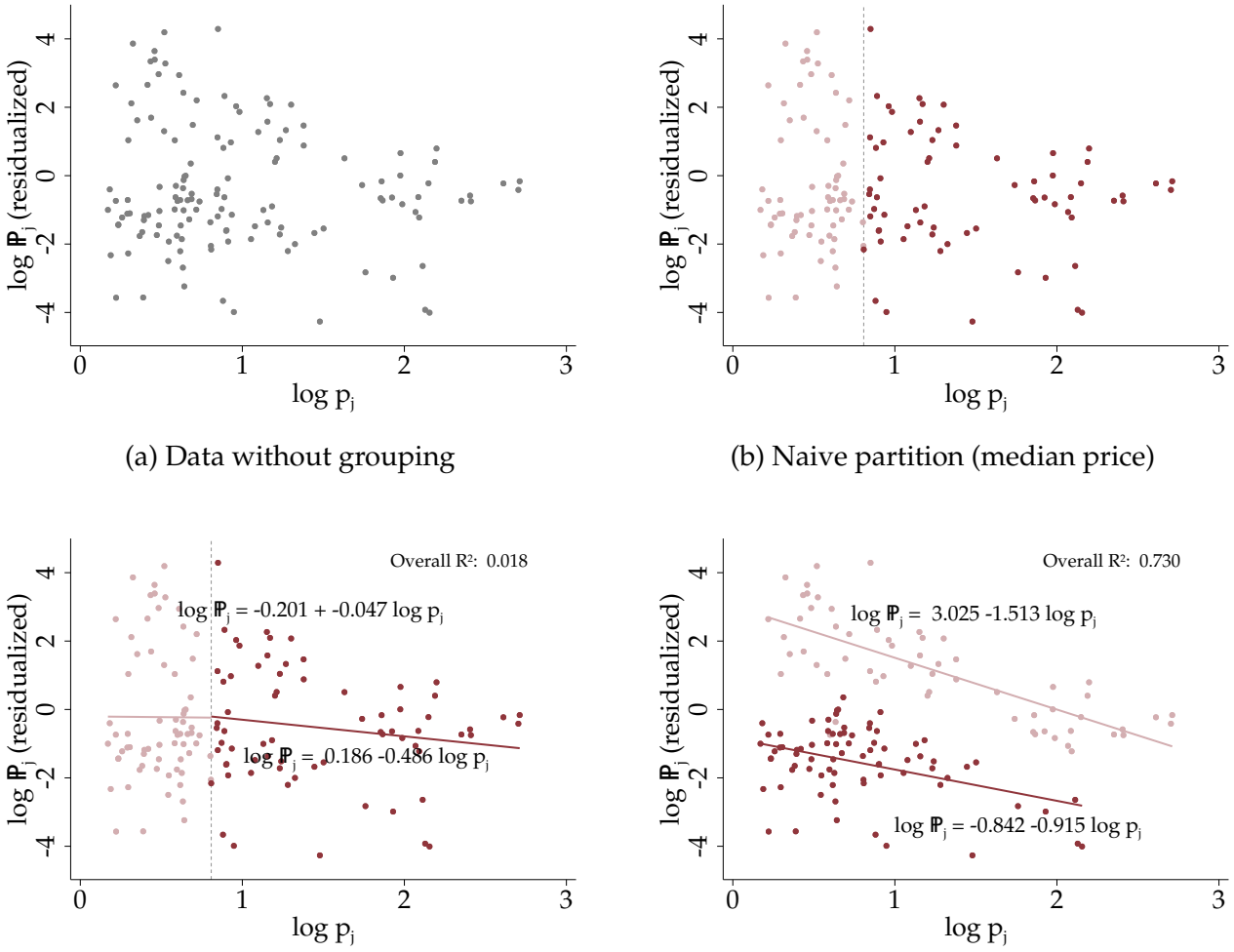
Figure 1 illustrates this idea for the case $K = 2$. Each point represents a product plotted in the space defined by price $p_j$ and residualized log market shares $\log P_j$. The first panel shows the cloud of observations without imposing any grouping structure. The second panel shows a naive candidate partition of products into two groups obtained by splitting the data according to the median price, which one might interpret as defining a high-price nest and a low-price nest. When we partition the data in this way, we find a very small $R^2$ after fitting two regression lines, indicating that this proposed nesting structure has little explanatory power of demand given prices and characteristics. The final panel shows the partition that minimizes within-sample error achieving an $R^2 = 0.73$ and overlays the corresponding demand curves.

[6] If $\sigma^k = \sigma^{k'}$ and $\lambda_m^k = \lambda_m^{k'}$ for all $m$, we cannot separate products across different nests across different demand curves. In this case, we would not be able to identify groups. However, this case corresponds to a very particular data-generating process in which two groups contain observationally identical products across all markets:

$$\forall j \in l, \exists j' \in \ell \text{ such that } \quad p_{jm} = p_{j'm},\, \mathbf{x}_{jm} = \mathbf{x}_{j'm}\ \forall m, \quad \text{for } l, \ell = k, k'.$$

Given that it is already extremely rare to observe two products with identical prices across markets, we believe this case is empirically unlikely.

Figure 1: Graphical representation of group identification




(a) Data without grouping

(b) Naive partition (median price)

(c) Naive partition demand curves

(d) Estimated demand curves

*Notes:* Each point represents a product plotted in the space defined by price and residualized log market shares. The top left first shows the raw cloud of observations. The top right panel shows a naive partition obtained by splitting products at the median price, corresponding to a high-price and a low-price nest. The bottom left panel corresponds to the best linear fit following from such price partition. The bottom right panel shows the partition and linear fit that minimizes the objective function.

## 2.3 Model Discussion

We now discuss the assumptions and limitations of our model.

Our asymptotic analysis considers a sequence of nested logit models in which both the number of products, $J$, and the number of markets, $M$, grow, but with a fixed number of nests $K$.[7] The corresponding asymptotic experiment is described in Section 4.

[7] Section 5.1 describes how $K$ can be selected via cross-validation.

First, A large product space is required because the $\lambda_m^k$ are random variables whose appropriately rescaled limits are well defined only as $J \to \infty$, by the law of large numbers. While one could instead hold $J$ fixed, doing so would require restricting product characteristics and instruments to have bounded support. We allow $J$ to grow both to avoid such restrictions and because recovering the nesting structure becomes intrinsically harder as $J$ increases as the number of admissible partitions grows exponentially in $J$. Large product spaces are also empirically relevant, as large choice sets are increasingly common in modern applications. In industrial organization, many canonical settings already feature $J \geqslant 100$: Berry et al. (1995) consider roughly 150 automobile models, Pinkse and Slade (2004) work with over 400 beer brands, Wollmann (2018) studies hundreds of commercial vehicle models, and Eizenberg (2014) examines a PC market with hundreds of hardware configurations. Scanner-data applications routinely involve even larger product spaces, with hundreds of products within narrowly defined categories across many markets (e.g., Bronnenberg et al., 2012; Dubé et al., 2010; Griffith et al., 2018; Miravete et al., 2018; Hendel and Nevo, 2006), and a growing methodological literature is explicitly designed for such high-dimensional settings (e.g., Armstrong, 2016). Large choice sets also arise naturally in urban and spatial economics, where products correspond to locations (e.g., Bayer et al., 2007; Ahlfeldt et al., 2015; Diamond et al., 2019), and more broadly in trade and spatial equilibrium models, where locations play a dual role as both products and markets (e.g., Eaton and Kortum, 2002; Allen and Arkolakis, 2022; Donaldson, 2018).

Second, large-$M$ asymptotics are required because, in the presence of unobserved components of $\delta_{jm}$ that vary across products and markets, correct classification relies on idiosyncratic noise averaging out at the product level as $M \to \infty$.

Third, no theoretical framework has yet established the consistency of clustering methods in nonlinear models, which restricts the model to utility $\delta_{jm}$ that is linear in characteristics. While this restriction prevents us from exploiting the full structure of the model in the classification step, it is helpful for several reasons. Firstly, 2SLS pricing projections are valid only when $\delta_{jm}$ is linear in parameters, as nonlinearities would introduce correlations between the pricing projection residuals and the structural error that do not vanish in expectation.[8] Secondly, departing from linearity forfeits closed-form estimators, potentially

[8] For instance, if $\delta_{jm} = \exp(\beta_p p_{jm} + \beta_x x_{jm}) + \xi_{jm}$, Jensen's inequality implies that $\mathbb{E}[\exp(\beta_p \hat{p}_{jm} + \beta_x x_{jm})\xi_{jm}]$ does not converge to zero.

increasing computational costs by orders of magnitude.

Finally, one might ask whether joint estimation of the model parameters and nesting structure via GMM is feasible. The answer is no: Choi and Okui (2024) establish that no GMM objective function can simultaneously identify both objects, since multiple combinations of parameters and nesting structures can set the same moments to zero. Together, these constraints confine us to the class of two-step clustering estimators. In the first step, we use excluded shifters to construct a price projection orthogonal to unobservable demand shocks. In the second step, we cluster on exogenous covariates and the price projection to recover the nesting structure.

# 3 Estimation

Our estimation follows a two-step procedure. In the first step, we elicit the nesting structure. In the second step, we recover the parameters conditional on the classification of the first step. In what follows, we denote by $\beta \equiv (\beta^1, \ldots, \beta^K)$ and $\lambda \equiv (\lambda_1^1, \ldots, \lambda_M^K)$.

## 3.1 First step: Classification

We use k-means to estimate the nesting structure. We consider two cases depending on the pricing projection structure. The first case assumes that the group structure of the pricing equation coincides with the GEV distribution of idiosyncratic taste shocks ($k' = k$) or that the pricing equation has no underlying group structure ($|\text{supp}(k')| = 1$). The second case considers prices with non-trivial group structure that differs from that of the taste shocks ($|\text{supp}(k')| > 1$ and $k \neq k'$). The latter case, while more robust, is computationally more demanding. In what follows, we describe both approaches.

### 3.1.1 Reduced-form Approach

When $k = k'$ or $\text{supp}(k') = 1$, that is when groupings in the demand equation and pricing equation coincide, we can substitute the pricing equation (5) into the structural market share equation (9) to obtain the following reduced-form equation:

$$\log \mathbb{P}_{jm} = \theta_z^{k(j)} z_{jm} + \theta_x^{k(j)} \mathbf{x}_{jm} + \alpha_m^{k(j)} + \nu_{jm}, \tag{11}$$

where $\theta_z^{k(j)} \equiv \beta_p^{k(j)}\gamma_z^{k(j)}$, $\theta_x^{k(j)} \equiv \beta_p^{k(j)}\gamma_x^{k(j)} + \tilde{\beta}_x^{k(j)}$, $\alpha_m^{k(j)} \equiv \beta_p^{k(j)}\kappa_m^{k(j)} + \lambda_m^{k(j)}$, and $\nu_{jm} \equiv \xi_{jm} + \beta_p^{k(j)}v_{jm}$.

In such case, we propose the following clustering problem:

$$\big(\hat{k}(1),\ldots,\hat{k}(J)\big) = \underset{k(1),\ldots,k(J)}{\operatorname{argmin}} \left\{ \min_{\boldsymbol{\theta},\boldsymbol{\alpha}} \sum_{j=1}^{J}\sum_{m=1}^{M} \big(\log \mathbb{P}_{jm} - \theta_z^{k(j)} z_{jm} - \theta_x^{k(j)}\mathbf{x}_{jm} - \alpha_m^{k(j)}\big)^2 \right\}. \quad (12)$$

Note that this step is only useful to obtain a classification, we treat $\boldsymbol{\theta}$ as nuisance parameters. Utility parameters, such as $\beta = (\beta_p, \beta_x)$ and $\sigma = (\sigma_1, \ldots, \sigma_K)$ will be recovered in the second step.

### 3.1.2 Two-stage K-means Approach

When $k \neq k'$ and $\text{supp}(k') > 1$, we follow Choi and Okui (2024) and consider a two-stage strategy to obtain the classification. That is, we first construct a prediction of prices, using demand shifters, to purge the endogenous part of prices, and then perform classification based on a second stage equation. That is, we solve the following classification problem:

$$\big(\hat{k}(1),\ldots,\hat{k}(J)\big) = \underset{k(1),\ldots,k(J)}{\operatorname{argmin}} \left\{ \min_{\tilde{\beta},\lambda} \sum_{j=1}^{J}\sum_{m=1}^{M} \big(\log \mathbb{P}_{jm} - \tilde{\beta}_p^{k(j)}\hat{p}_{jm} - \tilde{\beta}_x^{k(j)}\mathbf{x}_{jm} - \lambda_m^{k(j)}\big)^2 \right\}, \quad (13)$$

where $\hat{p}_{jm} = \hat{\gamma}_x^{\hat{k}'(j)}\mathbf{x}_{jm} + \hat{\gamma}_z^{\hat{k}'(j)} z_{jm} + \hat{\kappa}_m^{\hat{k}'(j)}$ and

$$\big(\hat{k}'(1),\ldots,\hat{k}'(J)\big) = \underset{k'(1),\ldots,k'(J)}{\operatorname{argmin}} \left\{ \min_{\gamma,\kappa} \sum_{j=1}^{J}\sum_{m=1}^{M} \big(p_{jm} - \gamma_x^{k'(j)}\mathbf{x}_{jm} + \gamma_z^{k'(j)} z_{jm} + \kappa_m^{k'(j)}\big)^2 \right\} \quad (14)$$

In Appendix C.1 we outline the algorithms to solve problems (12), (13), and (14).

## 3.2 Second Step: Structural Parameters

This second step conditions on the estimated classification from the first step and applies the method from Berry (1994). It follows an instrumental variable strategy. More precisely, using their inversion, log shares can be written as follows:

$$\log \mathbb{P}_{j,m} - \log \mathbb{P}_{0,m} = \beta_p p_{jm} + \beta_x x_{jm} + (1 - \sigma^{k(j)}) \log \mathbb{P}_{j|k(j),m} + \xi_{jm}, \tag{15}$$

where $k(j) \in \{1, \ldots, K\}$ indexes the group to which product $j$ belongs to. The regressor $\log \mathbb{P}_{j|k(j)}$ is unobserved, but when nests are known it is customary to construct it by computing market shares normalized by the sum of market shares in each group. Because now nests have been estimated in the first step, we construct this regressor as follows:

$$\log \mathbb{P}_{j|\hat{k}(j),m} = \log \left( \frac{\mathbb{P}_{jm}}{\sum_{j \in \hat{k}(j)} \mathbb{P}_{jm}} \right).$$

This regressor introduces a simultaneity problem in equation (15), as $\mathbb{P}_{jm}$ shows up in both the left hand side and the right hand side of the equation. To overcome this endogeneity problem, we use an instrumental variable approach following Berry (1994). From the structural expression of $\log \mathbb{P}_{j|k(j),m} = \frac{\exp(\beta_p p_{jm} + \beta_x x_{jm} + \xi_{jm})}{\sum_{j \in k(j)} \exp(\beta_p p_{jm} + \beta_x x_{jm} + \xi_{jm})}$, several natural instruments arise such as number of products in nest $k(j)$, average characteristics other products within group $k(j)$ (BLP-type instruments, Berry et al. (1995)), deviations in product characteristics relative to product $j$ within group $k(j)$ (Gandhi-Houde-type instruments, Gandhi and Houde (2019)), or the prediction of $\mathbb{P}_{j|k(j),m}$ just using exogenous characteristics (exogenous projections):

$$\frac{\exp(x_{jm})}{\sum_{j' \in k(j)} \exp(x_{j'm})}.$$

Let's denote the choice of our instrument by $w_{jm}$. Putting everything together, we can estimate equation (15) via two-stage least squares or via GMM using the following moment conditions:

$$\mathrm{E}\left[ \begin{bmatrix} x_{jm} \\ z_{jm} \\ w_{jm} \end{bmatrix} \cdot \xi_{jm}(\hat{k}) \right] = 0.$$

where $\xi_{jm}(\hat{k}) = \log \mathbb{P}_{j,m} - \log \mathbb{P}_{0,m} - \beta_p p_{jm} + \beta_x x_{jm} + (1 - \sigma^{k(j)}) \log \mathbb{P}_{j|\hat{k}(j),m}$. We call $\hat{\beta}$ the resulting estimator.

# 4 Statistical Properties

This section informally describes how we establish consistency and asymptotic normality of our estimator as $J$ and $M$ grow large. While our analysis is heavily based on the work of Bonhomme and Manresa (2015) (BM15, henceforth), several aspects of this setup do not make it a mere application of their results. In particular, $\mathbb{P}_{jm}$ goes to 0 as $J$ grows large, and at the same time $\lambda_m^{k(j)}$ is random and might drift to $-\infty$ as the number of products grow, hence making the nested logit model out of scope of the data generating processes (DGP) considered by BM15. The details of the theory can be found in Appendix B.

The main result in the paper is the characterization of the asymptotic distribution of the preference parameters estimator, $\hat{\beta}$, as being *asymptotically equivalent* to the one of an oracle estimator in which the nesting structure is *known*. That is, our estimator will have the same asymptotic distribution as the one in which the true structure of nests is known. As a consequence, the computation of the standard errors in $\hat{\beta}$ will not have to be adjusted by the classification. The main step in order to prove this result is to show that asymptotically, as the number of products and the number of markets grows, the probability of misclassification of any product goes to 0 sufficiently fast.[9] We next provide an intuition and a brief description of a few conditions that guarantee this result.

## 4.1 Consistent Classification

Let us start by rescaling our model by $\log J$ in order to work with fixed quantities, and not shares that are going to 0. For simplicity, let us also assume that the coefficients $\sigma_1, \ldots, \sigma_K$ are all the same, even though in our theory this is not necessary. We obtain:

$$\log J\mathbb{P}_{jm} = x_{jm}^\top \beta + \underbrace{\lambda_m^{k(j)} + \log J}_{\tilde{\lambda}_m^{k(j)}} + \nu_{jm}, \tag{16}$$

[9] It goes to zero faster than any polynomial rate, hence the first order asymptotic approximation is not affected by it.

where $\tilde{\lambda}_m^{k(j)} = \log\left(\frac{1}{J}\sum_{\{j|k(j)=k\}} e^{\frac{\delta_{jm}}{\sigma^k}}\right)$. Provided regularity conditions, the law of large numbers would apply to $\tilde{\lambda}_m^{k(j)}$ and the following decomposition would be granted:

$$\tilde{\lambda}_m^{k(j)} = \zeta_{k(j)m} + e_{k(j)m},$$

where $e_{km} \to 0$ as $J \to \infty$ for all $k$, since it is the difference between the log of the expectation and the sample mean. The reduce-form equation could be written as:

$$\log J\mathbb{P}_{jm} = x_{jm}^\top \beta + \zeta_{k(j)m} + \nu_{jm} + e_{k(j)m}$$

where now there are two sources of error: $\nu_{jm}$ and $e_{k(j)m}$.

Assuming $\beta$ is known, and assuming there are only two nests, we study the probability of misclassification of one product into the wrong nest. Let us also assume that the two nest-specific parameters are sufficiently separated. That is: $\frac{1}{M}\sum_{m=1}^M(\zeta_{1m} - \zeta_{2m})^2 > C$, where $C > 0$. Given the structure of the estimator, misclassification of a product can be characterized as the probability that the product specific residuals when assigned to one group is smaller than when the product is assigned to the other group:

$$\mathbb{P}\left(\hat{k}(j) = 2|k(j) = 1\right) = \mathbb{P}\left(\sum_{m=1}^M(\log J\mathbb{P}_{jm} - x_{jm}^\top\beta - \tilde{\lambda}_m^{k(2)})^2 < \sum_{m=1}^M(\log J\mathbb{P}_{jm} - x_{jm}^\top\beta - \tilde{\lambda}_m^{k(1)})^2\right)$$

Now, this probability can be bounded by three terms:

$$\begin{aligned}\mathbb{P}\left(\hat{k}(j) = 2|k(j) = 1\right) \leqslant & \underbrace{\mathbb{P}\left(2\sum_{m=1}^M \nu_{jm}(\zeta_{1m} - \zeta_{2m}) < \frac{1}{3}\sum_{m=1}^M(\zeta_{1m} - \zeta_{2m})^2\right)}_{(1)} + \\ & + \underbrace{\mathbb{P}\left(2\sum_{m=1}^M(\zeta_{1m} - \zeta_{2m})(e_{1m} - e_{2m}) < \frac{1}{3}\sum_{m=1}^M(\zeta_{1m} - \zeta_{2m})^2\right)}_{(2)} + \\ & + \underbrace{\mathbb{P}\left(\sum_{m=1}^M(e_{1m} - e_{2m})^2 < \frac{1}{3}\sum_{m=1}^M(\zeta_{1m} - \zeta_{2m})^2\right)}_{(3)}.\end{aligned}$$

If we are able to show that each of these terms is going to 0 we would have established that misclassification goes to 0. Let us start with term (3): the left hand side of the inequality, can be bounded by the maximum of $e^2_{k(j)m}$ over all $m$ and $k(j) \in \{1, 2\}$. This quantity goes to 0 as the number of products grows. Since $\frac{1}{M}\sum_{m=1}^{M}(\zeta_{1m} - \zeta_{2m})^2$ is bounded away from 0, (3) tends to 0 when $J \to \infty$. The argument for term (2) is similar to that of (3). Finally, term (1) holds a more nuanced intuition: it requires that the covariance between the error term, $\nu_{jm}$, and the grouping structure goes to 0. In particular, if the error term has limited cross-sectional dependence, this implies absence of correlation with any grouping structure. This sufficient condition justifies working with a price projection that ensures that the residual is not only uncorrelated with the endogenous variable, but in addition has limited cross-sectional dependence. Controlling for group heterogeneity in the price projection equation makes this condition more plausible.

From here, we adapt the proofs of BM15 to establish consistent estimation of groups with minor modifications to accommodate the uniform $o_P(1)$ error terms. The details are given in Appendix section B where we prove the essential asymptotic results in BM15 with this change.

**Theorem 4.1** (misclassification)**.** *Suppose that Assumptions 1, 2 and 3 hold (see Appendix section B). Then, as $J, M \to \infty$ and as $\log M/J \to 0$, and for any $\delta > 0$:*

$$\mathbb{P}(\max_{j\in[J]} |\widehat{k}(j) - k(j)| > 0) = o(1) + o(JM^{-\delta}).$$

*where $\widehat{k}$ is the group estimator described in 3.2.*

This result states that products can be classified with high accuracy to their true groups, even when the number of markets is much smaller than the number of products asymptotically. This justifies the use of estimated nests in the second-step estimator, and formally leads to asymptotic equivalence between our estimator and an infeasible one where the nesting structure is observed.

We then check that the market shares equations of the reduced-form and two-stage approaches satisfy the hypothesis of theorem 4.1, hence proving consistent classification of both approaches.

**Corollary 1** (Asymptotic Normality)**.** *Suppose the conditions of theorem 4.1 are satisfied, plus*

*additional regularity conditions stated in Assumption 3 of Bonhomme and Manresa (2015) . It holds*

$$\sqrt{JM}\left(\widehat{\beta}-\beta\right)\rightarrow\mathcal{N}(0,\Omega),$$

*where $\widehat{\beta}$ is defined in the previous section with $\Omega$ being the oracle asymptotic variance of the infeasible estimator where the nesting structure is known.*

# 5 Extensions

In this section, we show how to estimate the number of groups and present a model that allows for price endogeneity in the form of product fixed effects when no exogenous price shifters are available.

## 5.1 Estimating the Number of Groups

So far, the number of groups is assumed to be known and fixed. In this section, we present how to estimate the number of groups. It is well known that choosing the optimal number of groups in a clustering algorithm is a hard problem. We propose to use a cross validation criterion on out-of-sample data. Concretely, we do cross-validation using the "Elbow Method", a widely used heuristic for choosing the right number of groups in clustering algorithms. Other indexes and criterion, such as the Silhouette method, could be applied following a similar logic, but the Elbow Method has the advantage of being computationally simple to implement.

Our proposed algorithm is as follows. First, divide products $\mathcal{P}$ into $N$ equal parts, $\mathcal{P}_1,...,\mathcal{P}_N$. Second, iterate for all $k\in\{1,...,K\}$ the following procedure:

1. For every $n=1,..,N$:

   (a) Remove $\mathcal{P}_n$ products and estimate grouping structure $\mathcal{B}_{-n}(k)$ and grouping parameters $(\beta_{-n}(k),\lambda_{-n}(k))$ in the remaining $\mathcal{P}_{-n}$ products.

   (b) Classify products $j\in\mathcal{P}_n$ using $(\beta_{-n}(k),\lambda_{-n}(k))$:

$$k^*(j)=\underset{k\in\{1,...,K\}}{\arg\min}\sum_{m=1}^{M}(\log\mathbb{P}_{jm}-\log\mathbb{P}_{0m}-\beta_{-n}^{k}x_j-\lambda_{m,-n}^{k})^2.$$

(c) Compute out-of-sample MSE given the previous classification:

$$MSE_n(k) = \frac{1}{J \cdot M} \sum_{m=1}^{M} \sum_{j \in \mathcal{P}_n} (\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} - \beta_{-n}^{k} x_{jm} - \lambda_{m,-n}^{k})^2.$$

2. Take average across $N$ folds:

$$MSE(k) = \frac{1}{N} \sum_{n=1}^{N} MSE_n(k).$$

Finally, we select $k^*$ as the point at which the elbow occurs, defined by the following criterion:

$$k^* = \arg\max_k \left\{ \Delta^2 \, MSE(k) \;\text{ or }\; d\big(k,\, \overline{k=2,\, k=\bar{K}}\big) \right\},$$

where $\Delta^2 \, MSE(k)$ denotes the second difference (i.e., the reduction in slope) of $MSE(k)$, and $d(\cdot)$ denotes the distance from point $k$ to the line connecting the endpoints $k = 2$ and $k = \bar{K}$. Intuitively, the elbow criterion identifies the number of clusters beyond which adding further groups yields diminishing reductions in $MSE(k)$, trading off fit against model complexity in a manner analogous to information criteria such as AIC or BIC in maximum likelihood estimation. The Elbow method has two main advantages. First, because it relies on out-of-sample fit, it reduces the risk of overfitting. Second, it is straightforward to compute.

In general, this heuristic — or the addition of further selection criteria — may yield multiple candidates for $k^*$. In such cases, we proceed as follows. First, we discard any candidate $k$ for which the second-step estimation produces parameters inconsistent with utility maximization, namely any estimate of $\sigma$ falling outside the interval $(0, 1)$.[10] If multiple candidates remain after this pruning step, we select the largest $k$ among them. The rationale is that overestimating the number of groups does not compromise asymptotic consistency, whereas underestimating it does (Moon and Weidner, 2015).

[10] If no candidate survives this elimination step, we report the $k^*$ associated with the lowest p-value rejecting the null hypothesis that all $\hat{\sigma}$ lie within $(0, 1)$.

## 5.2 Endogenous Product Fixed Effects

Assume that:

$$\xi_{jm} = \mu_j + \nu_{jm},$$

where $\mathrm{E}[\nu_{jm}|p_{jm}, x_{jm}] = 0$ but where there could be correlation between $\mu_j$ and $p_{jm}$. In this case, average utility is given by:

$$\delta_{jm} = \beta_p p_{jm} + \beta_x x_{jm} + \mu_j + \nu_{jm}. \tag{17}$$

We modify the estimation procedure outlined in Sections 3.1 and 3.2 by taking mean differences within products as this transformation differentiates out the time-invariant demand shocks $\xi_j$ from $\delta_{jm}$:

$$\begin{aligned}\nabla\delta_{jm} &= \delta_{jm} - \frac{1}{M}\sum_{m=1}^{M}\delta_{jm} \\ &= \beta_p \nabla p_{jm} + \beta_x \nabla x_{jm} + \nabla\nu_{jm},\end{aligned}$$

where

$$\nabla x_{jm} \equiv x_{jm} - \frac{1}{M}\sum_{m=1}^{M} x_{jm} \quad \text{and} \quad \nabla p_{jm} \equiv p_{jm} - \frac{1}{M}\sum_{m=1}^{M} p_{jm}.$$

It is easy to see from the previous equation and under our assumptions that:

$$\mathrm{E}[\nabla x_{jm}\nabla\nu_{jm}] = 0 \quad \text{and} \quad E[\nabla p_{jm}\nabla\nu_{jm}] = 0.$$

Finally define $\nabla\lambda_m^{k(j)} \equiv \lambda_m^{k(j)} - \frac{1}{M}\sum_{m=1}^{M}\lambda_m^{k(j)}$. The classification algorithm is modified as follows:

1. Let $B_0^1, \ldots, B_0^K$ be a starting group structure.

2. Compute:

$$(\beta_s, \lambda_s) = \underset{(\beta,\lambda)}{\arg\min} \sum_{j=1}^{J}\sum_{m=1}^{M}\left(\nabla \log \mathbb{P}_{jm} - \nabla \log \mathbb{P}_{0m} - \begin{bmatrix}\nabla p_{jm} \\ \nabla x_{jm}\end{bmatrix}\beta^{k(j)} - \nabla\lambda_m^{k(j)}\right)^2.$$

3. For $\theta^s$ , compute for all $j \in J$:

$$k(j)^{s+1} = \underset{k \in \{1,...,K\}}{\arg\min} \sum_{m=1}^{M} \left( \nabla \log \mathbb{P}_{jm} - \nabla \log \mathbb{P}_{0m} - \begin{bmatrix} \nabla p_{jm} \\ \nabla x_{jm} \end{bmatrix} \beta_s^k - \nabla \lambda_{s,m}^k \right)^2,$$

to recover grouping structure $\mathcal{B}^{s+1}$.

4. Repeat until convergence of parameters.

Finally, the moment conditions of the second step are modified as follows:

$$\mathrm{E}\left[ \begin{bmatrix} \nabla x_{jm} \\ \nabla p_{jm} \\ \nabla w_{jm} \end{bmatrix} \cdot \nabla \nu_{jm} \right] = 0.$$

## 5.3 Adding Consumer Heterogeneity

We can naturally extend our method to incorporate consumer heterogeneity. Let's start with the case where both consumer heterogeneity and individual choices are observable. Let's assume that consumer's heterogeneity can be described or approximated by types, $\omega^i \in \{\omega^1, ..., \omega^L\}$ where $l$ indexes type. For example, a practitioner could consider multiple income bins as a way to approximate the income distribution. In such case, let's consider a model of utility given by:

$$\delta_{jm}^l = \beta^l x_{jm} + \xi_{jm}^l.$$

Define:

$$\mathbb{P}_{jm}^l = \mathbb{P}_m(j|\omega = \omega^l).$$

Let's first consider a general model where all consumers have a different nesting structures. It follows:

$$\log \mathbb{P}_{jm}^l - \log \mathbb{P}_{0m}^l = \frac{1}{\sigma^{k^l(j)}} \left( \beta^l x_{jm} + \xi_{jm}^l \right) + (\sigma^{k^l(j)} - 1) \log IV_m^{k^l(j)}.$$

Therefore, our objective function to estimate nests becomes:

$$\min_{k(j)\in\{1,...,K\}} \min_{(\beta,\lambda)} \sum_{l=1}^{L} \pi_l \sum_{j=1}^{J} \left( \log \mathbb{P}^l_{jm} - \log \mathbb{P}^l_{0m} - \beta^{k(j),l} x_{jm} - \lambda_m^{k(j),l} \right)^2, \tag{18}$$

where $\pi_l$ determines type's $l$ weight in the objective function. It is worth noting that in this general model because both the nesting structure as well as errors are type-specific, one could solve the problem type by type and apply the algorithm outlined in Section 3.1 in parallel across types.

It is straightforward to see that equation (18) also applies in the case where the practitioner wants to impose the same nesting structure across consumer types. In such case, it is only necessary to impose such common nesting structure when solving for equation (18).

## 5.4 Identification with Multiple Levels

In our previous discussion, we assumed just one level of nests. However, NL models can have multiple upper levels of groups. In this section, we extend our discussion to a model with two levels of nests. Generalizing to more levels is straightforward.

Assume we have upstream nests given by $A_1, ..., A_N$, and downstream nests given by $B_1, ..., B_K$, where $N < K$. We denote by $k(j)$ the downstream nest to which product $j$ belongs and by $n(k)$ the upstream nest to which downstream nest $k$ belongs. In this case, choice probabilities can be written as:

$$\mathbb{P}_{jm} = \mathbb{P}_{j|k(j),m} \mathbb{P}_{k(j)|n(k),m} \mathbb{P}_{n(k),m},$$

with conditional probabilities given by:

$$\mathbb{P}_{j|k,m} = \frac{e^{\frac{\delta_{jm}}{\sigma^k}}}{\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}}$$
$$\mathbb{P}_{k|n,m} = \frac{\left(\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}\right)^{\sigma^k}}{\sum_{k\in A_n}\left(\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}\right)^{\sigma^k}}$$
$$\mathbb{P}_{n,m} = \frac{\left(\sum_{n\in A_n}\left(\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}\right)^{\sigma^k}\right)^{\sigma^n}}{\sum_k\left(\sum_{k\in A_n}\left(\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}\right)^{\sigma^k}\right)^{\sigma^n}}.$$

Taking logs and under the assumption that $\delta_{0m} = 0$ with the outside option being a singleton nest, it follows:

$$\ln \mathbb{P}_{jm} - \ln \mathbb{P}_{0m} = \frac{\delta_{jm}}{\sigma^k} + (\sigma^k - 1)\ln IV_m^k + (\sigma^n - 1)\ln IV_m^n,$$

with inclusive values defined as

$$IV_m^k = \sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}$$
$$IV_m^n = \sum_{k\in A_n}\Big(\sum_{j\in B_k} e^{\frac{\delta_{jm}}{\sigma^k}}\Big)^{\sigma^k}.$$

Assume that $\delta_{jm} = \beta x_{jm}$ and define $\beta^k \equiv \frac{\beta}{\sigma^k}$, $\lambda_m^k \equiv (\sigma^k - 1)\ln IV_m^k$, $\lambda_m^n = (\sigma^n - 1)\ln IV_m^n$, $\alpha \equiv -\ln IV_m$. Substituting,

$$\ln \mathbb{P}_{jm} - \ln \mathbb{P}_{0m} = \beta^k x_{jm} + \lambda_m^k + \mu_m^n.$$

Observe that we cannot directly estimate all $\lambda_m^k$ and $\mu_m^n$ in the previous model because of co-linearity. Therefore, we impose the restriction $\tilde{\lambda}_m^{k(n)} = 0$ for some $k \in A_n$, say $k(n)$. In such case, our model becomes:

$$\ln \mathbb{P}_{jm} - \ln \mathbb{P}_{0m} = \beta^k x_{jm} + \tilde{\lambda}_m^k + \tilde{\mu}_m^n,$$

where

$$\tilde{\mu}_m^n = \lambda_m^{k(n)} + \mu_m^n$$

and

$$\tilde{\lambda}_m^{k(n)} = \begin{cases} 0 & \text{for } k = k(n) \in A_n \\ \lambda_m^k - \lambda_m^{k(n)} & \text{for } k \in A_n, k \neq k(n). \end{cases}$$

We modify our classification algorithm of Section 3.1 as follows:

1. Let $B_0^1, \ldots, B_0^K, A_0^1, ..., A_0^N$ be a starting group structure.

2. For every $n = 1, ..., N$, pick some $k(n)$ such that $B_k \subset A_n$ and set $\tilde{\lambda}_m^{k(n)} = 0$ for all $m = 1, .., M$.

3. Compute:

$$(\beta_s, \lambda_s, \mu_s) = \underset{\beta,\lambda,\mu}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \left( \log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} - \beta^{k(j)} x_{jm} - \tilde{\lambda}_m^{k(j)} - \tilde{\mu}_m^{n(j)} \right)^2$$
$$\text{subject to } \tilde{\lambda}_m^{k(n)} = 0 \quad \forall n = 1, ..., N, \forall m = 1, ..., M.$$

4. For $(\beta_s, \lambda_s, \mu_s)$, construct $\theta_s^k = \lambda_s^k + \mu_s^n$, and compute for all $j \in J$:

$$k(j)^{s+1} = \underset{k \in \{1,...,K\}}{\arg\min} \sum_{m=1}^{M} \left( \log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} - \beta_p^k p_{jm} - \beta_x^k x_{jm} - \theta_m^k \right)^2,$$

   with $n(j)^s = n(k(j)^{s+1})$ to recover grouping structure $\mathcal{B}^{s+1}$ and $\mathcal{A}^{s+1}$.

5. Repeat until convergence of parameters.

## 5.5 Type II EV error: Nested CES Demand

Under nested CES, a representative consumer maximizes a two-level CES utility function over products $j \in \{1, \ldots, J\}$ organized into nests $B_1, \ldots, B_K$:

$$U_m = \left( \sum_{k=1}^{K} \left( \sum_{j \in B_k} (A_{jm}\, c_{jm})^{(\sigma^k - 1)/\sigma^k} \right)^{\frac{\sigma^k}{\sigma^k - 1} \cdot \frac{\sigma^0 - 1}{\sigma^0}} \right)^{\sigma^0/(\sigma^0 - 1)},$$

where $c_{jm}$ is consumption of product $j$, $A_{jm}$ is a quality shifter, $\sigma^k > 1$ is the within-nest elasticity of substitution, and $\sigma_0 \geqslant 1$ is the across-nest elasticity. An equivalent microfoundation

arises from a random utility model where idiosyncratic shocks are drawn from a nested Fréchet (Type II EV) distribution with within-nest shape parameter $\sigma^{k(j)} - 1$ and across-nest shape parameter $\sigma^0 - 1$, directly paralleling the nested logit derivation where shocks follow a GEV distribution. Standard expenditure minimization gives the market share:

$$\mathbb{P}_{jm} = \frac{(p_{jm}/A_{jm})^{1-\sigma^{k(j)}}}{\sum_{d \in B_{k(j)}} (p_{dm}/A_{dm})^{1-\sigma^{k(j)}}} \cdot \frac{\left(\sum_{d \in B_{k(j)}} (p_{dm}/A_{dm})^{1-\sigma^{k(j)}}\right)^{(1-\sigma^0)/(1-\sigma^{k(j)})}}{\sum_{l=1}^{K} \left(\sum_{d \in B_l} (p_{dm}/A_{dm})^{1-\sigma^l}\right)^{(1-\sigma^0)/(1-\sigma^l)}} \tag{19}$$

$$= (p_{jm}/A_{jm})^{1-\sigma^{k(j)}} \Lambda_m^{k(j)}, \tag{20}$$

where $\Lambda_m^{k(j)}$ collects all nest- and market-level terms. Taking logs and setting $\log A_{jm} = \beta_x x_{jm} + \xi_{jm}$:

$$\log \mathbb{P}_{jm} = \underbrace{(\sigma^{k(j)} - 1)\beta_x}_{\beta_x^{k(j)}} x_{jm} - \underbrace{(\sigma^{k(j)} - 1)}_{-\beta_p^{k(j)}} \log p_{jm} + \lambda_m^{k(j)} + \xi_{jm}, \tag{21}$$

which exactly maps to the first-step equation of our baseline model.

# 6 Monte Carlo

In this section, we assess the performance of our estimator using a Monte Carlo exercise. Across our different Monte Carlo simulations, we vary the number of products $J$ and the number of markets $M$, while fixing the rest of the fundamentals. In what follows, we describe our data generating process for fixed $J, M,$ and $K$.

## 6.1 Data Generating Process for Panel Data

First, we fix the number of groups $K = 3$ and classify products using uniform draws $k(j) \sim \mathcal{U}\{1, ..., K\}$, requiring that each group has at least twenty percent of products.[11] We fix the vector of $\sigma = (\sigma^1, ..., \sigma^K)$ to certain values in the interval $[0, 1]$ and order them in ascending order. In our main Monte Carlo exercise, we fix $\sigma = (0.2, 0.3, 0.6)$ unless otherwise specified.

[11] We need at least two products per group. The reason is that we need within market-group variance in product characteristics to identify the group-specific parameters.

The average utility $\delta_{jm}$ is given by:

$$\delta_{jm} = \beta_p p_{jm} + \beta_x x_{jm} + \xi_{jm},$$

where, $\beta_p = -1, \beta_x = 1$, $x_{jm}$ are exogenous characteristics and $p_{jm}$ are prices that can be correlated with unobservable demand shocks $\xi_{jm}$. Concretely, we assume that prices are given by $p_{jm} = mc_{jm} + \rho \cdot \xi_{jm}$, where $mc_{jm}$ is orthogonal to $\xi_{jm}$.[12] We set $\rho = 0.3$.

To construct product characteristics, first we draw the mean of characteristics $mc$ and $x$ as follows:

$$\begin{bmatrix} \mu_x^k \\ \mu_{mc}^k \end{bmatrix} \overset{i.i.d.}{\sim} \mathcal{N}\Big( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix} \Big),$$

across groups $k$.

Next, for a given vector of means $(\mu_x^k, \mu_{mc}^k)$, we draw the vector of characteristics as follows:

$$\begin{bmatrix} x_{jm} \\ mc_{jm} \\ \xi_{jm} \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \mu_x^k \\ \mu_{mc}^k \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right),$$

across $j$ and $m$.

Once we have generated $\delta_{jm}$ for all $j$ and $m$ and in combination with the vector $(\sigma^1, ..., \sigma^k)$, we proceed to construct inclusive values according to the following expressions:

$$IV_{k,m} = \Big( \sum_{d \in B_k} e^{\frac{\delta_{dm}}{\sigma^k}} \Big).$$

Finally, leveraging the closed-form solution of choice probabilities in NL models, we construct choices probabilities using equation (1).

## 6.2 Estimation

Our estimation proceeds in two steps. In the first step, we take data generated as described above in the form of

$$\{(\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}, p_{jm}, x_{jm}, z_{jm})\}_{j,m},$$

[12] Monte Carlo simulations under Nash–Bertrand pricing yield very similar results. We nevertheless focus on this specification because price endogeneity persists as $J$ grows large.

and classify products into $K$ groups according to the algorithm described in Section 3.1.

The output of this classification step is a partition of products into $K$ groups $\hat{\mathcal{B}} = \{\hat{B}_k\}_{k=1}^K$, as well as estimated parameters:

$$(\hat{\beta}_{k,p}, \hat{\beta}_{k,x}) \quad \text{and} \quad \hat{\lambda}_{k,m} \qquad \forall k, m,$$

which are non-linear functions of the structural parameters. Given $\hat{\mathcal{B}}$, in the second step we mostly follow Berry (1994). First, we construct the following conditional choice probabilities:

$$\hat{\mathbb{P}}_{j|k(j),m} = \frac{\mathbb{P}_{j,m}}{\sum_{j \in \hat{B}_k} \mathbb{P}_{j,m}}.$$

We then estimate the following linear equation:

$$\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m} = \beta_p p_{jm} + \beta_x x_{jm} + (\sigma^k - 1) \log \hat{\mathbb{P}}_{j|k(j),m} + \xi_{j,m}.$$

As mentioned earlier $\log \hat{\mathbb{P}}_{j|k(j),m}$ is endogenous because it is also function of the unobservable component $\xi_{j,m}$ by construction. To overcome this problem, we instrument $\log \hat{\mathbb{P}}_{j|k(j),m}$ using the projection of $\log \hat{\mathbb{P}}_{j|k(j),m}$ on exogenous characteristics as an instrument:

$$w_{j,m} = \log \Big( \frac{\exp(x_{jm})}{\sum_{j' \in k(j)} \exp(x_{j'm})} \Big).$$

### 6.2.1 Results

In this section we present our Monte Carlo results for various cases. We vary the number of products to be equal to $J = 100, 500, 1000$ to assess the performance of our estimator in markets of different size. We also set the number of markets equal to $M = 10, 50, 100$ to asses the properties of our estimator in short and large panels. We set the number of random initial conditions $I = 1000$. The results of these Monte Carlo simulations can be found in Table 1.

First, we observe that with as few as ten markets, the match rate between the estimated groups and the true groups is no less than 99%. Second, when $J$ and $M$ grow the precision of the estimates increases as expected. Also, conditional on some $J$, increasing the number of markets $M$ leads to higher match rates. However, this is not always the case for fixed $M$ and higher $J$. The reason is that by increasing $J$ two forces are at play that go in oppo-

Table 1: Monte Carlo Results

| | | | | | $\beta_p$ | $\beta_c$ | $\sigma^1$ | $\sigma^2$ | $\sigma^3$ |
|---|---|---|---|---|---|---|---|---|---|
| J | M | Runtime | Matched | True | -1 | 1 | 0.2 | 0.3 | 0.6 |
| 100 | 10 | 00:02 | 0.996 | Mean $\beta$ | -0.992 | 0.991 | 0.189 | 0.297 | 0.602 |
| | | | | Std $\beta$ | 0.032 | 0.033 | 0.034 | 0.024 | 0.007 |
| 100 | 50 | 00:25 | 1.0 | Mean $\beta$ | -1.001 | 0.998 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.01 | 0.011 | 0.001 | 0.002 | 0.003 |
| 100 | 100 | 01:07 | 1.0 | Mean $\beta$ | -1.0 | 1.0 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.006 | 0.007 | 0.001 | 0.001 | 0.003 |
| 500 | 10 | 00:06 | 0.995 | Mean $\beta$ | -1.0 | 0.998 | 0.199 | 0.298 | 0.6 |
| | | | | Std $\beta$ | 0.015 | 0.02 | 0.006 | 0.016 | 0.003 |
| 500 | 50 | 07:14 | 1.0 | Mean $\beta$ | -1.0 | 0.999 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.004 | 0.004 | 0.0 | 0.001 | 0.001 |
| 500 | 100 | 29:24 | 1.0 | Mean $\beta$ | -1.0 | 0.999 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.003 | 0.003 | 0.0 | 0.0 | 0.001 |
| 1000 | 10 | 00:12 | 1.0 | Mean $\beta$ | -1.0 | 0.999 | 0.199 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.007 | 0.007 | 0.003 | 0.001 | 0.002 |
| 1000 | 50 | 44:57 | 1.0 | Mean $\beta$ | -1.0 | 1.0 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.002 | 0.003 | 0.0 | 0.0 | 0.001 |
| 1000 | 100 | 2:11:14 | 1.0 | Mean $\beta$ | -1.0 | 0.999 | 0.2 | 0.3 | 0.6 |
| | | | | Std $\beta$ | 0.002 | 0.002 | 0.0 | 0.0 | 0.001 |

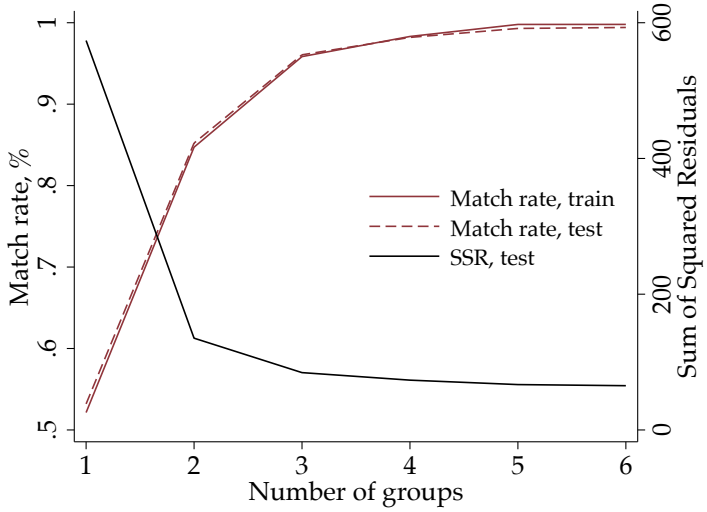
*Notes:* This table shows average Monte Carlo results when over 50 bootstrap samples. Column 1 reports the number of products $J \in \{100, 500, 1000\}$. Column 2 reports the number of market $M \in \{10, 50, 100\}$. Column 3 reports the average running time in seconds for one bootstrap sample in a 32-core machine with 16GB of RAM per core. Column 4 reports the average percentage of products that are correctly matched with their true group. The rest of the columns report the average estimates and and their standard deviations for the structural parameters.

site directions: On the one hand, as $J$ increases the estimated nesting structure is closer to the true nesting structure and thus will exhibit lower mis-classification rate. On the other hand, increasing the number of products also mechanically increases the probability of mis-classification. Overall, we take these results as evidence that our estimation procedure can correctly recover the groups when the market share data is generated using a NL model.

### 6.2.2 Estimating the Number of Groups

In this section, we show using Monte Carlo simulations how to estimate the number of groups. Our data generating process in this case also follows Section 6.1. The true number of groups is $K = 3$, the total number of products $J = 500$, and the total number of markets is $M = 10$. We do cross-validation exercise described in Section 5.1 on ten folds on data. That is, we train the model in 90% of the data and leaving out 10% of the observations for out of sample testing. The results are presented in Figure 2. We clearly observe very marginal improvements in the predictive power of the model when $K > 3$ as the SSR function becomes almost flat after $K = 4$. Therefore, our cross-validation procedure suggests that the optimal number of groups should be $K^* = 3$, which corresponds to the population number of groups with which the data was generated. We take these results as suggestive evidence that our cross-validation procedure can recover the underlying number of groups.

Figure 2: Average cross-validation results across bootstrap samples




*Notes:* The figures above show the out-of-sample sum of squared errors on 10% of the data across different number of groups. Parameters are set to be J = 500, M = 10, with true number of groups K = 3. The data generating process follows the procedure described in Section 6.1.

# 7 Empirical application: Demand for Beer in the US

In this section, we describe how to apply our estimator to the US beer market.

## 7.1 Data Description

For our main application, we use NielsenIQ store panel data for 2022. This dataset captures purchases from all selling points covered by NielsenIQ scanner data, including supercenters and supermarkets. Each entry records the purchase of a single retail product, including its unit price, the number of units sold, and a store identifier. We define products as Universal Product Codes (UPCs), which combine a retail product with its packaging type.[13] We focus on the beer category, which contains 24,188 unique UPCs. NielsenIQ also provides auxiliary datasets with information on volume size per unit (e.g., whether the product contains 330-ml or 500-ml units), package configuration (e.g., 6-pack cans or 4-pack bottles), beer type (e.g., ale or stout), packaging type (e.g., bottle, can, box), brand, and whether the beer is imported or domestic.

We consider 48 markets, defined as the continental US states. To aggregate sales within a state, we sum all units sold across selling points within that state. We construct a market-specific UPC price by taking a weighted average, where weights are the number of UPCs sold within each store.[14] Finally, because our method relies on longitudinal observations, we retain only products sold in at least ten markets, leaving a total of 2806 unique UPCs.[15] These remaining 11.6% of UPCs cover most of the sales volume, ranging from 61.5% to 99% of total units sold within a state. We combine the sales of all these infrequent UPCs into the outside option. Our final dataset is an unbalanced panel of 64497 observations.[16] Given its unbalanced nature, we adapt our algorithm to allow for the entry and exit of products across markets.

[13] For example, a single 330-ml can of a particular beer has a different UPC than a 6-pack of 330-ml cans of the same beer.

[14] Before state-level aggregation, we drop sales with prices below the 1st or above the 99th percentiles to remove price outliers and potential data entry errors. After state aggregation, we also drop products whose descriptions contain "MALTA GOYA" because these are misclassified as alcoholic beer. We also drop 23 products with missing data on their characteristics.

[15] Many UPCs are sold in very few locations. For example, 11,904 UPCs are sold in only one state and 3,923 UPCs are sold in only two states.

[16] Only 136 UPCs are sold across all 48 states. After applying our filter, the average UPC is sold in 23.32 states, implying an unbalanced panel with approximately 50% of markets missing per product on average.

Table 2 reports summary statistics for our main dataset. These statistics are weighted by sales volume. Even when aggregated at the national level, we observe substantial variation in prices and liquid content.

Table 2: Summary statistics

**Panel A: Full Sample Summary Statistics**

| Variable | Mean | Std. Dev. | 10th pct | Median | 90th pct |
|---|---|---|---|---|---|
| Price ($) | 10.618 | 6.758 | 2.596 | 10.298 | 20.308 |
| Unit quantity (units) | 8.838 | 8.229 | 1.000 | 6.000 | 24.000 |
| Price per ml ($) | 0.004 | 0.001 | 0.002 | 0.004 | 0.005 |
| Unit size (ml) | 535.627 | 450.385 | 354.882 | 354.882 | 739.338 |
| Market share | 0.859% | 1.152% | 0.061% | 0.498% | 1.967% |

**Panel B: Median Characteristics of the top ten national brands**

| Brand | Market share | Unit price/ml | Unit price | Unit quantity | Unit size |
|---|---|---|---|---|---|
| Bud Light | 12.53% | 0.003 | 8.63 | 6.00 | 473.18 |
| Miller | 7.81% | 0.003 | 12.47 | 12.00 | 354.88 |
| Modelo | 7.63% | 0.005 | 10.23 | 6.00 | 354.88 |
| Coors | 7.11% | 0.003 | 12.62 | 12.00 | 354.88 |
| Bud | 6.73% | 0.003 | 3.22 | 1.00 | 739.34 |
| Michelob | 5.96% | 0.004 | 15.20 | 12.00 | 354.88 |
| Corona | 4.59% | 0.004 | 16.52 | 12.00 | 354.88 |
| Busch | 4.43% | 0.002 | 10.87 | 12.00 | 354.88 |
| Natural | 3.63% | 0.002 | 6.49 | 6.00 | 473.18 |
| New Belgium | 1.98% | 0.005 | 10.28 | 1.00 | 567.81 |

*Notes:* Panel A summarizes key statistics of product characteristics across beers, weighting each product by its sales volume. Panel B reports the median characteristics of the top ten national brands in 2022.

## 7.2 Demand Model and Estimation

In our baseline demand specification, we include only three continuous product characteristics—log unit size, log units per UPC, and log prices—and no categorical variables. This setup can be viewed as a conservative benchmark for assessing the performance of our method in uncovering the nesting structure. In Section 7.3, we report a robustness exercise that incorporates categorical variables describing beer type in the estimation of the grouping structure.

We also include market fixed effects to capture demand shocks common to all products within a market (e.g., temperature), as well as market-level variation in the attractiveness

of the outside option, since it is not uniformly defined by the same set of products across markets. The resulting mean utility is:

$$\delta_{jm} = \lambda_m + \beta_s \log \mathrm{s}_j + \beta_n \log \mathrm{n}_j + \beta_p \log \mathrm{p}_{jm} + \xi_{jm},$$

where $s_j$ is unit size, $n_j$ is the number of units per UPC, and $p_j$ is price.

Given that the utility levels are not identified (Train, 2009), we normalize the utility of the outside option $j = 0$ to zero, namely:

$$\delta_{0m} = 0 \text{ for all } m = 1, ..., M.$$

For simplicity, we also assume that the outside option constitutes its own nest for all $m = 1, ..., M$ following many empirical IO applications. Given this assumption, we will classify only on inside goods during our classification step.

To address potential endogeneity of prices due to correlation with $\xi_{jm}$, we construct Gandhi–Houde instruments (Gandhi and Houde, 2019). Specifically, we compute the average squared differences in log unit size and log units relative to competing brands:

$$\mathrm{gh}^1_{jm} = \frac{1}{J_m} \sum_{j' \notin \text{Brand}(j,m)} (\log \mathrm{s}_j - \log \mathrm{s}_{j'})^2 \quad \text{and} \quad \mathrm{gh}^2_{jm} = \frac{1}{J_m} \sum_{j' \notin \text{Brand}(j,m)} (\log \mathrm{n}_j - \log \mathrm{n}_{j'})^2,$$

where $J_m$ is the number of products in market $m$, and $\text{Brand}(j, m)$ is the set of products in market $m$ that belong to $j's$ brand.

Although product characteristics are constant across markets, the set of available products varies. Thus, spatial variation in the assortment of products provides cross-market variation in these instruments. The relevance condition relies on the relationship between product differences and prices: products that are more vertically differentiated in their characteristics from their competitors face weaker competitive pressure and therefore result in higher equilibrium prices. The exclusion restriction requires that product characteristics of competitors, $\log \mathrm{s}_{j'}$ and $\log \mathrm{n}_{j'}$ with $j' \notin \text{Brand}(j, m)$ are not a function of unobservable demand shocks $\xi_{jm}$. This condition would hold, for example, if demand shocks are realized after product entry and design decisions are made. By contrast, it would fail if firms adjust product characteristics in response to competitors' demand shocks.

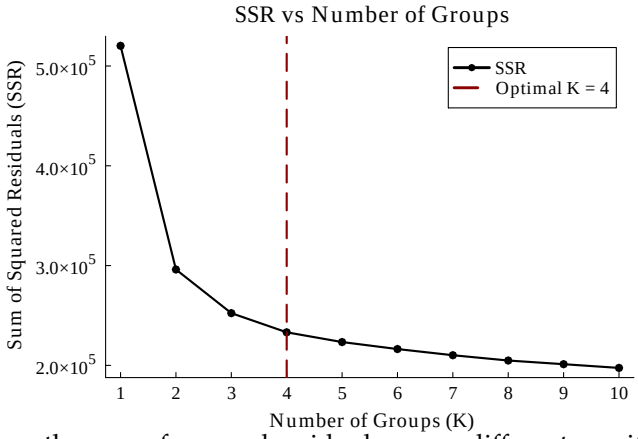
We begin by estimating the first stage. With three product characteristics and 48 market fixed effects, estimating a model with $k$ groups requires $(3 + 48) \cdot k$ parameters. Given the large parameter space, it is important to use many starting values. In general, we set the number of initial conditions to $k \cdot (M + P) \cdot 50$, where $M$ is the number of markets and $P$ is the number of covariates.[17]

### 7.2.1 First-Step Results

As part of the first stage, we begin by determining the number of groups. To do so, we follow the cross-validation procedure described in Section 5.1 and run the classification algorithm for $k = 1, \ldots, 10$. The results, shown in Figure 3, indicate that improvements in predictive performance are negligible for $k \geqslant 5$.

Applying the Elbow method yields two candidate values, $k = 3$ and $k = 4$. Both specifications produce within-nest independence parameters, $\sigma^k$, that are consistent with utility maximization. In this case, we select the largest admissible number of groups as picking more groups does not compromise asymptotic consistency (Moon and Weidner, 2015). Accordingly, our preferred specification sets $k^* = 4$.

Figure 3: Sum of squared residuals across different number of groups




*Notes:* The figure above shows the sum of squared residuals across different specifications with $k = 2, 3, \ldots, 8$. The results are estimated using Gandhi-Houde instruments. The optimal $k^*$ is the one with the largest distance to the uniform line.

Next, we proceed to described the average product characteristics conditional on every nest for $k^* = 4$. Results can be seen in Table 3. Overall, we find four differentiated groups

[17] For our baseline classification and K = 2,...,5, results become rather stable after 2,000 initial conditions.

in terms of median characteristics. Group 1 seems to correspond to cheaper average regular beer (e.g. lager). Group 2 on the other hand, represents a group of light beers with top ten brands representing 75.5% of all their UPCs, significantly above the rest of the other groups. Groups 3 and 4 contain more ales, less light beers, and less as well top-ten brands compared to the other groups. While they look similar, we can see that group 4 is more expensive, both in terms of unit price as well as price per ml.

Table 3: Characteristics of beer across groups

| Characteristics | Unconditional values | Conditional values by group 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Unit size (ml) | 354.882 | 473.176 | 354.882 | 354.882 | 354.882 |
| Unit quantity | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| Unit price | 10.209 | 9.857 | 10.156 | 12.367 | 10.509 |
| Price per ml | 0.004 | 0.003 | 0.004 | 0.004 | 0.005 |
| Share of domestic | 0.794 | 0.820 | 0.773 | 0.830 | 0.861 |
| Share of ale | 0.158 | 0.225 | 0.098 | 0.291 | 0.377 |
| Share of regular beer | 0.442 | 0.469 | 0.430 | 0.442 | 0.326 |
| Share of stout and porter | 0.015 | 0.019 | 0.011 | 0.018 | 0.046 |
| Share of light beer | 0.369 | 0.266 | 0.446 | 0.238 | 0.239 |
| Share of other types | 0.016 | 0.022 | 0.015 | 0.010 | 0.012 |
| Share top-ten brands | 0.624 | 0.460 | 0.755 | 0.382 | 0.334 |
| # of products | 2806 | 659 | 242 | 1102 | 803 |
| Label | | Regular Lager | Light Beer | Special/ Craft | Boutique Craft |

*Notes:* Rows 1–4 report median values of unit size, unit quantity, unit price, and price per ml, presented both unconditionally and conditional on group membership. Rows 5–11 report unconditional and conditional mean shares. Groups are estimated using Gandhi-Houde instruments.

Since classification is performed at the UPC level, individual brands may span multiple groups. To obtain a unique group assignment for each brand, we assign it to the group with the highest likelihood. Figure 4 displays the resulting brand classification, where the size of each logo reflects the brand's relative volume within its assigned group. Group 2 is clearly populated by mainstream light beers, including Bud Light, Michelob Ultra, Coors Light, and Miller Lite. Group 1, by contrast, contains many of the non-light counterparts of those same brands. Group 3 is more heterogeneous, combining some international lagers like Tecate with less mainstream offerings such as Founders. Group 4 consists predominantly of craft

beers from local and boutique breweries. Based on the average product characteristics and the most prominent brands within each group, we label Group 1 as *Regular Lager*, Group 2 as *Light Beer*, Group 3 as *Special/Craft*, and Group 4 as *Boutique Craft*.

As an additional validation exercise, we compute the relative popularity of each group across U.S. counties with respect to the national average. Despite the classification incorporating no spatial information, Figures 5 and 6 reveal a geographic patterns that exhibit a strong spatial autocorrelation in the popularity of these groups: *Light Beer* is disproportionately popular in the Southern states, while *Boutique Craft* is concentrated in the Northeast, Oregon, and Washington State.

### 7.2.2 Out-of-sample Validation: Bud Light Boycott

We validate the substitution patterns implied by our estimated nesting structure using the 2023 Bud Light boycott as an exogenous demand shock. In April 2023, a social media promotion of Bud Light featuring Dylan Mulvaney, a transgender actress and influencer, triggered a widespread backlash and calls for a boycott. The episode received extensive media coverage and had measurable effects on Bud Light sales, providing a plausibly exogenous negative demand shock to a single product, namely Bud Light.

If our first-step classification captures substitution patterns reasonably well, we should see more substitution toward products that share a group with Bud Light—*Light Beer*—than toward products in other nests following the boycott. To test this prediction, we augment our state-product panel with data from 2023 and estimate the following difference-in-differences specification:

$$\log \mathbb{P}_{jmt} = \kappa_{jm} + \eta_t + \sum_{k=1}^{K} \alpha_k \, \mathbf{1}\{b_{k(j)} = k\} \, \mathbf{1}\{t = 2023\} + \varepsilon_{jmt}, \tag{22}$$

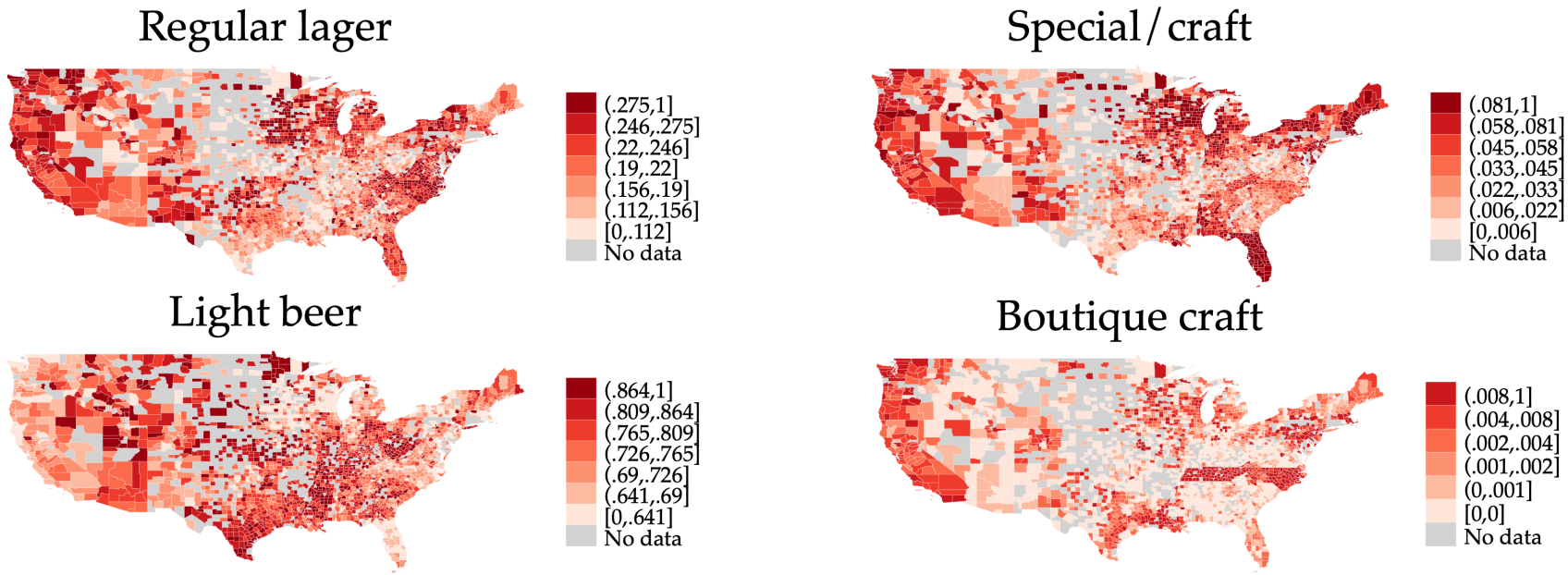
where $\kappa_{jm}$ are product-by-market fixed effects, $\eta_t$ is a year fixed effect, and the interaction terms capture differential changes in log market shares across nests between 2022 and 2023. The inclusion of product fixed effects $\kappa_{jm}$ is important: identification comes entirely from temporal variation within product and market, which was not used in the first-step classification of nests, effectively making this an out-of-sample validation of the estimated nesting structure.

Figure 4: Top 10 brands by group



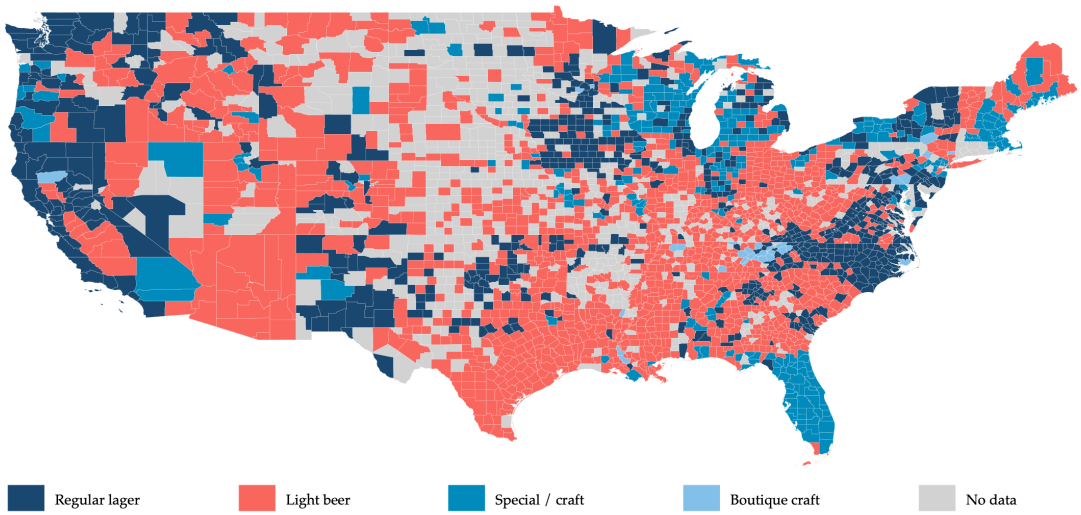

*Notes:* This figure shows the top ten most national sold beer brands at the national level for each group. The bubble size of each brand reflects the relative size of the market share in each particular group.

Figure 5: Spatial correlation of county market shares by group




*Notes:* These maps above show the market share of units sold of each beer group over the total sales at the county level.

Figure 6: Most popular group relative to national average




*Notes:* This map displays the most popular group relative to the national trend in terms of market share of units sales at the county level.

Table 4: Diff-in-Diff results of Bud Light boycott

| Dependent Variable: Log Share | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | First stage | | DiD | | | | Placebo | |
| | (1) | | (2) | | (3) | | (4) | |
| t=2023 | -0.364*** | (0.005) | -0.408*** | (0.005) | -0.449*** | (0.007) | -0.362*** | (0.010) |
| Treatment (Budlight) | -0.157*** | (0.027) | | | | | | |
| Treatment same group | | | 0.422*** | (0.015) | | | | |
| Regular beer × t=2023 | | | | | 0.173*** | (0.013) | | |
| Light beer × t=2023 | | | | | 0.463*** | (0.016) | | |
| Non-lager/craft × t=2023 | | | | | -0.001 | (0.012) | | |
| Boutique craft × t=2023 | | | | | 0.018 | (0.017) | | |
| Placebo Regular beer × t=2023 | | | | | | | -0.006 | (0.015) |
| Placebo Light beer × t=2023 | | | | | | | -0.011 | (0.015) |
| Placebo Non-lager/craft × t=2023 | | | | | | | 0.003 | (0.014) |
| Placebo Boutique craft × t=2023 | | | | | | | 0.001 | (0.014) |
| Observations | 219728 | | 214353 | | 172778 | | 172778 | |
| UPC-Market FE | YES | | YES | | YES | | YES | |
| Exclude Bud Light | No | | YES | | YES | | YES | |
| Baseline group | Non-Budlight | | Other groups | | Outside option | | Outside option | |

Notes: *** p<0.01, ** p<0.05, * p<0.1

This table reports difference-in-differences estimates of the effect of the April 2023 Bud Light boycott on log market shares, defined as the log of the ratio of units sold for a given UPC in a state to total units sold in that state. Column (1) includes all products and estimates the average effect on Bud Light relative to all other products. Columns (2)–(4) exclude Bud Light from the sample. Column (2) defines treatment as all products in the Light Beer group (excluding Bud Light) and control as products in all remaining groups; the baseline group is other data-driven groups. Column (3) estimates group-specific effects relative to the outside option. Column (4) reports a placebo test in which products are randomly assigned to five groups. All specifications include UPC-by-market fixed effects. Heteroskedasticity-robust standard errors in parentheses.

Table 4 reports the results. The first specification shows that the share of Bud Light fell significantly in 2023 relative to other beers, confirming that the boycott had a statistically significant effect. The second specification sets treated products as those who are in the *Light Beer* group other than Bud Light and control products as those in other groups as well as in the outside option, showing a positive and significant increase in the sales of the treated products relative to control groups. The third specification estimates a different treatment effects for every group, where the control group is defined as those products in the outside options. The coefficient on *Light Beer* is large and statistically significant ($\hat{\alpha}_{\text{Light}} = 0.463$), indicating that products in the same nest as Bud Light experienced a significant increase in market share following the boycott. By contrast, the coefficients for *Regular Lager* ($\hat{\alpha}_{\text{Regular}} = 0.173$), *Special/Craft* ($\hat{\alpha}_{\text{Special}} = -0.001$), and *Boutique Craft* ($\hat{\alpha}_{\text{Boutique}} = 0.018$) are substantially smaller. These results are consistent with the substitution patterns predicted by the nested logit model: following a negative shock to Bud Light, consumers substituted primarily toward other products in the *Light Beer* group rather than toward products in other nests. Finally, the last specification reports a placebo test using randomly drawn groups that confirms that our results are not driven by anything systematic in the data.

### 7.2.3 Second-Step Results

Table 5: Estimation Results for Nested Logit with $k^* = 4$

| | OLS | IV |
|---|---|---|
| log unit size | 0.070 (0.006)*** | 0.339 (0.025)*** |
| log upc quantity | 0.080 (0.005)*** | 0.441 (0.026)*** |
| log price | -0.054 (0.006)*** | -0.416 (0.033)*** |
| $\sigma_{\text{regular/lager}}$ | 0.142 (0.001)*** | 0.523 (0.014)*** |
| $\sigma_{\text{light beer}}$ | 0.343 (0.002)*** | 0.807 (0.016)*** |
| $\sigma_{\text{special/craft}}$ | -0.015 (0.001)*** | 0.336 (0.013)*** |
| $\sigma_{\text{boutique craft}}$ | -0.320 (0.001)*** | 0.042 (0.015)*** |
| Mean Own-price Elasticity | 1.351 (0.003) | -3.561 (0.013) |
| Mean Cross-price Elasticity | -0.003 (0.008) | 0.011 (0.033) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Notes:* This table reports second-step estimates of the nested logit demand model with $k^* = 4$ groups recovered in the first step. The dependent variable is the log ratio of a product's market share to the outside option share. The OLS column estimates equation (15) without instrumenting for price or the within-group share. The IV column instruments for log price using Gandhi-Houde instruments—the average squared difference in log unit size and log unit quantity relative to competing brands—and for the log within-group share using the exogenous share projection described in Section 3.2. $\sigma_k$ denotes the within-nest correlation parameter for group k, where $\sigma_k \in (0, 1)$ is consistent with utility maximization. Mean own-price and cross-price elasticities are computed by averaging across products and markets, weighted by national sales volume; standard deviations of the elasticity distribution across products are reported in parentheses. All specifications include market fixed effects. Standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

We now show the estimation results of preference parameters using $k^* = 4$ as the number of nests. We estimate preference parameters by estimating the second-stage equation:

$$\log \mathbb{P}_{j,m} - \log \mathbb{P}_{0,m} = \lambda_m + \beta_s \log \mathrm{s}_j + \beta_q \log \mathrm{q}_j + \beta_p \log \mathrm{p}_{jm} + (1 - \sigma^{k(j)}) \log \mathbb{P}_{j|k(j),m} + \xi_{jm}.$$

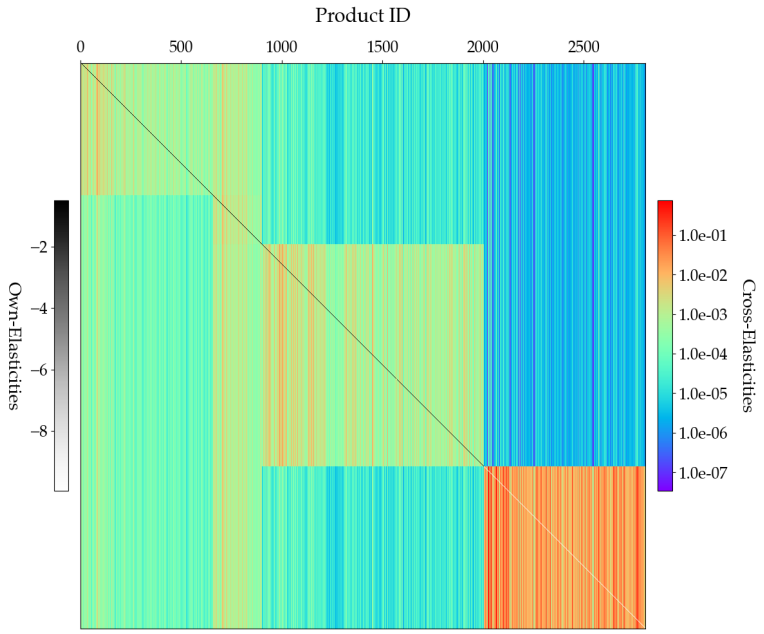
To estimate $\sigma^k$ we need to instrument $\log \mathbb{P}_{j|k(j),m}$ because those are mechanically correlated with $\xi_{jm}$. We construct the following instruments:

$$\frac{\exp(\log \mathrm{s}_j)}{\sum_{j \in k(j,m)} \exp(\log \mathrm{s}_j)} \quad \text{and} \quad \frac{\exp(\log \mathrm{n}_j)}{\sum_{j \in k(j,m)} \exp(\log \mathrm{n}_j)},$$

where $k(j,m)$ denotes the products in group $k(j)$ that are also present in market $m$. Results can be seen in Table 5. As expected, OLS yields an upward-biased price coefficient of -0.054 and some negative within-nest correlations, implying a positive average own-price elasticity. With Gandhi-Houde instruments, the price coefficient increases in magnitude to -0.416, roughly an order of magnitude larger. Notably, the estimates of within-nest correlations all take values that are consistent with utility maximization as they all lie between zero and one. These estimation results were obtained without imposing any constraints on $\sigma^k$'s. For example, we also find that *Light Beer* has the largest $\hat{\sigma}^k$ equal to 0.807 and that *Boutique Craft* has the lowest sigma equal to 0.042. These results imply that for *Light Beer* independent idiosyncratic shocks dominate choices relatively more than for *Boutique Craft*, where shocks are nearly identical and choices are mostly determined by differences in $\delta_{jm}$. Consequently, *Light Beer* is the most inelastic group and with the lowest within-group substitution as opposed to *Boutique Craft* the most elastic group and with the largest within-group substitution.

Next, we move on to analyze price elasticities. Because we have many products that are not available in all markets, we construct average elasticities by taking averages across markets. When doing so, we obtain an average own-price elasticity of -3.561, similar to previous estimates in the literature studying demand for beer (Pinkse and Slade, 2004; Slade, 2004; Romeo, 2016; Miller and Weinberg, 2017). The average cross-price elasticity is 0.011 with standard deviation 0.033. The full graph of price elasticities can be seen in Figure 8a, where we have ordered products first by group and then by UPC identifier. Following the interpretation on $\hat{\sigma}^k$, we see a lot of within-group substitution for *Boutique Craft*, in contrast to *Light Beer* that behaves close to a simple Logit model with proportional substitution patterns.

Figure 7: Price elasticity matrices of Data-Driven NL




*Notes:* The figures above display the price elasticity matrices estimated using the nested logit specification and the BLP procedure. Diagonal cells each matrix denote products' own-price elasticities, whereas the off-diagonal cells denote cross-price elasticities. Concretely, the value in row $i$, column $j$ denotes $\frac{\partial \log \mathbb{P}_i}{\partial \log p_j}$.

## 7.3 Robustness

In this section, we present a battery of robustness checks.

### 7.3.1 Hausman instruments

We also estimate the first step using Hausman instruments for prices (Hausman, 1996):

$$\mathrm{h}_{jm} = \frac{1}{N_j - 1} \sum_{m' \neq m} \log p_{jm'},$$

where $N_j$ denotes the number of markets in which product $j$ is sold. Relevance relies on cost shocks that are common across markets, such as input costs. The exclusion restriction requires that $\xi_{jm}$ be uncorrelated with prices in other markets conditional on covariates. This condition holds if demand shocks are independent across markets but may fail if, for example, marketing campaigns are targeted at specific regions.

Results are reported in Appendix C.2.1. The estimated nesting structure is broadly simi-

lar to the baseline: we again recover four groups as the optimal specification, and the conditional group characteristics closely resemble those in the baseline in terms of median product attributes. The structural parameter estimates remain consistent with utility maximization, although we fail to reject that one of the nesting parameters is statistically indistinguishable from one. The implied own-price elasticity is somewhat smaller at -1.612, falling slightly outside the range of $-6.3$ to $-2.4$ typically reported in the literature. Overall, while both instrument sets produce qualitatively similar groupings, we adopt the Gandhi-Houde specification as our baseline given the stronger empirical support for that model and the more credible exclusion restriction in this setting.

#### 7.3.2 Results with three and five groups

When changing the number of groups, we find a somewhat hierarchical structure in the grouping and we increase the number of groups. For example, when $k^* = 3$, we find that our baseline *Regular Lager* and *Special/Craft* merge into one group. Overall, we label these three groups as "Light Beer", "Regular Lager", and "Craft Beer" groups. On the other hand, when $k^* = 5$, we obtain similar groups as when $k^* = 4$, but the "Regular Beer" groups splits into two. In terms of the structural parameters, we still find $\sigma$ estimates that are consistent with utility maximization, however in both cases we cannot reject the null that they are outside the interval $(0, 1)$. For further details see Appendix C.2.2.

#### 7.3.3 Classification with additional categorical variables

We perform a robustness exercise in which we augment the baseline specification with categorical variables describing beer type provided by NielsenIQ. We consider five categories: Light Beer, Regular/Lager, Stout and Porter, Ale, and Others.[18]

Including these categorical variables, the cross-validation procedure selects $k^* = 3$ groups, one fewer than in the baseline. As shown in Table A7, the three groups correspond roughly to *Regular Lager*, *Light Beer*, and *Craft* beer. Second-step estimates are reported in Table A8 and are consistent with utility maximization, yielding an average own-price elasticity of $-2.343$, somewhat more inelastic than the baseline but within the range of estimates in the

[18] The original data contains nine categories. We consolidate four smaller categories into Others and combine two categories of regular lager into one for simplicity. Results are similar using the nine disaggregated categories.

literature.

That fewer groups emerge when categorical variables are included is intuitive. When beer type categories are excluded from the classification step, the group-by-market fixed effects $\lambda_m^{k(j)}$ must absorb not only the unobserved nest-specific demand shifters for which they are designed, but also the systematic variation in demand attributable to beer type. The classifier responds by forming additional groups, each with its own set of market fixed effects, until the residual variation is sufficiently small.

#### 7.3.4 K-means Classification in Product Characteristics

A common approach in practice is to classify products using $k$-means clustering on observed characteristics, without exploiting variation in demand across markets. We apply $k$-means to the same characteristics used in our baseline. The cross-validation procedure again selects $k^* = 4$ groups. Average product characteristics by group are reported in Table A9 and second-step estimates in Table A10.

While the number of groups coincides with our baseline, the groups are less heterogeneous both in conditional product characteristics and in implied substitution patterns. This reflects a fundamental distinction between the two approaches: $k$-means groups products by proximity in characteristics space, whereas our method classifies products by how consumers substitute among them across markets as implied by the structure of a nested-logit model. The two need not coincide, and the second-step estimates show that exploiting market share variation recovers a nesting structure that delivers more heterogeneous substitution patterns. Moreover, there is no statistical justification for clustering solely on product characteristics to guarantee the recovery of the population grouping structure. This will occurs only when the $k$-means solution coincides with the solution to problem (12). A necessary condition is that the characteristics used in the clustering procedure fully determine the substitution patterns across products. In particular, two conditions must hold. First, substitution patterns must be functions of product characteristics:

$$(\beta_x^k, \beta_p^k, \lambda_m^k) = g(x_j), \qquad \text{or} \qquad \sigma^{k(j)} = h(x_j).$$

Second, choice cannot be driven by unobserved attributes such as brand reputation or marketing campaigns, that is, $\xi_{jm} = 0$. We view these conditions as rather strong in many

empirical settings.

## 7.4 Comparison to other demand models

### 7.4.1 "Naive" Nested Logit

As a benchmark, we estimate a standard nested logit model using the ex-ante product classification provided by NielsenIQ. Specifically, we impose that products are grouped according to *beer type*, yielding nests such as regular/lager, light beer, ale, stout/porter, and a residual category for other types. This approach mirrors common empirical practice, in which researchers select a plausible partition based on observable product categories and estimate the model conditional on that grouping. We refer to this specification as the "naive" nested logit.

Following this naive approach, we find that the naive nested logit performs worse than the data-driven nested logit along several dimensions. First, when we implement the Bud Light boycott validation exercise, the naive specification predicts substantially less substitution toward the group to which Bud Light belongs (labeled *light beer*) compared to the substitution toward the data-driven light beer group using our out-of-sample exercise that leverages the Bud Light boycott. Concretely, under the data-driven nested logit, products in the same group as Bud Light experience a large and statistically significant increase in demand following the boycott. The coefficient on *Treatment same group* $\times\, t = 2023$ is 0.422, indicating substantial substitution toward products within the same estimated nest. By contrast, the naive nested logit produces much weaker substitution patterns. The coefficient on *Naive light beer* $\times\, t = 2023$ is only 0.13, substantially smaller than the data-driven estimate.

This difference is confirmed in the direct comparison column. In this specification, we construct three treatment subgroups: products that belong only to the Light Beer group identified by the data-driven classification, products that belong only to the Light Beer group defined by NielsenIQ, and products that belong to both groups. We find that the substitution captured by the Naive Light Beer group is entirely driven by products that also belong to the data-driven Light Beer group. By contrast, products that belong only to the Naive Light Beer group do not experience any statistically significant demand response following the boycott.

These results indicate that the substitution patterns implied by the data-driven nesting

structure more closely match the observed demand responses following the boycott. One possible explanation is that NielsenIQ classifies products as Light Beer based on low caloric content, which may arise either because the beer is truly light or because it is sold in smaller containers. As a result, the NielsenIQ classification may group together products that are not close substitutes from a demand perspective.

Table 6: Diff-in-Diff results of Bud Light boycott with Naive Nested Logit

| Dependent Variable: Log Share | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Data-driven NL | | | | Naive NL | | | | Interaction | |
| t=2023 | -0.408*** | (0.005) | -0.449*** | (0.007) | -0.329*** | (0.006) | -0.449*** | (0.007) | -0.381*** | (0.007) |
| Treatment same group | 0.422*** | (0.015) | | | | | | | | |
| Regular beer × t=2023 | | | 0.173*** | (0.013) | | | | | | |
| Light beer × t=2023 | | | 0.463*** | (0.016) | | | | | | |
| Non-lager/craft × t=2023 | | | -0.001 | (0.012) | | | | | | |
| Boutique craft × t=2023 | | | 0.018 | (0.017) | | | | | | |
| Naive light beer × t=2023 | | | | | 0.130*** | (0.018) | 0.250*** | (0.018) | | |
| Naive ale × t=2023 | | | | | | | 0.028** | (0.012) | | |
| Naive other × t=2023 | | | | | | | 0.009 | (0.041) | | |
| Naive lager × t=2023 | | | | | | | 0.249*** | (0.012) | | |
| Naive stout × t=2023 | | | | | | | -0.041* | (0.023) | | |
| DD Light only | | | | | | | | | 0.385*** | (0.018) |
| Naive Light only | | | | | | | | | 0.033 | (0.022) |
| DD and Naive Light | | | | | | | | | 0.422*** | (0.027) |
| Observations | 214353 | | 172778 | | 214353 | | 172778 | | 214353 | |
| UPC-Market FE | YES | | YES | | YES | | YES | | YES | |
| Exclude Bud Light | YES | | YES | | YES | | YES | | YES | |
| Baseline group | Other groups | | Outside option | | Outside option | | Outside option | | Outside option | |

Notes: *** p<0.01, ** p<0.05, * p<0.1

Each column reports OLS estimates from a difference-in-differences specification with log market share as the dependent variable, comparing 2022 to 2023 after the Bud Light boycott. Bud Light products are excluded from all columns. Columns (1)–(2) use the data-driven nesting structure ($k^* = 4$); Columns (3)–(4) use a naive nesting structure based on beer type (light beer, ale, lager, stout/porter, other). In Column (1), the treatment indicator equals one for products in Bud Light's data-driven group (light beer) interacted with $t = 2023$; the baseline comprises all other groups. In Column (2), group-specific post-boycott effects are estimated relative to the outside option. Columns (3) and (4) replicate this design using naive beer-type groups. Column (5) interacts data-driven and naive light-beer indicators to decompose substitution into products classified as light beer by only the data-driven method, only the naive method, or both. All specifications include UPC-by-market (state) fixed effects. Robust (HC1) standard errors are reported in parentheses.

Second, when examining the structural parameter estimates, we find that one group yields a negative $\hat{\sigma}^k$, which lies outside the range consistent with utility maximization. This provides empirical evidence that the nesting structure implied by the naive specification is rejected by the data.

Table 7: Estimation Results for Naive Nested Logit

| | OLS | IV |
|---|---|---|
| log unit size | 0.054 (0.005)*** | 0.686 (0.041)*** |
| log upc quantity | 0.059 (0.004)*** | 0.694 (0.042)*** |
| log price | -0.039 (0.005)*** | -0.753 (0.050)*** |
| $\sigma_{\text{ale}}$ | -0.029 (0.001)*** | 0.239 (0.012)*** |
| $\sigma_{\text{light beer}}$ | 0.087 (0.001)*** | 0.348 (0.010)*** |
| $\sigma_{\text{other}}$ | -0.356 (0.002)*** | -0.098 (0.015)*** |
| $\sigma_{\text{regular/lager}}$ | 0.080 (0.001)*** | 0.322 (0.010)*** |
| $\sigma_{\text{stout/porter}}$ | -0.296 (0.001)*** | 0.040 (0.017)** |
| Mean Own-price Elasticity | 0.360 (0.003) | -3.752 (0.104) |
| Mean Cross-price Elasticity | -0.001 (0.003) | 0.007 (0.111) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Note:* The table reports OLS and IV estimates of a nested logit demand model where products are grouped by beer type (regular beer/lager, light beer, ale, stout/porter, and other), yielding $k = 5$ nests. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and five group-specific nesting parameters ($\sigma_g$). The IV specification instruments for log price using Gandhi-Houde differentiation IVs and for the within-group log share using the exogenous share projection. Market fixed effects are included. Mean own-price and cross-price elasticities are computed from the IV estimates and weighted by national unit sales. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

### 7.4.2 Mixed Logit

In this section, we estimate a mixed logit model using PyBLP (Conlon and Gortmaker, 2020). Concretely, we specify that the utility of an individual $i$ from purchasing product $j$ in market $m$ is given by

$$u_{ijm} = \delta_{jm} + \mu_{ijm} + \epsilon_{ijm}, \tag{23}$$

where $\epsilon_{ijm}$ follows the Type I extreme distribution. The common component of utility is given by

$$\delta_{jm} = \lambda_m + \beta_p \log \mathrm{p}_{jm} + \beta_s \log \mathrm{s}_j + \beta_q \log \mathrm{q}_j + \xi_{jm}, \tag{24}$$

and the individual-specific component $\mu_{ijm}$ is given by

$$\mu_{ijm} = (\sigma_p \nu_i) \log \mathrm{p}_{jm}, \tag{25}$$

where $\sigma_p$ governs the variance of the idiosyncratic component. $\nu_i \sim N(0,1)$ follows a standard normal distribution. We estimate the parameters $\{\boldsymbol{\beta}, \sigma_p\}$ by solving the following GMM problem:[19]

$$\min_{\boldsymbol{\beta},\sigma_p} \quad \bar{g}\left(\boldsymbol{\beta},\sigma_p\right)' W \bar{g}\left(\boldsymbol{\beta},\sigma_p\right), \quad \bar{g}\left(\boldsymbol{\beta},\sigma_p\right) = \frac{1}{N}\sum_{j,m}\left(Z'_{jm}\xi_{jm}(\boldsymbol{\beta},\sigma_p)\right) \tag{26}$$

where $\boldsymbol{\beta} = [\beta_p, \beta_s, \beta_q]$ and $Z_{jm}$ denote instruments. Specifically, we use the Gandhi-Houde instruments to make results as comparable as possible.[20] $W$ is a weighting matrix whose initial values are set by the 2SLS estimate and then updated using a two-step optimal GMM weights.[21] The estimation results are reported in Table 8.

Table 8: Mixed Logit estimates

| **Panel A: Linear Coefficients** | $\boldsymbol{\beta}$ |
|---|---|
| log price | -3.775*** (0.090) |
| log unit size | 3.077*** (0.062) |
| log unit quantity | 3.222*** (0.062) |
| **Panel B: Non-linear Coefficient** | $\sigma_p$ |
| log price | 0.002*** (0.0003) |
| Own-price elasticity | -4.683 (2.863) |
| Cross-price elasticity | 0.002 (0.005) |
| Market Fixed Effects | ✓ |
| IV Type | Gandhi-Houde |
| Number of Products | 2806 |
| Number of Observations | 64497 |

*Notes:* Panel A displays the estimates of $\boldsymbol{\beta}$ in equation (26). Panel B displays the estimates of $\sigma_p$. Robust standard errors in parentheses.

With these coefficients, the average price elasticity is -4.683 with a standard deviation of 2.863 in the average market. For cross-price elasticities, we find an average value of 0.002 with standard deviation of 0.005. Despite being a less general model, for this particular application, we find that the data-driven NL model delivers more heterogeneity in substitution

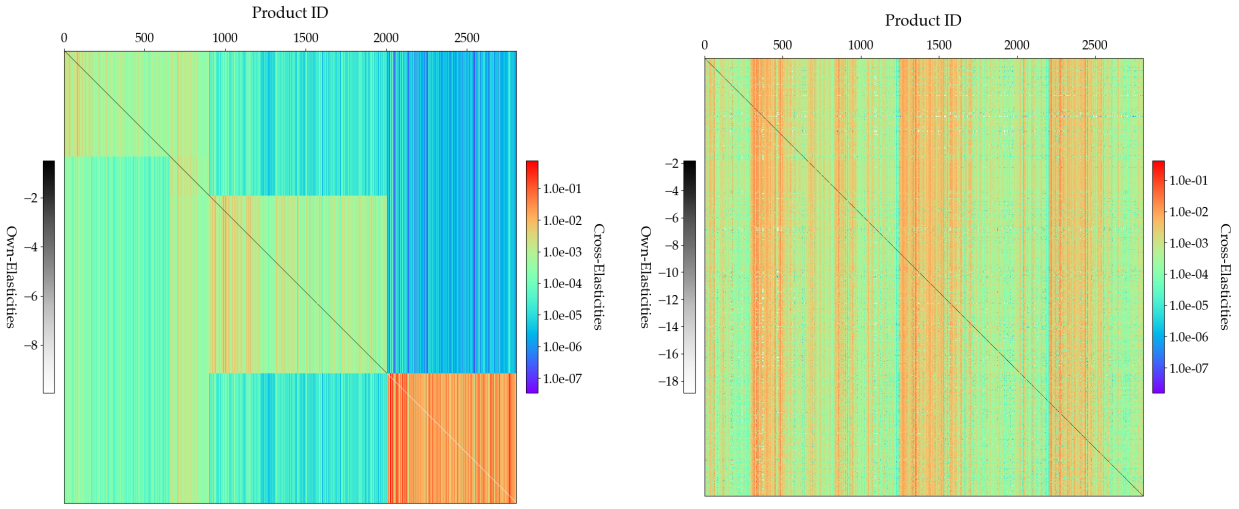
[19] In practice, we use the `slsqp` optimization routine with box contraints in PyBLP.

[20] We also estimate the previous model using optimal instruments as suggested by Conlon and Gortmaker (2020). However, using the same optimization routine and same initial value as in our baseline specification, we obtain an estimate for the variance coefficient that converges to zero.

[21] See Conlon and Gortmaker (2020) for details.

patterns across products with450% more variation in cross-price elasticities.

Figure 8: Price elasticity matrices




(a) Nested logit ($k^* = 4$)

(b) Mixed Logit

*Notes:* The figures above display the price elasticity matrices estimated using the nested logit specification and the BLP procedure. Diagonal cells each matrix denote products' own-price elasticities, whereas the off-diagonal cells denote cross-price elasticities. Concretely, the value in row $i$, column $j$ denotes $\frac{\partial \log \mathbb{P}_i}{\partial \log p_j}$.

We can also compare the full matrix of price elasticities for both models, as depicted in Figure 8. In the figure, products are ordered first by group and then by UPC identifier, and elasticities are averaged across markets. The left panel plots the estimated elasticities from the data-driven nested logit model, while the right panel reports elasticities from the mixed logit model.

The data-driven nested logit displays substantial variation in substitution patterns with a clear block-diagonal structure, reflecting stronger substitution within nests than across nests. By contrast, the mixed logit produces elasticity patterns that closely resemble those of a simple logit model, with relatively little variation across products.

One possible explanation is that the market shares of many products are extremely small and exhibit limited variation across markets, which makes it difficult to identify substantial heterogeneity in price sensitivities. As a result, $1 - \hat{\mathbb{P}}_{jm}(p^i)$ is close to one for most products, implying that the elasticity expression—which involves an integral over the distribution of random coefficients in the mixed logit—is well approximated by the simple logit formula.

Overall, we conclude that despite being in theory a somewhat restrictive model, the data-driven nested logit model is able to empirically capture very heterogeneous substitution

patterns when applied to this particular setting.

# 8 Conclusion and Future Work

We develop a method to jointly estimate utility parameters and the nesting structure in nested logit models when the grouping of products is unknown. Exploiting the restrictions imposed by the nested logit model, we provide consistency results in settings with endogenous prices. We then propose a two-step estimation strategy. In the first step, we recover the grouping structure using a clustering procedure applied to model-implied demand curves. In the second step, we estimate preference parameters conditional on the estimated groups as if they were known. Under appropriate regularity conditions, we show that this procedure delivers an estimator that is asymptotically equivalent to the case in which the true nesting structure is known.

We illustrate our approach by estimating demand in the U.S. beer market, where the relevant product groups are not known ex ante. Our method identifies four distinct groups. Imposing the estimated grouping structure, we estimate an average own-price elasticity of -3.561and within-group parameters ranging between 0.042 and 0.807, values that are consistent with utility maximization without imposing any additional constraints. The large dispersion in within-group correlations suggests substantial heterogeneity in substitution patterns across products. We also compare the data-driven nested logit with a naive nested logit and a mixed logit model with random coefficients. In this empirical setting, the data-driven nested logit outperforms both alternatives, both in an out-of-sample validation exercise and in its ability to capture richer heterogeneity in substitution patterns.

Several directions for future research remain. First, researchers may wish to apply the clustering procedure using alternative moments that arise from non-linear estimation problems, such as non-linear least squares. Developing the corresponding theoretical results for non-linear models is therefore an important next step. Second, in some applications researchers wish to combine a nesting structure with a mixed logit specification but lack prior information about the relevant nests.[22] Although the theory for such non-linear settings has not yet been developed, it may still be possible to apply our approach using the linear approximation to mixed logit proposed by Salanié and Wolak (2022). Understanding

[22] See, for example, Conlon and Rao (2023).

the conditions under which this approximation delivers consistent estimates of the nesting structure is an important avenue for future work. More broadly, incorporating consumer heterogeneity by combining BLP-type demand estimation with our clustering procedure represents a promising direction for extending the framework.

Finally, while we focus on an application in industrial organization, our approach may prove useful in other contexts where the relevant grouping structure is not directly observable. Examples include defining labor markets using worker mobility data, identifying neighborhoods in residential choice models, or studying product segmentation in international trade following Broda and Weinstein (2006). Even when data are available only in cross-section, the bilateral nature of trade flows may allow researchers to construct the panel variation needed to apply our method.

# References


Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the Berlin wall. *Econometrica*, 83(6):2127–2189.

Allcott, H., Kane, R., Maydanchik, M. S., Shapiro, J. S., and Tintelnot, F. (2024). The effects of "Buy American": Electric vehicles and the Inflation Reduction Act. Technical report, National Bureau of Economic Research.

Allen, T. and Arkolakis, C. (2022). The welfare effects of transportation infrastructure improvements. *Review of Economic Studies*, 89(6):2911–2957.

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

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031.

Azar, J. A., Berry, S. T., and Marinescu, I. (2022). Estimating labor market power. Technical report, National Bureau of Economic Research.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Berry, S., Eizenberg, A., and Waldfogel, J. (2016). Optimal product variety in radio markets. *The RAND Journal of Economics*, 47(3):463–497.

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

Björnerstedt, J. and Verboven, F. (2016). Does merger simulation work? evidence from the swedish analgesics market. *American Economic Journal: Applied Economics*, 8(3):125–164.

Bonhomme, S., Lamadon, T., and Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, 90(2):625–643.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bordeu, O. (2024). Commuting infrastructure in fragmented cities.

Bresnahan, T. F. (1987). Competition and collusion in the American automobile industry: The 1955 price war. *Journal of Industrial Economics*, 35(4):457–482.

Broda, C. and Weinstein, D. E. (2006). Globalization and the gains from variety. *The Quarterly journal of economics*, 121(2):541–585.

Bronnenberg, B. J., Dubé, J.-P. H., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *American Economic Review*, 102(6):2472–2508.

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.

Choi, J. and Okui, R. (2024). Latent group structure in linear panel data models with endogenous regressors. *arXiv preprint arXiv:2405.08687*.

Ciliberto, F. and Williams, J. W. (2014). Does multimarket contact facilitate tacit collusion? inference on conduct parameters in the airline industry. *The RAND journal of economics*, 45(4):764–791.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. *RAND Journal of Economics*, 51(4):1108–1161.

Conlon, C. and Rao, N. L. (2023). The cost of curbing externalities with market power: Alcohol regulations and tax alternatives. Technical report, National Bureau of Economic Research Cambridge, MA, USA.

Couture, V., Dingel, J. I., Green, A. E., and Handbury, J. (2025). Demographic preferences and income segregation. Technical report, National Bureau of Economic Research.

Couture, V., Gaubert, C., Handbury, J., and Hurst, E. (2024). Income growth and the distributional effects of urban spatial sorting. *Review of Economic Studies*, 91(2):858–898.

Davis, D. R., Dingel, J. I., Monras, J., and Morales, E. (2019). How segregated is urban consumption? *Journal of Political Economy*, 127(4):1684–1738.

Diamond, R., McQuade, T., and Qian, F. (2019). The effects of rent control expansion on tenants, landlords, and inequality. *American Economic Review*, 109(9):3365–3394.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Dubé, J.-P., Fox, J. T., and Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80(5):2231–2267.

Dubé, J.-P., Hitsch, G. J., and Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. *RAND Journal of Economics*, 41(3):417–445.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Einav, L. (2007). Seasonality in the us motion picture industry. *The Rand journal of economics*, 38(1):127–145.

Eizenberg, A. (2014). Upstream innovation and product variety in the US home PC market. *Review of Economic Studies*, 81(3):1003–1045.

Fajgelbaum, P., Grossman, G. M., and Helpman, E. (2011). Income distribution, product quality, and international trade. *Journal of Political Economy*, 119(4):721–765.

Fosgerau, M., Monardo, J., and De Palma, A. (2019). The inverse product differentiation logit model. *Available at SSRN 3141041*.

Fosgerau, M., Monardo, J., and de Palma, A. (2024a). Biases in nested logit estimation under misspecified structure. *American Economic Journal: Applied Economics*. evaluates biases in multi-level nested logit vs RC logit.

Fosgerau, M., Monardo, J., and De Palma, A. (2024b). The inverse product differentiation logit model. *American Economic Journal: Microeconomics*, 16(4):329–370.

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

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the US automobile industry. *Econometrica: Journal of the Econometric Society*, pages 891–951.

Grennan, M. (2013). Price discrimination and bargaining: Empirical evidence from medical devices. *American Economic Review*, 103(1):145–177.

Griffith, R., Nesheim, L., and O'Connell, M. (2018). Income effects and the welfare consequences of tax in differentiated product oligopoly. *Quantitative Economics*, 9(1):305–341.

Grigolon, L. and Verboven, F. (2014). Nested logit or random coefficients logit? a comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics*, 96(5):916–935.

Grigsby, J. R. (2022). Skill heterogeneity and aggregate labor market dynamics.

Hahn, J. and Moon, H. R. (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory*, 26(3):863–881.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In Bresnahan, T. and Gordon, R. J., editors, *The Economics of New Goods*, pages 207–248. University of Chicago Press, Chicago.

Hendel, I. and Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673.

Hortacsu, A., Lieber, J., Monardo, J., and de Paula, A. (2023). Estimating nesting structures. Work in progress. No publicly available draft.

Khandelwal, A. (2010). The long and short (of) quality ladders. *The Review of Economic Studies*, 77(4):1450–1476.

Knittel, C. R. and Metaxoglou, K. (2014). Estimation of random-coefficient demand models: two empiricists' perspective. *Review of Economics and Statistics*, 96(1):34–59.

Kovach, M. and Tserenjigmid, G. (2022). Behavioral foundations of nested stochastic choice and nested logit. *Journal of Political Economy*, 130(9):2411–2461.

McFadden, D. (1978). Modelling the choice of residential location. In *Spatial Interaction Theory and Planning Models*. North Holland.

McFadden, D. and Train, K. (2000). Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470.

Miller, N. H. and Weinberg, M. C. (2017). Understanding the price effects of the MillerCoors joint venture. *Econometrica*, 85(6):1763–1791.

Miravete, E. J., Seim, K., and Thurk, J. (2018). Market power with a simple pricing rule: The case of alcohol. *Review of Economic Studies*, 85(4):2295–2332.

Moon, H. R. and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579.

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

Parry, I. W. and Small, K. A. (2005). Does britain or the united states have the right gasoline tax? *American Economic Review*, 95(4):1276–1289.

Pinkse, J. and Slade, M. E. (2004). Mergers, brand competition, and the price of a pint. *European Economic Review*, 48(3):617–643.

Romeo, C. J. (2016). Incorporating prior information into a GMM objective for mixed logit demand systems. *Journal of Industrial Economics*, 64(2):336–363.

Salanié, B. and Wolak, F. A. (2022). Fast, detail-free, and approximately correct: Estimating mixed demand systems. Technical report, Working paper.

Slade, M. E. (2004). Market power and joint dominance in UK brewing. *Journal of Industrial Economics*, 52(1):133–163.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Verboven, F. (1996). International price discrimination in the european car market. *The RAND Journal of Economics*, pages 240–268.

Verhoogen, E. A. (2008). Trade, quality upgrading, and wage inequality in the mexican manufacturing sector. *The Quarterly Journal of Economics*, 123(2):489–530.

Wollmann, T. G. (2018). Trucks without bailouts: Equilibrium product characteristics for commercial vehicles. *American Economic Review*, 108(6):1364–1406.

# A Appendix

## A.1 Substitution patterns and elasticity of demand

For simplicity, let's assume that $\delta_{jm} = \alpha \log p_{jm} + \beta \log X_{jm} + \xi_{jm}$.
Own-price elasticities are given by:

$$\mathcal{E}_j^j = \frac{\alpha}{\sigma^{k(j)}}\Big(1 - \sigma^{k(j)}\mathbb{P}_{jm} - (1-\sigma^{k(j)})\mathbb{P}_{jm|k(j)}\Big)$$

For cross-price elasticites, recall that if product $q \in B_{k'} \neq B_k$

$$\begin{aligned}\frac{\partial}{\partial p_q}\mathbb{P}_j &= -\frac{e^{\frac{v_j}{\sigma^k}}\big(\sum_{d\in B_k} e^{\frac{v_d}{\sigma^k}}\big)^{\sigma^k-1}\sigma^{k'}\big(\sum_{d\in B_{k'}} e^{\frac{v_d}{\lambda_{k'}}}\big)^{\lambda_{k'}-1}\Big(-\frac{\alpha}{\lambda_{k'}}e^{\frac{v_q}{\lambda_{k'}}}\Big)}{\Big(\sum_{l=1}^K\big(\sum_{d\in B_l} e^{\frac{v_d}{\sigma^l}}\big)^{\sigma^l}\Big)^2} \\ &= -\alpha\mathbb{P}_j\mathbb{P}_q.\end{aligned}$$

On the other hand, if product $q \in B_k$

$$\begin{aligned}\frac{\partial}{\partial p_q}\mathbb{P}_j &= \frac{e^{\frac{v_j}{\sigma^k}}(\sigma^k-1)\big(\sum_{d\in B_k} e^{\frac{v_d}{\sigma^k}}\big)^{\sigma^k-2}\Big(-\frac{\alpha}{\sigma^k}e^{\frac{v_q}{\sigma^k}}\Big)}{\sum_{l=1}^K\big(\sum_{d\in B_l} e^{\frac{v_d}{\sigma^l}}\big)^{\sigma^l}} \\ &\quad - \frac{e^{\frac{v_j}{\sigma^k}}\big(\sum_{d\in B_k} e^{\frac{v_d}{\sigma^k}}\big)^{\sigma^k-1}\sigma^k\big(\sum_{d\in B_k} e^{\frac{v_d}{\sigma^k}}\big)^{\sigma^k-1}\Big(-\frac{\alpha}{\sigma^k}e^{\frac{v_q}{\sigma^k}}\Big)}{\Big(\sum_{l=1}^K\big(\sum_{d\in B_l} e^{\frac{v_d}{\sigma^l}}\big)^{\sigma^l}\Big)^2} \\ &= -\alpha\Big((\sigma^k-1)\frac{1}{\sigma^k}\mathbb{P}_j\mathbb{P}_{q|k} + \mathbb{P}_j\mathbb{P}_q\Big).\end{aligned}$$

Therefore, price elasticities are given by

$$\mathcal{E}_q^j = \frac{\frac{\partial}{\partial p_q}\mathbb{P}_j}{\frac{\mathbb{P}_j}{p_q}} = \begin{cases} \frac{\alpha}{\sigma^{k(j)}}\Big(1 - \sigma^{k(j)}\mathbb{P}_{jm} - (1-\sigma^{k(j)})\mathbb{P}_{jm|k(j)}\Big) & \text{if } j = q \\ -\alpha\mathbb{P}_q & \text{if } q \in B_{k'} \neq B_k \\ -\alpha\Big((\sigma^k-1)\frac{1}{\sigma^k}\mathbb{P}_{q|k} + \mathbb{P}_q\Big) & \text{if } q \in B_k \end{cases}$$

## A.2 Approximation of Pricing Models by Linear Pricing Projections

### A.2.1 Primitives

We index products by $j$ and markets by $m$. Below we describe the baseline primitives of demand, supply, and marginal costs.

First, demand is a function of a nesting structure defined by the *demand partition* $k(j)$. This partition governs substitution patterns under the nested logit or nested CES demand systems. Second, for simplicity, we assume that marginal costs are a linear function of observed product characteristics $\mathbf{x}_{jm}$ and exogenous price shifters $\mathbf{z}_{jm}$ (Berry, 1994; Berry et al., 1995; Nevo, 2001):[23]

$$mc_{jm} = \boldsymbol{\gamma}_x^{mc,c(j)\prime}\mathbf{x}_{jm} + \boldsymbol{\gamma}_z^{mc,c(j)\prime}\mathbf{z}_{jm} + \kappa_m^{mc,c(j)} + \omega_{jm}, \tag{28}$$

where $c(j)$ denotes a partition of products along cost structures and $\omega_{jm}$ are unobservable cost shocks.

Products in the same cost group share the coefficient vectors $\boldsymbol{\gamma}_x^{mc,c(j)}$ and $\boldsymbol{\gamma}_z^{mc,c(j)}$, so that the pass-through of observed characteristics and input prices into marginal cost is homogeneous within the group but may differ across groups. The fixed effect $\kappa_m^{mc,c(j)}$ absorbs cost determinants common to all products in the same group within a market, such as segment-specific local input prices or group-level productivity shocks. The residual $\omega_{jm}$ captures unobserved cost heterogeneity at the product level, which we assume to be mean-independent of the other components.

The goal of this appendix is to show how certain common pricing models map or can be approximated by the following empirical pricing projections:

$$p_{jm} = \boldsymbol{\gamma}_x^{k'(j)\prime}\mathbf{x}_{jm} + \boldsymbol{\gamma}_z^{k'(j)\prime}\mathbf{z}_{jm} + \kappa_m^{k'(j)} + v_{jm}, \tag{29}$$

and to spell out the assumptions required in each case. Here $k'(j)$ is the *pricing partition*: the group structure under which products share common coefficient vectors $\boldsymbol{\gamma}_x^{k'(j)}$ and $\boldsymbol{\gamma}_z^{k'(j)}$ and

[23] This marginal cost function implies constant returns to scale because it is not a function of quantities. It is straightforward to depart from linearity or make marginal costs a function of quantities $Q_{jm}$:

$$mc_{jm} = \phi^{c(j)}(x_{jm}, z_{jm}, Q_{jm}, \omega_{jm}), \tag{27}$$

a group-by-market fixed effect $\kappa_m^{k'(j)}$.

We will start with the case of Nash-Bertrand in single-product firms but will generalize our results to other common competition models and a general ownership structure. For each model we derive $k'(j)$ as a function of the primitive partitions $k(j)$ and $c(j)$, and of the ownership structure $f(j)$ when firms are multiproduct. When two partitions $P$ and $Q$ are both relevant, we write $P \cap Q$ for their *common refinement*: the partition under which two products share the same group if and only if they share the same group under both $P$ and $Q$.

### A.2.2 Nash-Bertrand with Single-Product Firms

Under Bertrand–Nash competition with single-product firms, the first-order condition gives:

$$p_{jm} = \frac{1}{1 - \frac{1}{\mathcal{E}^j_{jm}}} \cdot mc_{jm}. \tag{30}$$

Under nested logit with nest-specific $\sigma^{k(j)} \in (0, 1]$, from (2) and utility expression (3), we can see that:

$$\mathcal{E}^j_{jm} = \frac{\beta_p}{\sigma^{k(j)}} + \mathcal{O}_p(J^{-1}), \tag{31}$$

where the remainder vanishes as the number of products $J$ grows large and individual shares become small. We can see that the main component of the elasticity, $\frac{\beta}{\sigma^{k(j)}}$ is nest-specific and decreasing in $\sigma^{k(j)}$. Substituting into prices:

$$\log p_{jm} = \log mc_{jm} - \log(1 - \frac{\sigma^{k(j)}}{\beta_p}) + \mathcal{O}_p(\log J^{-1}).$$

As $\sigma^{k(j)} \to 0$, products within the nest become more differentiated from each other, consumers become more elastic, competition intensifies, prices converge to marginal costs, and markups go to zero.[24] When $\sigma^{k(j)} = 1$, the nesting structure is irrelevant. Substituting (28)

[24] We can equivalently derive a similar expression from prices in levels. From first order conditions, it follows:

$$p_{jm} = mc_{jm} + \mu^{own}_{jm}, \qquad \mu^{own}_{jm} = \frac{\mathbb{P}_{jm}}{-\partial \mathbb{P}_{jm}/\partial p_{jm}}. \tag{32}$$

In the limit, $\mu^{own}_{jm} = \frac{\sigma^{k(j)}}{\beta}$.

and (31) into (30):

$$\log p_{jm} = \boldsymbol{\gamma}_x' \mathbf{x}_{jm} + \boldsymbol{\gamma}_z' \mathbf{z}_{jm} + \underbrace{\kappa_m^{mc} - \log\left(1 - \frac{\sigma^{k(j)}}{\beta}\right)}_{\kappa_m^{k'(j)}} + \underbrace{\omega_{jm}}_{v_{jm}} + \mathcal{O}_p(\log J^{-1}), \tag{33}$$

which is (5) with pricing partition $k'(j) = k(j)$ and coefficient vectors equal to their cost counterparts. The nest-specific price component that is absorbed into the group-by-market fixed effect $\kappa_m^{k'(j)}$, leaving the slope coefficients unaffected.

### A.2.3 Extensions

**Extension to multiproduct firms.** Starting from the Nash-Bertrand first-order condition for a multi-product firm $f$ in market $m$,

$$p_{jm} = mc_{jm} + \frac{\mathbb{P}_{jm}}{\partial \mathbb{P}_{jm}/\partial p_{jm}} - \sum_{l \in \mathcal{F}_{fm}, l \neq j} \mu_{lm} \frac{\partial \mathbb{P}_{lm}/\partial p_{jm}}{\partial \mathbb{P}_{jm}/\partial p_{jm}}, \tag{34}$$

where $\mathcal{F}_{fm}$ is the portfolio of firm $f$ in market $m$ and $\mu_{jm} \equiv p_{jm} - mc_{jm}$.

Let $D_{jm} \equiv 1 - (1 - \sigma^{k(j)})\mathbb{P}_{j|k(j),m} - \sigma^{k(j)}\mathbb{P}_{jm}$, where $k(j)$ denotes the nest of product $j$, $\mathbb{P}_{j|k(j),m}$ is the within-nest share, and $\mathbb{P}_{k(j),m}$ is the nest-level share. The own-price derivative gives $\mathbb{P}_{jm}/(\partial \mathbb{P}_{jm}/\partial p_{jm}) = \sigma^{k(j)}/(\beta_p D_{jm})$. Using $\mathbb{P}_{jm} = \mathbb{P}_{j|k(j),m}\mathbb{P}_{k(j),m}$, the cross-to-own derivative ratios are

$$\frac{\partial \mathbb{P}_{lm}/\partial p_{jm}}{\partial \mathbb{P}_{jm}/\partial p_{jm}} = \begin{cases} -\dfrac{(1-\sigma^{k(j)}) + \sigma^{k(j)}\mathbb{P}_{k(j),m}}{\mathbb{P}_{k(j),m}\, D_{jm}} \cdot \mathbb{P}_{lm} & l \in k(j),\ l \neq j, \\ -\dfrac{\sigma^{k(j)}}{D_{jm}} \cdot \mathbb{P}_{lm} & l \notin k(j). \end{cases}$$

Substituting into (34), the pricing equation becomes

$$p_{jm} = mc_{jm} + \frac{1}{\beta_p D_{jm}} \left[ \sigma^{k(j)} + \frac{(1-\sigma^{k(j)}) + \sigma^{k(j)}\mathbb{P}_{k(j),m}}{\mathbb{P}_{k(j),m}} \sum_{\substack{l \in \mathcal{F}_{fm} \cap k(j) \\ l \neq j}} \mu_{lm}\, \mathbb{P}_{lm} + \sigma^{k(j)} \sum_{l \in \mathcal{F}_{fm} \setminus k(j)} \mu_{lm}\, \mathbb{P}_{lm} \right]. \tag{35}$$

Since $\mu_{jm} = p_{jm} - mc_{jm}$, equation (35) defines the linear fixed point

$$\boldsymbol{\mu}_m = \mathbf{b}_m + \mathbf{A}_m \boldsymbol{\mu}_m, \tag{36}$$

where $\mathbf{b}_m$ has entries $b_{jm} = \sigma^{k(j)}/(\beta_p D_{jm})$ and $\mathbf{A}_m$ has entries

$$A_{jl} = \begin{cases} \dfrac{(1-\sigma^{k(j)}) + \sigma^{k(j)}\mathbb{P}_{k(j),m}}{\beta_p \, D_{jm} \, \mathbb{P}_{k(j),m}} \cdot \mathbb{P}_{lm} & l \in \mathcal{F}_{fm} \cap k(j),\ l \neq j, \\ \dfrac{\sigma^{k(j)}}{\beta_p \, D_{jm}} \cdot \mathbb{P}_{lm} & l \in \mathcal{F}_{fm} \setminus k(j), \\ 0 & \text{otherwise.} \end{cases}$$

**Asymptotic expansion.** Assume $(I - \mathbf{A}_m)^{-1}$ is finite and that its limit as $J \to \infty$ exists. The solution to (36) is then given by the Neumann expansion

$$\boldsymbol{\mu}_m = (I - \mathbf{A}_m)^{-1}\mathbf{b}_m = \mathbf{b}_m + \mathbf{A}_m\mathbf{b}_m + \mathbf{A}_m^2\mathbf{b}_m + \cdots$$

As $J \to \infty$, $\mathbb{P}_{jm} \to 0$ and $\mathbb{P}_{j|k(j),m} \to 0$, so $D_{jm} = 1 + O(J^{-1})$. This gives $b_{jm} = \sigma^{k(j)}/\beta_p + O(J^{-1})$, and every entry $A_{jl} = O(J^{-1})$, so $\mathbf{A}_m^k = O(J^{-k})$ for each $k \geqslant 1$. The expansion therefore truncates at first order:

$$\boldsymbol{\mu}_m = \mathbf{b}_m + \mathbf{A}_m\mathbf{b}_m + O(J^{-2}).$$

Substituting $b_{jm} = \sigma^{k(j)}/\beta_p + O(J^{-1})$ and $\mathbf{A}_m = \bar{\mathbf{A}}_m + O(J^{-2})$, where $\bar{\mathbf{A}}_m$ is obtained by setting $D_{jm} = 1$:

$$\boldsymbol{\mu}_m = \frac{\sigma^{k(j)}}{\beta_p} + \bar{\mathbf{A}}_m\frac{\sigma^{k(j)}}{\beta_p} + O(J^{-1}) = \frac{\sigma^{k(j)}}{\beta_p}\left(1 + \bar{A}_m\right) + O(J^{-1}),$$

where $\bar{A}_m$ denotes the row sum of $\bar{\mathbf{A}}_m$. Recognizing that the full geometric series $(I - \bar{\mathbf{A}}_m)^{-1} = I + \bar{\mathbf{A}}_m + \bar{\mathbf{A}}_m^2 + \cdots$ sums to $1/(1-\bar{A}_m)$ under symmetry, we obtain $\boldsymbol{\mu}_m = \bar{\boldsymbol{\mu}}_m + O(J^{-1})$ where $\bar{\mu} = (\sigma^{k(j)}/\beta_p)/(1 - \bar{A}_m)$. The row sum of $\bar{\mathbf{A}}_m$ is

$$\bar{A}_m = \frac{(1-\sigma^{k(j)}) + \sigma^{k(j)}\mathbb{P}_{k(j),m}}{\beta_p \, \mathbb{P}_{k(j),m}} \sum_{\substack{l \in \mathcal{F}_{fm} \cap k(j) \\ l \neq j}} \mathbb{P}_{lm} \; + \; \frac{\sigma^{k(j)}}{\beta_p} \sum_{l \in \mathcal{F}_{fm} \setminus k(j)} \mathbb{P}_{lm},$$

Substituting $\mathbb{P}_{k(j),m} \to 1/K_k$ under balanced nests and splitting the first term:

$$\bar{A}_m = \frac{K_k(1-\sigma^{k(j)})}{\beta_p} \sum_{\substack{l \in \mathcal{F}_{fm} \cap k(j) \\ l \neq j}} \mathbb{P}_{lm} + \frac{\sigma^{k(j)}}{\beta_p} \sum_{\substack{l \in \mathcal{F}_{fm} \\ l \neq j}} \mathbb{P}_{lm}.$$

Assume the firm's total market share converges to a fixed limit, $\sum_{l \in \mathcal{F}_{fm}, l \neq j} \mathbb{P}_{lm} \to \bar{S}_{fm}$. If the firm's products are assigned to nests independently and uniformly, then by a law of large numbers each nest receives a share $1/K_k$ of the firm's products, so

$$\sum_{\substack{l \in \mathcal{F}_{fm} \cap k(j) \\ l \neq j}} \mathbb{P}_{lm} \;\to\; \frac{\bar{S}_{fm}}{K_k}.$$

Substituting:

$$\bar{A}_m \;\to\; \frac{K_k(1-\sigma^{k(j)})}{\beta_p} \cdot \frac{\bar{S}_{fm}}{K_k} + \frac{\sigma^{k(j)}}{\beta_p} \bar{S}_{fm} = \frac{\bar{S}_{fm}}{\beta_p},$$

where $\sigma^{k(j)}$ cancels out. The limiting markup is therefore

**Result.** Since every entry of $\mathbf{A}_m$ is $O(J^{-1})$, we have $\mathbf{A}_m^2 = O(J^{-2})$ and the Neumann expansion truncates at first order:

$$\boldsymbol{\mu}_m = (I + \mathbf{A}_m + O(J^{-2}))\mathbf{b}_m = \mathbf{b}_m + \mathbf{A}_m \mathbf{b}_m + O(J^{-2}).$$

We now substitute the limits of $\mathbf{b}_m$ and $\mathbf{A}_m$. First, $b_{jm} = \sigma^{k(j)}/\beta_p + O(J^{-1})$. Second, assume nests grow proportionally so that $\mathbb{P}_{k(j),m} \to 1/K_k$ with $K_k/K_{k'} = O(1)$ for all $k, k'$, and that the ownership structure follows a fixed partition so that $\sum_{l \in \mathcal{F}_{fm} \cap k(j), l \neq j} \mathbb{P}_{lm} \to \bar{\Theta}^{in}_{fm}$ and $\sum_{l \in \mathcal{F}_{fm} \setminus k(j)} \mathbb{P}_{lm} \to \bar{\Theta}^{out}_{fm}$. Then $(\mathbf{A}_m \mathbf{b}_m)_j$ converges to

$$(\bar{\mathbf{A}}_m \bar{\mathbf{b}}_m)_j = \frac{\sigma^{k(j)}}{\beta_p} \left[ \frac{K_k(1-\sigma^{k(j)}) + \sigma^{k(j)}}{\beta_p} \bar{\Theta}^{in}_{fm} + \frac{\sigma^{k(j)}}{\beta_p} \bar{\Theta}^{out}_{fm} \right] \equiv \frac{\sigma^{k(j)}}{\beta_p} \bar{A},$$

so that $\mu_{jm} = \sigma^{k(j)}/\beta_p \, (1 + \bar{A}) + O(J^{-2})$. Summing the full geometric series gives the exact constant:

$$p_{jm} = mc_{jm} + \bar{\mu} + O(J^{-1}), \qquad \bar{\mu} = \frac{\sigma^{k(j)}/\beta_p}{1 - \bar{S}_{fm}/\beta_p}, \tag{37}$$

where $\bar{A} = \frac{K_k(1-\sigma^{k(j)})+\sigma^{k(j)}}{\beta_p}\,\bar{\Theta}^{in}_{fm} + \frac{\sigma^{k(j)}}{\beta_p}\,\bar{\Theta}^{out}_{fm}$. Competition ($\bar{\Theta}^{in}_{fm} = \bar{\Theta}^{out}_{fm} = 0$) gives $\bar{\mu} = \sigma^{k(j)}/\beta_p$; monopoly ($\bar{\Theta}^{in}_{fm} = 1/K_k$, $\bar{\Theta}^{out}_{fm} = (K_k - 1)/K_k$) gives $\bar{\mu} = \sigma^{k(j)}/(\beta_p - 1)$; any intermediate ownership yields a markup strictly between the two.

**Bertrand–Nash Competition with Nested CES.** Under nested CES with nest-specific within-group substitution elasticity $\sigma^{k(j)} > 1$ and across-nest elasticity $\sigma_0 \geqslant 1$, the own-price elasticity of product $j$ is:

$$\varepsilon_{jm} = \sigma^{k(j)}\big(1 - \mathbb{P}_{j|k(j),m}\big) + \sigma_0\,\mathbb{P}_{j|k(j),m}\big(1 - \mathbb{P}_{k(j),m}\big). \tag{38}$$

Note that when $J$ grows large, we can write:

$$\varepsilon_{jm} = \sigma^{k(j)} + \mathcal{O}_p(J^{-1}), \tag{39}$$

which implies:

$$\frac{\varepsilon_{jm}}{\varepsilon_{jm} - 1} = \frac{\sigma^{k(j)}}{\sigma^{k(j)} - 1} + \mathcal{O}_p(J^{-1}).$$

Defining $\lambda_{k(j)} \equiv \sigma^{k(j)}/(\sigma^{k(j)} - 1)$, it follows:

$$p_{jm} = \lambda_{k(j)}\, mc_{jm} + \mathcal{O}_p(J^{-1}). \tag{40}$$

Substituting (28) into (40):

$$p_{jm} = \underbrace{\lambda_{k(j)}\,\boldsymbol{\gamma}_x^{mc,c(j)}}_{\boldsymbol{\gamma}_x^{k'(j)}}{}'\mathbf{x}_{jm} + \underbrace{\lambda_{k(j)}\,\boldsymbol{\gamma}_z^{mc,c(j)}}_{\boldsymbol{\gamma}_z^{k'(j)}}{}'\mathbf{z}_{jm} + \underbrace{\lambda_{k(j)}\,\kappa_m^{mc,c(j)}}_{\kappa_m^{k'(j)}} + \underbrace{\lambda_{k(j)}\,\omega_{jm}}_{v_{jm}} + \mathcal{O}_p(J^{-1}), \tag{41}$$

which is (5) with pricing partition $k'(j) = k(j) \cap c(j)$, coefficient vectors $\boldsymbol{\gamma}_x^{k'(j)} = \lambda_{k(j)}\,\boldsymbol{\gamma}_x^{mc,c(j)}$ and $\boldsymbol{\gamma}_z^{k'(j)} = \lambda_{k(j)}\,\boldsymbol{\gamma}_z^{mc,c(j)}$, and fixed effect $\kappa_m^{k'(j)} = \lambda_{k(j)}\,\kappa_m^{mc,c(j)}$.

### A.2.4 Other pricing models

**Perfect Competition.** Under perfect competition every firm is a price taker and equilibrium requires $p_{jm} = mc_{jm}$. There is no strategic markup, so the demand partition $k(j)$ plays no role. Substituting (28) directly into the pricing condition gives (5) with pricing

partition $k'(j) = c(j)$, coefficient vectors $\boldsymbol{\gamma}_x^{k'(j)} = \boldsymbol{\gamma}_x^{mc,c(j)}$ and $\boldsymbol{\gamma}_z^{k'(j)} = \boldsymbol{\gamma}_z^{mc,c(j)}$, fixed effect $\kappa_m^{k'(j)} = \kappa_m^{mc,c(j)}$, and residual $v_{jm} = \omega_{jm}$.

**Monopolistic Competition.** Under monopolistic competition, each firm has market power over its differentiated product but is small relative to the market. Under nested logit, the nest-level inclusive value is defined as:

$$IV_{k(j),m} \equiv \log \sum_{d \in B_{k(j)}} e^{\delta_{dm}/\sigma^{k(j)}},$$

and under nested CES the nest-level price index is:

$$P_{k(j),m} \equiv \left( \sum_{d \in B_{k(j)}} (p_{dm}/A_{dm})^{1-\sigma^{k(j)}} \right)^{1/(1-\sigma^{k(j)})}.$$

Under monopolistic competition, firm $f$ takes $IV_{k(j),m}$ (resp. $P_{k(j),m}$) as given when maximizing profits $(p_{jm} - mc_{jm})\mathbb{P}_{jm}$, ignoring the effect of its own price on the aggregate nest-level term. This yields the pricing condition:

$$p_{jm} = \frac{\varepsilon_{jm}^{MC}}{\varepsilon_{jm}^{MC} - 1} mc_{jm}, \tag{42}$$

where $\varepsilon_{jm}^{MC} \equiv -(\partial \mathbb{P}_{jm}/\partial p_{jm})^{MC}(p_{jm}/\mathbb{P}_{jm})$ is the own-price elasticity computed holding the nest-level term fixed. Under nested logit, holding $IV_{k(j),m}$ fixed means the within-nest share $\mathbb{P}_{j|k(j),m}$ responds to $p_{jm}$ but the nest-level share $\mathbb{P}_{k(j),m}$ does not, giving:

$$\varepsilon_{jm}^{MC} = \frac{\alpha\, p_{jm}}{\sigma^{k(j)}} \big(1 - \sigma^{k(j)}\, \mathbb{P}_{jm}\big).$$

Under nested CES, holding $P_{k(j),m}$ fixed eliminates the cross-nest substitution term, giving:

$$\varepsilon_{jm}^{MC} = \sigma^{k(j)} \big(1 - \mathbb{P}_{j|k(j),m}\big).$$

In both cases, as $J \to \infty$ the elasticity converges to the same nest-specific limit as under Bertrand–Nash: $\varepsilon_{jm}^{MC} \to \alpha p_{jm}/\sigma^{k(j)}$ under nested logit and $\varepsilon_{jm}^{MC} \to \sigma^{k(j)}$ under nested CES. The pricing equations therefore map to the same (5) as (33) and (41), with pricing parti-

tion $k'(j) = k(j) \cap c(j)$ and identical coefficient vectors and fixed effects. The difference between monopolistic competition and Bertrand–Nash arises only at order $\mathcal{O}(J^{-1})$: under Bertrand–Nash the firm internalizes the effect of $p_{jm}$ on the nest-level term, whereas under monopolistic competition it does not.

**Perfect Collusion and Partial Cartels.** Under full collusion, a single decision maker maximizes joint industry profits. The FOC is identical to (6) with the firm-specific portfolio $\mathcal{F}_{fm}$ replaced by the entire product set. In the large-market limit, all portfolio corrections vanish by the same argument as in the extension with multiproduct firms, and the pricing equation takes the same form as under Bertrand competition.

Under partial collusion, a subset of firms forms a cartel $\mathcal{C}$ and jointly maximizes cartel profits. The cartel portfolio replaces the firm portfolio in the FOC. In the large-market limit the portfolio corrections vanish as before, and the pricing partition is again $k'(j) = k(j) \cap c(j)$.

**Conduct Parameter Model.** Bresnahan's conduct parameter model parametrizes the degree of market power by a scalar $\theta \in [0, 1]$, interpolating between perfect competition ($\theta = 0$) and full collusion ($\theta = 1$):

$$p_{jm} = mc_{jm} + \theta\, \mu_{jm}^{\text{monopoly}} + (1 - \theta)\, \mu_{jm}^{\text{competitive}}.$$

Under the large-market limit, the monopoly and competitive markups converge to their respective nest-specific limiting values. For nested logit, the limiting markup is $\theta\bar{\mu}_{k(j)}^{\text{mono}} + (1-\theta)\bar{\mu}_{k(j)}^{\text{comp}}$, a nest-specific additive constant that depends on $\theta$ and $\sigma^{k(j)}$. For nested CES, the markup factor is $\theta\bar{\lambda}_{k(j)}^{\text{mono}} + (1-\theta)\bar{\lambda}_{k(j)}^{\text{comp}}$, a nest-specific multiplicative scalar. If $\theta$ is common across all products, the pricing partition remains $k'(j) = k(j) \cap c(j)$. If different groups are assigned different conduct parameters $\theta^{h(j)}$ under the menu approach of Bresnahan (1987), the conduct partition $h(j)$ enters and the pricing partition becomes $k'(j) = k(j) \cap c(j) \cap h(j)$.

**Cournot Competition.** Under Cournot competition, firms choose quantities rather than prices. For a single-product firm owning product $j$, the first-order condition equates marginal revenue to marginal cost:

$$p_{jm} + q_{jm} \frac{\partial p_{jm}}{\partial q_{jm}} = mc_{jm},$$

which rearranges to the Lerner condition:

$$\frac{p_{jm} - mc_{jm}}{p_{jm}} = \frac{\mathbb{P}_{jm}}{\varepsilon_{jm}}, \tag{43}$$

where $\varepsilon_{jm} > 0$ is the own-price elasticity of demand. Solving for price:

$$p_{jm} = \frac{\varepsilon_{jm}}{\varepsilon_{jm} - \mathbb{P}_{jm}} mc_{jm}. \tag{44}$$

As $J \to \infty$, $\mathbb{P}_{jm} \to 0$, so the Cournot markup factor satisfies:

$$\frac{\varepsilon_{jm}}{\varepsilon_{jm} - \mathbb{P}_{jm}} = 1 + \frac{\mathbb{P}_{jm}}{\varepsilon_{jm} - \mathbb{P}_{jm}} = 1 + \mathcal{O}_p(J^{-1}),$$

under both nested logit (where $\varepsilon_{jm}$ remains bounded away from zero for $\sigma^{k(j)} \in (0, 1]$) and nested CES (where $\varepsilon_{jm} \to \sigma^{k(j)} > 1$). Therefore:

$$p_{jm} = mc_{jm} + \mathcal{O}_p(J^{-1}). \tag{45}$$

Substituting (28) into (45):

$$p_{jm} = \boldsymbol{\gamma}_x^{mc,c(j)\prime} \mathbf{x}_{jm} + \boldsymbol{\gamma}_z^{mc,c(j)\prime} \mathbf{z}_{jm} + \kappa_m^{mc,c(j)} + \underbrace{\omega_{jm}}_{v_{jm}} + \mathcal{O}_p(J^{-1}), \tag{46}$$

which is (5) with pricing partition $k'(j) = c(j)$, exactly as under perfect competition.

**Equivalence with perfect competition in large markets.** Equation (46) establishes a key result: in the large-$J$ limit, Cournot competition is observationally equivalent to perfect competition. The Cournot markup is proportional to $\mathbb{P}_{jm}$, which vanishes as $J \to \infty$, whereas under Bertrand competition with nested logit or nested CES demand, markups converge to a strictly positive nest-specific constant $\mu_{k(j)}$ or scaling factor $\lambda_{k(j)}$.

**Cost-Plus Pricing and Regulated Prices.** Under cost-plus pricing, a firm sets price as a fixed markup over marginal cost:

$$p_{jm} = (1 + \mu^{k'(j)})\, mc_{jm},$$

where $\mu^{k'(j)} \geqslant 0$ is a group-specific markup rate. Substituting (28):

$$p_{jm} = \underbrace{(1+\mu^{k'(j)})\,\boldsymbol{\gamma}_x^{mc,c(j)}}_{\boldsymbol{\gamma}_x^{k'(j)}}{}'\mathbf{x}_{jm} + \underbrace{(1+\mu^{k'(j)})\,\boldsymbol{\gamma}_z^{mc,c(j)}}_{\boldsymbol{\gamma}_z^{k'(j)}}{}'\mathbf{z}_{jm} + \underbrace{(1+\mu^{k'(j)})\,\kappa_m^{mc,c(j)}}_{\kappa_m^{k'(j)}} + \underbrace{(1+\mu^{k'(j)})\,\omega_{jm}}_{v_{jm}}, \tag{47}$$

which is (5) with pricing partition $k'(j) \cap c(j)$. The demand partition $k(j)$ plays no role since the markup does not depend on demand primitives. As under nested CES, the markup scales the entire cost structure, generating the testable restriction that the ratio of any pricing coefficient to its cost counterpart equals $(1+\mu^{k'(j)})$ within each group.

When prices are set by a regulator as a function of observed costs, the pricing rule takes the form $p_{jm} = \phi^{k'(j)}(\mathbf{x}_{jm}, \mathbf{z}_{jm}) + \omega_{jm}$, where $\phi^{k'(j)}(\cdot)$ is a group-specific regulatory formula. If $\phi^{k'(j)}$ is approximated by a polynomial in $(\mathbf{x}_{jm}, \mathbf{z}_{jm})$, this maps directly to (5) with $v_{jm} = \omega_{jm}$ and pricing partition $k'(j)$ determined by which products face the same regulatory formula. As under perfect competition, the demand partition $k(j)$ plays no role.

# B Appendix for Statistical Theory of Estimator

## B.1 Assumptions on the nested-logit model

### B.1.1 Notation and definitions

For vectors $x$, $\|x\|_p$ denotes its $\ell^p$-norm $(\sum_{k=1}^K |x_k|^p)^{1/p}$ for $1 \leqslant p < \infty$ and $\|x\|_\infty = \max_{k\in\{1,\dots,K\}} |x_k|$. For a random vector $X$, $\|X\|_{L^p} := (\mathrm{E}\,|X|^p)^{1/p}$. We abuse notation slightly and use $m \in M$ to mean $m \in \{1,\dots,M\}$, and use $\kappa \in K^J$ to mean $\kappa \in \{1,\dots,K\}^J$, etc. $K$ is fixed throughout.

**Definition 1. (Subgaussian norm.)** The subgaussian norm of a scalar random variable $X$ is defined as

$$\|X\|_{\psi_2} := \inf\{K > 0 : \mathrm{E}\exp(X^2/K^2) \leqslant 2\}.$$

The subgaussian norm of a random vector $X$ is defined as

$$\|X\|_{\psi_2} = \sup_{\|u\|_2=1} \| X^\top u \|_{\psi_2}.$$

$X$ is said to be subgaussian if $\|X\|_{\psi_2} < \infty$.

If $\|X\|_{\psi_2} \leqslant K$, then $\mathbb{P}(|X| > u) \leqslant 2\exp(-cu^2/K^2)$ for any $u \geqslant 0$ where $c > 0$ is a universal constant.

**Definition 2. (Subexponential norm.)** Thesubexponential norm of a scalar random variable $X$ is defined as

$$\|X\|_{\psi_1} := \inf\{K > 0 : \mathrm{E}\exp(|X|/K) \leqslant 2\}.$$

### B.1.2 Assumptions on the data generating process

Consider the following assumptions on the nested logit model

**Assumption 1.** *Assume that*

***(i).*** $(k_j : j \in \mathbb{N})$ *is a deterministic sequence with support* $\{1,\dots,K\}$ *and* $\sum_{j=1}^J \mathbb{1}\{k_j = \kappa\}/J \to \varpi_\kappa \in (0,1)$ *for each* $\kappa \in K$.

***(ii).*** $(x, z, \xi) := (x_j, z_j, \xi_j : j \in \mathbb{N})$ *is iid across* $j$, *and* $p := ((p_{j,J})_{j\in J} : J \in \mathbb{N})$ *is a row-wise iid triangular array satisfying*

*(a)* $\mathrm{E}[\xi_j] = 0, \mathrm{E}[x_j\xi_j] = 0,$ *and* $\mathrm{E}[z_j\xi_j] = 0.$

*(b)* $\operatorname{var} \tilde{x}_j$ *is positive definite, where* $\tilde{x}_j := (x_j^\top, z_j^\top)^\top$.

*(c)* $(x_j^\top, z_j^\top, \xi_j, p_{j,J})^\top$ *is a subgaussian random vector for each* $J$*, and has subgaussian norm uniformly bounded over* $J \in \mathbb{N}$.

***(iii).*** *Market shares* $\mathbb{P} := ((\mathbb{P}_{j,J})_{j\in J})_{J\in\mathbb{N}}$ *are generated according to*

$$\begin{cases} \mathbb{P}_{j,J} = \dfrac{\exp(\delta_j/\sigma_{k_j J})}{\sum_{i=1}^J \mathbb{1}\{k_i=k\}\exp(\delta_i/\sigma_{k_j J})} \dfrac{\left(\sum_{i=1}^J \mathbb{1}\{k_i=k\}\exp(\delta_i/\sigma_{k_j J})\right)^{\sigma_{k_j J}}}{1+\sum_{\ell=1}^K \left(\sum_{i=1}^J \mathbb{1}\{k_i=\ell\}\exp(\delta_i/\sigma_{\ell J})\right)^{\sigma_{\ell J}}} & \\ & \text{for } j \in \{1,\dots,J\} \\ \text{where } \delta_{j,J} := \beta_x^\top x_j + \beta_p p_{j,J} + \xi_j & \end{cases}$$

*for some parameters* $((\sigma_{1J},\dots\sigma_{KJ})_{J\in\mathbb{N}}, \beta_x, \beta_p)$ *where* $\sigma_{\kappa J} \in (\underline{\sigma}, \bar{\sigma}) \subset (0,1)$ *for all* $\kappa \in K, J \in \mathbb{N}$ *and* $\beta := (\beta_x, \beta_p) \in \mathcal{B} \subset \mathbb{R}^{d_x+1}$ *with* $\mathcal{B}$ *bounded. Moreover,* $\sigma_{\kappa J} \to \sigma_*$ *for all* $\kappa \in K$ *as* $J \to \infty$*, with* $\sigma_{\kappa J} - \sigma_{\tilde{\kappa} J} = o(1/\log J)$ *for every* $\kappa \neq \tilde{\kappa}$.

***(iv)*** *For each market* $m \in \mathbb{N}$, $((x_{jm}, z_{jm}, \xi_{jm})_{j\in\mathbb{N}}, ((p_{jm,J})_{j\in J})_{J\in\mathbb{N}}, ((\mathbb{P}_{jm,J})_{j\in J})_{J\in\mathbb{N}})$ *is an iid copy of* $((x,z,\xi),p,\mathbb{P})$ *described above. The dataset of* $M$ *markets and* $J$ *products available to the analyst is*

$$(x_{jm}, z_{jm}, p_{jm,J}, \mathbb{P}_{jm,J} : j \in J, m \in M).$$

**Assumption 2.** *In addition to Assumption 1, also assume that prices admits the representation*

$$p_{j,J} = \pi_{k'_j,J}^\top x_j + \gamma_{k'_j,J}^\top z_j + \psi_{k'_j,J} + \omega_{j,J}, \quad j \in J \tag{48}$$

*where* $(k'_j : j \in \mathbb{N})$ *has support* $\{1,\dots,K'\}$*, with* $J^{-1}\sum_{j=1}^J \mathbb{1}(k'_j = \kappa) \to (0,1)$ *for each* $\kappa \in K'$*, and* $\psi_{k'_j}$ *and* $\omega_j$ *are latent random variables satisfying*

*(i)* $\operatorname{E}\omega_{j,J} = 0, \omega_{j,J}$ *is independent of* $(x_j, z_j, \xi_j)$, $\operatorname{E}[\omega_{j,J}\psi_{\kappa,J}] = 0$*, and* $\omega_{j,J}$ *are subgaussian, with uniformly bounded subgaussian norm over* $\kappa \in K', J \in \mathbb{N}$.

*(ii)* $\{\psi_{\kappa,J} : \kappa \in K', J \in \mathbb{N}\} \subset \Psi$ *almost surely for some bounded set* $\Psi \subset \mathbb{R}$*, and* $\{\pi_{\kappa,J} : \kappa \in K', J \in \mathbb{N}\}$ *and* $\{\gamma_{\kappa,J} : \kappa \in K', J \in \mathbb{N}\}$ *are bounded sets.*

*(iii) Cross-sectional dependence are restricted according to* $\sup_{J\in\mathbb{N}} \frac{1}{J}\sum_{j=1}^J \sum_{i=1}^J |\operatorname{E}[\omega_{j,J}\omega_{i,J}]| < L$ *for some* $L < \infty$.

*(iv) Separation in the grouped fixed effect representation:* $\liminf_{J\to\infty} \operatorname{var}(\psi_{kJ} - \psi_{\ell J}) > 0$ *for every* $(k,\ell) \in K'^2$ *and* $k \neq \ell$*, and* $\liminf_{J\to\infty} \operatorname{var}\left(\psi_{kmJ} - \psi_{\ell mJ} + x_{jm}^\top(\pi_{k,J} - \pi_{\ell,J}) + z_{jm}^\top(\gamma_{k,J} - \gamma_{\ell,J})\right) > 0$.

If Assumptions 1 and 2 holds, by applying logs to $J\mathbb{P}_{jm,J}$,

$$\log J\mathbb{P}_{jm,J} = \sigma_{k_j}^{-1}\delta_{jm,J} + \zeta_{k_jm,J}$$

$$\delta_{jm,J} := (\beta_x^\top x_j + \beta_p p_{jmJ} + \xi_{jm})$$

$$\zeta_{km,J} := \lambda_{km,J} + \log J = \log \frac{\left(\frac{1}{J}\sum_{i=1}^{J} \mathbb{1}\{k_i = k\} \exp(\delta_{im,J}/\sigma_k)\right)^{\sigma_{kJ}-1}}{\sum_{\ell=1}^{K} J^{\sigma_{\ell J}-\sigma_{kJ}} \left(\frac{1}{J}\sum_{i=1}^{J} \mathbb{1}\{k_i = \ell\} \exp(\delta_{im,J}/\sigma_\ell)\right)^{\sigma_{\ell J}}}, \quad k \in K$$

By substituting out $p_{jm,J}$ with the pricing equation from Assumption 2, one obtains

$$\begin{aligned}
\log \mathbb{P}_{jm,J} &= \sigma_{k_j}^{-1}(\beta_x^\top x_{jm} + \beta_p(\pi_{k'_j,J}^\top x_{jm} + \gamma_{k'_j,J}^\top z_{jm} + \psi_{k'_j m,J} + \omega_{jm,J}) + \xi_{jm}) + \zeta_{k_jm,J} \\
&= \left(\sigma_{k_j}^{-1}(\beta_x + \beta_p \pi_{k'_j,J})\right)^\top x_{jm} + (\sigma_{k_j}^{-1}\beta_p \gamma_{k'_j,J})^\top z_{jm} \\
&\quad + (\sigma_{k_j}^{-1}\beta_p \psi_{k'_j m,J} + \zeta_{k_jm,J}) + \sigma_{k_j}^{-1}(\beta_p \omega_{jm,J} + \xi_{jm})
\end{aligned}$$

We assume for now either $K' = 1$ or there exists a permutation of $k'$ such that $k_j = k'_j$ for all $j$, almost surely.

**Assumption 3.** *$k$ from assumption 1 with support $K$ and $k'$ from assumption 2 with support $K'$ satisfy either*

*(i) $K' = 1$.*

*(ii) $K' = K$ and $k = k'$ up to permutation, almost surely.*

Under this assumption, we may write the previous display formally as

$$\log \mathbb{P}_{jm,J} = \tilde{x}_{jm}^\top \theta_{k_j,J} + \breve{\alpha}_{k_jm,J} + v_{jm,J}, \quad \tilde{x}_{jm} := (x_{jm}^\top, z_{jm}^\top)^\top \tag{49}$$

## B.2 General Theory

The model 49 formally looks like the baseline model considered in Bonhomme & Manresa (2015) although it has 3 important differences: the triangular-array nature of the data, that $\breve{\alpha}_{g_it,N}$ may not be almost surely in a bounded set, and group-specific $\theta$.

We first consider the following simplified DGP

$$y_{it,N} = x'_{it}\theta_N + \breve{\alpha}_{g_i t,N} + v_{it,N} \tag{50}$$

$$= x'_{it}\theta_N + \alpha_{g_i t,N} + v_{it,N} + \varepsilon_{it,N} \tag{51}$$

where the only difference with 49 is that $\theta$ are homogeneous across groups. We assume that each $\breve{\alpha}_{\gamma t,N}$ writes as $\breve{\alpha}_{\gamma t,N} = \alpha_{\gamma t,N} + \varepsilon_{\gamma t,N}$ with $\{\alpha_{\gamma t,N}\}$ are all contained in a bounded set, and the tail probabilities of $\varepsilon_{\gamma t,N}$ are controlled (in a sense made precise in Assumption 4(h) below). This is motivated by the fact that under some regularity, $\breve{\alpha}_{km,J}$ from the nested logit model may be decomposed this way. As in Bonhomme and Manresa (2015), we first provide theory and proofs for the case when $\theta$ is homogeneous across group, and then generalize $\theta$ to be possibly group-specific afterwards. To reduce clutter, we will suppress the $N$ subscripts in $v_{it,N}$ if relevant $N$ is apparent from context.

### B.2.1 Consistency of coefficients $(\hat{\theta}, \hat{\alpha}_{\hat{g}_i t})$

The estimators considered here $\hat{\theta}, \hat{\alpha}, \hat{g}$ are exactly the same estimators in the main specification of Bonhomme & Manresa (2015):

$$(\hat{\theta}, \hat{\alpha}, \hat{g}) := (\hat{\theta}_{NT}, \hat{\alpha}_{NT}, \hat{g}_{NT}) := \arg \min_{(\vartheta, \mathrm{a}, \nu) \in \Theta \times \mathcal{A}^{GT} \times G} \sum_{i=1}^{N} \sum_{i=t}^{T} \left( y_{it,N} - x'_{it}\vartheta - \mathrm{a}_{\nu_i t} \right)^2$$

We show they are consistent as an intermediate step to prove the misclassification theorem in the next subsection.

Consider the following assumptions on 50:

**Assumption 4.** *Assume there exists a constant $M > 0$ such that*
*(a) $\{\theta_N : N \in \mathbb{N}\} \subset \Theta$ and $\{\alpha_{\gamma t,N} : \gamma \in G, t, N \in \mathbb{N}^2\} \subset \mathcal{A}$ for bounded sets $\Theta \subset \mathbb{R}^d$ and $\mathcal{A} \subset \mathbb{R}$.*
*(b) $\mathrm{E}\,\|x_{it}\|^2 < M$ for every $i, t$.*
*(c) $\mathrm{E}\, v_{it} = 0$ and $\mathrm{E}\, v_{it}^4 < M$ for every $i, t, N$.*
*(d) $|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{E}\left[v_{it}v_{is}x'_{it}x_{is}\right]| < M$ for every $N, T$.*
*(e) $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left[v_{it}v_{jt}\right]| < M$ for every $N, T$.*
*(f) $|\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{cov}(v_{it}v_{jt}, v_{is}v_{js})| < M$ for every $N, T$.*

*(g)* $\operatorname{plim}_{N,T\to\infty} \min_{\nu\in G} \widehat{\rho}(\nu) = \rho > 0$ *where* $\widehat{\rho}(\nu)$ *is the minimum eigenvalue of the matrix*

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_{g_i\wedge\nu_i,t}\right)\left(x_{it} - \bar{x}_{g_i\wedge\nu_i,t}\right)',$$

*and* $\bar{x}_{g_i\wedge\gamma_i,t}$ *denote the sample mean of* $\{x_{jt} : g_j = g_i, \nu_j = \nu_i, j \in N\}$.
*(h) For every* $\eta > 0$*, it holds that*

$$\sup_{\gamma\in G, t\in\mathbb{N}} \mathrm{P}(|\varepsilon_{\gamma t,N}| > \eta) = O(r(N)),$$

*with* $r(N) \to 0$ *as* $N \to \infty$.

Besides the triangular array aspect in (a), Assumptions 4(a)-(g) are identical to Assumption 1 in Bonhomme & Manresa (2015)

**Theorem 1.** *If Assumption 4 holds, then,*

$$\widehat{\theta} - \theta_N \to_p 0,$$
$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{\alpha}_{\widehat{g}_i t} - \alpha_{g_i t,N})^2 \to_p 0,$$

*as* $N, T \to \infty$ *and* $T = o(1/r(N))$.

The GFE estimator $\hat{\theta}, \hat{\alpha}, \hat{g}$ under the modified DGP minimizes,

$$\begin{aligned}\check{Q}(\vartheta, \mathrm{a}, \nu) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it,N} - x_{it}'\vartheta - \mathrm{a}_{\nu_i t}\right)^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{g_i t,N} + v_{it} + x_{it}'(\theta_N - \vartheta) + \alpha_{g_i t,N} - \mathrm{a}_{\nu_i t}\right)^2.\end{aligned}$$

Denote the same objects as in Appendix A of Bonhomme & Manresa (2015)

$$\widehat{Q}(\vartheta, \mathrm{a}, \nu) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it} + x_{it}'(\theta_N - \vartheta) + \alpha_{g_i t,N} - \mathrm{a}_{\nu_i t}\right)^2$$
$$\widetilde{Q}(\vartheta, \mathrm{a}, \nu) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}'(\theta_N - \vartheta) + \alpha_{g_i t} - \mathrm{a}_{\nu_i t}\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2.$$

**Lemma 1.** *Under Assumption 4, with 4(h) holding for $r(N)$,*

$$\sup_{\vartheta\in\Theta,\, \mathrm{a}\in\mathcal{A}^{GT},\, \nu\in G} |\check{Q}(\vartheta,\mathrm{a},\nu) - \widetilde{Q}(\vartheta,\mathrm{a},\nu)| \to_p 0.$$

*as $N,T\to\infty$ and $T = o(1/r(N))$.*

*Proof.* We bound

$$|\check{Q}(\vartheta,\mathrm{a},\nu) - \widetilde{Q}(\vartheta,\mathrm{a},\nu)| \leqslant |\check{Q}(\vartheta,\mathrm{a},\nu) - \widehat{Q}(\vartheta,\mathrm{a},\nu)| + |\widehat{Q}(\vartheta,\mathrm{a},\nu) - \widetilde{Q}(\vartheta,\mathrm{a},\nu)|.$$

Lemma A.1 in Bonhomme & Manresa (2015) implies the second term converges to $0$ in probability uniformly as $N,T\to\infty$. The first term is

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{g_it,N}^2 + \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{g_it,N}\big(v_{it} + x_{it}'(\theta_N - \vartheta) + \alpha_{g_it,N} - \mathrm{a}_{\nu_it}\big).$$

By Cauchy-Schwarz Inequality, the absolute value of the second term is bounded by

$$\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{g_it,N}^2}\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(v_{it} + x_{it}'(\theta_N - \vartheta) + \alpha_{g_it,N} - \mathrm{a}_{\nu_it}\big)^2}$$

and moreover,

$$\begin{aligned}
\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{g_it,N}^2} &= \sqrt{\sum_{\gamma=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}\{g_i=\gamma\}\varepsilon_{\gamma t,N}^2} \\
&\leqslant \sqrt{\sum_{\gamma=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{\gamma t,N}^2} \\
&\leqslant \sqrt{G}\max_{\gamma\in G, t\in T}|\varepsilon_{\gamma t,N}|
\end{aligned}$$

Thus

$$\begin{aligned}
&\sup_{\vartheta\in\Theta,\,\mathrm{a}\in\mathcal{A}^{GT},\,\nu\in G}|\check{Q}(\vartheta,\mathrm{a},\nu)-\widetilde{Q}(\vartheta,\mathrm{a},\nu)|\\
&\leqslant 2\sqrt{G}\left(\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|\right)\sup_{\vartheta\in\Theta,\,\mathrm{a}\in\mathcal{A}^{GT},\,\nu\in G}\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(v_{it}+x_{it}'(\theta_N-\vartheta)+\alpha_{g_it,N}-\mathrm{a}_{\nu_it}\right)}\\
&\quad+G\left(\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|^2\right)
\end{aligned}$$

By Assumption 1(a)-(c), the sup term is $O_{\mathrm{P}}(1)$. Note that by the union bound, for any $\eta>0$

$$\begin{aligned}
\mathrm{P}\left(\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|>\eta\right)&\leqslant\sum_{\gamma\in G,t\in T}\mathrm{P}\left(|\varepsilon_{\gamma t,N}|>\eta\right)\\
&\leqslant GT\sup_{\gamma\in G,t\in T}\mathrm{P}\left(|\varepsilon_{\gamma t,N}|>\eta\right)\\
&\leqslant GTO(r(N))
\end{aligned}$$

Assuming $T=o(1/r(N))$, it follows that this term is $o(1)$, i.e., $\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|$ and $\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|^2$ are $o_{\mathrm{P}}(1)$. Thus $\sup_{\vartheta\in\Theta,\,\mathrm{a}\in\mathcal{A}^{GT},\,\nu\in G}|\check{Q}(\vartheta,\mathrm{a},\gamma)-\widetilde{Q}(\vartheta,\mathrm{a},\gamma)|=o_{\mathrm{P}}(1)$, as desired. □

*Proof. (of Theorem 1.)* Observe that

$$\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\tilde{Q}(\theta_N,\alpha_N,g)=\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\check{Q}(\hat{\theta},\hat{\alpha},\hat{g})+\underbrace{\check{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\check{Q}(\theta_N,\alpha_N,g)}_{\leqslant 0}+\check{Q}(\theta_N,\alpha_N,g)-\tilde{Q}(\theta_N,\alpha_N,g)$$

which is $o_{\mathrm{P}}(1)$ by Lemma 1 above. Applying Lemma A.2 in Bonhomme & Manresa (2015) yields $\|\widehat{\theta}-\theta_N\|=o_{\mathrm{P}}(1)$. Also, we can bound

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\alpha_{g_it,N}-\hat{\alpha}_{\hat{g}_it})^2&=\tilde{Q}(\theta_N,\hat{\alpha},\hat{g})-\tilde{Q}(\theta_N,\alpha_N,g)\\
&=\tilde{Q}(\theta_N,\hat{\alpha},\hat{g})+\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\tilde{Q}(\theta_N,\alpha_N,g)\\
&\leqslant|\tilde{Q}(\theta_N,\hat{\alpha},\hat{g})+\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})|+|\tilde{Q}(\hat{\theta},\hat{\alpha},\hat{g})-\check{Q}(\hat{\theta},\hat{\alpha},\hat{g})|\\
&\quad+|\check{Q}(\theta_N,\alpha_N,g)-\tilde{Q}(\theta_N,\alpha_N,g)|
\end{aligned}$$

where the final inequality uses the penultimate display. The first term is shown to be $o_{\mathrm{P}}(1)$ in proof of Theorem 1 in Bonhomme & Manresa (2015). The last two terms are $o_{\mathrm{P}}(1)$ by Lemma

1. Thus, we conclude that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\alpha_{g_i t,N} - \hat{\alpha}_{\hat{g}_i t})^2 = o_{\mathrm{P}}(1)$. □

### B.2.2 Consistency of classification $\hat{g}_i$

**Assumption 5.**

*(a) For all $\gamma \in G$, it holds* $\operatorname{plim}_{N\to\infty} N^{-1}\sum_{i=1}^{N} \mathbb{1}_{g_i=\gamma} > 0$.

*(b) For all $(\gamma, \widetilde{\gamma}) \in G^2$ with $\gamma \neq \widetilde{\gamma}$, it holds* $\liminf_{N\to\infty} \operatorname{plim}_{T\to\infty} T^{-1}\sum_{t=1}^{T}(\alpha_{\gamma t,N} - \alpha_{\widetilde{\gamma} t,N})^2 > 0$.

*(c) For all $(\gamma, \widetilde{\gamma}) \in G^2$ with $\gamma \neq \widetilde{\gamma}$ and $N \in \mathbb{N}$, it holds that $(v_{itN})_{t\in\mathbb{N}}$, $(\alpha_{\gamma tN} - \alpha_{\widetilde{\gamma} tN})_{t\in\mathbb{N}}$, $((\alpha_{\gamma tN} - \alpha_{\widetilde{\gamma} tN})v_{itN})_{t\in\mathbb{N}}$ are $\alpha$-strongly mixing processes with coefficient $\alpha[t] \leqslant \exp(-at^{d_1})$. Moreover* $\mathrm{E}[(\alpha_{\gamma tN} - \alpha_{\widetilde{\gamma} tN})v_{itN}] = 0$ *for every $N \in \mathbb{N}$.*

*(d) There exists constants $b > 0$ and $d_2 > 0$ such that* $\mathrm{P}(|v_{it}| > m) \leqslant \exp(1 - (m/b)^{d_2})$ *for all $i, t$ and $m > 0$.*

*(e) There exists a constant $M^* > 0$ such that, as $N, T \to \infty$,*

$$\max_{i\in N} \mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\| \geqslant M^*\right) = o(T^{-\delta}) \quad \text{for all } \delta > 0.$$

**Theorem 2.** *Under Assumptions 4 and 5, for every $\delta > 0$,*

$$\mathrm{P}\left(\max_{i\in N}\left|\widehat{g}_i(\widehat{\theta}, \widehat{\alpha}) - g_i\right| > 0\right) = o(1) + o(NT^{-\delta}),$$

*as $N, T \to \infty$ and $T = o(1/r(N))$.*

**Lemma 2.** *For $\eta > 0$ small enough, we have, for all $\delta > 0$*

$$\max_{i\in N} \mathrm{P}\left((\widehat{\theta}, \widehat{\alpha}) \in \mathcal{N}_{NT}(\eta),\ \widehat{g}_i(\widehat{\theta}, \widehat{\alpha}) \neq g_i,\ \max_{\gamma\in G, t\in T}|\varepsilon_{\gamma t,N}| \leqslant \sqrt{\eta}\right) = o(T^{-\delta}) \text{ as } N, T \to \infty.$$

*where* $\mathcal{N}_{NT}(\eta) := \{(\vartheta, \mathrm{a}) : \|\vartheta - \theta_N\|^2 < \eta$ *and* $\frac{1}{T}\sum_{t=1}^{T}(\alpha_{\gamma t,N} - \mathrm{a}_{\gamma t})^2 < \eta \quad \forall \gamma \in G\}$

*Proof.* Write the probability of interest as

$$\max_{i\in N} \mathrm{E}\left[\mathbb{1}\left(\{(\widehat{\theta}, \widehat{\alpha}) \in \mathcal{N}_{NT}(\eta)\} \cap \{\max_{\gamma\in G, t\in T}|\varepsilon_{\gamma t,N}| \leqslant \sqrt{\eta}\}\right)\sum_{\gamma=1}^{G} Z_{i\gamma}(\hat{\theta}, \hat{\alpha})\right]$$

where

$$
\begin{aligned}
Z_{i\gamma}(\vartheta, \mathrm{a}) &= \mathbb{1}\{g_i \neq \gamma\}\, \mathbb{1}\left\{ \sum_{t=1}^{T} (\mathrm{a}_{g_i t} - \mathrm{a}_{\gamma t}) \left( \varepsilon_{g_i t,N} + v_{it} + x_{it}' (\theta_N - \vartheta) + \alpha_{g_i t,N} - \frac{\mathrm{a}_{\gamma t} + \mathrm{a}_{g_i t}}{2} \right) \leqslant 0 \right\} \\
&\leqslant \max_{\tilde{\gamma} \neq \gamma} \mathbb{1}\left\{ \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \left( \varepsilon_{g_i t,N} + v_{it} + x_{it}' (\theta_N - \vartheta) + \alpha_{g_i t,N} - \frac{\mathrm{a}_{\gamma t} + \mathrm{a}_{\tilde{\gamma} t}}{2} \right) \leqslant 0 \right\} \\
&\leqslant \max_{\tilde{\gamma} \neq \gamma} \mathbb{1}\left\{ \sum_{t=1}^{T} (\alpha_{\tilde{\gamma} t,N} - \alpha_{\gamma t,N}) \left( v_{it} + \alpha_{\tilde{\gamma} t,N} - \frac{\alpha_{\gamma t,N} - \alpha_{\tilde{\gamma} t,N}}{2} \right) \leqslant \check{A}_T \right\} \\
&\leqslant \max_{\tilde{\gamma} \neq \gamma} \mathbb{1}\left\{ \sum_{t=1}^{T} (\alpha_{\tilde{\gamma} t,N} - \alpha_{\gamma t,N}) \left( v_{it} + \alpha_{\tilde{\gamma} t,N} - \frac{\alpha_{\gamma t,N} - \alpha_{\tilde{\gamma} t,N}}{2} \right) \leqslant A_T + \left| \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \varepsilon_{g_i t,N} \right| \right\}
\end{aligned}
$$

where we define

$$
\begin{aligned}
\check{A}_T &= \left| \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \left( \varepsilon_{g_i t,N} + v_{it} + x_{it}' (\theta_N - \vartheta) + \alpha_{g_i t,N} - \frac{\mathrm{a}_{\gamma t} + \mathrm{a}_{\tilde{\gamma} t}}{2} \right) \right. \\
&\quad \left. - \sum_{t=1}^{T} (\alpha_{\tilde{\gamma} t,N} - \alpha_{\gamma t,N}) \left( v_{it} + \alpha_{\tilde{\gamma} t,N} - \frac{\alpha_{\gamma t,N} - \alpha_{\tilde{\gamma} t,N}}{2} \right) \right| \\
&= \left| \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \left( v_{it} + x_{it}' (\theta_N - \vartheta) + \alpha_{g_i t,N} - \frac{\mathrm{a}_{\gamma t} + \mathrm{a}_{\tilde{\gamma} t}}{2} \right) \right. \\
&\quad \left. - \sum_{t=1}^{T} (\alpha_{\tilde{\gamma} t,N} - \alpha_{\gamma t,N}) \left( v_{it} + \alpha_{\tilde{\gamma} t,N} - \frac{\alpha_{\gamma t,N} - \alpha_{\tilde{\gamma} t,N}}{2} \right) + \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \varepsilon_{g_i t,N} \right| \\
&\leqslant A_T + \left| \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \varepsilon_{g_i t,N} \right|,
\end{aligned}
$$

with $A_T$ is identical to the one in the proof of Lemma B4 in Bonhomme & Manresa (2015). Moreover, by Cauchy Schwarz Inequality, for $(\vartheta, \mathrm{a}) \in \mathcal{N}_\eta$, we can bound

$$
\left| \sum_{t=1}^{T} (\mathrm{a}_{\tilde{\gamma} t} - \mathrm{a}_{\gamma t}) \varepsilon_{g_i t,N} \right| \leqslant T C_4 \max_{\gamma \in G, t \in T} |\varepsilon_{\gamma t,N}|.
$$

In the event $\{(\widehat{\theta}, \widehat{\alpha}) \in \mathcal{N}_{NT}(\eta)\} \cap \{\max_{\gamma \in G, t \in T} |\varepsilon_{\gamma t,N}| \leqslant \sqrt{\eta}\}$, we therefore have, as in the proof

of Lemma B4 in Bonhomme & Manresa (2015), that

$$Z_{i\gamma}(\hat{\theta},\hat{\alpha}) \leqslant \max_{\tilde{\gamma}\neq\gamma} \mathbb{1}\left\{ \sum_{t=1}^{T}(\alpha_{\tilde{\gamma}t,N}-\alpha_{\gamma t,N})\left(v_{it}+\alpha_{\tilde{\gamma}t,N}-\frac{\alpha_{\gamma t,N}-\alpha_{\tilde{\gamma}t,N}}{2}\right)\right.$$
$$\left. + TC_1\sqrt{\eta}\left(\frac{1}{T}\sum_{t=1}^{T} v_{it}^2\right)^{1/2} + TC_2\sqrt{\eta}\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|\right) + T(C_3+C_4)\sqrt{\eta}\right\} =: \check{Z}_{ig}.$$

Following the steps in the proof of Lemma B4 in Bonhomme & Manresa (2015), for $\widetilde{M} > \max(\sqrt{M}, M^*)$ given by Assumptions 1 and 2.e.,

$$\mathrm{P}\left(\check{Z}_{i\gamma}=1\right) \leqslant \sum_{\tilde{\gamma}\neq\gamma}\left[\mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\| \geqslant \widetilde{M}\right) + \mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}t,N}-\alpha_{\gamma t,N})^2 \leqslant \frac{c_{\gamma,\tilde{\gamma}}}{2}\right) + \mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T} v_{it}^2 \geqslant \widetilde{M}\right)\right.$$
$$\left. + \mathrm{P}\left(\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}t,N}-\alpha_{\gamma t,N})v_{it} \leqslant -T\frac{c_{\gamma,\tilde{\gamma}}}{4} + TC_1\sqrt{\widetilde{M}}\sqrt{\eta} + TC_1\widetilde{M}\sqrt{\eta} + T(C_3+C_4)\sqrt{\eta}\right)\right].$$

By applying Lemma B5 in Bonhomme & Manresa (2015), for every $\delta > 0$, we can bound

$$\mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}t,N}-\alpha_{\gamma t,N})^2 \leqslant \frac{c_{\gamma,\tilde{\gamma}}}{2}\right) = o(T^{-\delta}),$$
$$\mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T} v_{it}^2 \geqslant \widetilde{M}\right) = o(T^{-\delta}),$$
$$\mathrm{P}\left(\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}t,N}-\alpha_{\gamma t,N})v_{it} \leqslant -T\frac{c_{\gamma,\tilde{\gamma}}}{4} + TC_1\sqrt{\widetilde{M}}\sqrt{\eta} + TC_1\widetilde{M}\sqrt{\eta} + T(C_3+C_4)\sqrt{\eta}\right)\Bigg] = o(T^{-\delta}).$$

Using this and Assumption (2e), the probability of interest is thus bounded as

$$\max_{i\in N}\sum_{\gamma=1}^{G}\mathrm{P}\left(\check{Z}_{i\gamma}=1\right) \leqslant G(G-1)\max_{i\in N}\mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\| \geqslant \widetilde{M}\right) + o(T^{-\delta}) = o(T^{-\delta}).$$

□

*Proof. (of Theorem 2.)* Estimate the probability of interest as

$$\begin{aligned}
\mathrm{P}\left(\max_{i\in N}\left|\widehat{g}_i(\widehat{\theta},\widehat{\alpha})-g_i\right|>0\right) &= \mathrm{P}\left(\left(\bigcup_{i=1}^{N}\{\widehat{g}_i(\widehat{\theta},\widehat{\alpha})\neq g_i\}\right)\cap\left\{(\widehat{\theta},\widehat{\alpha})\in\mathcal{N}_{NT}(\eta)\right\}\cap\left\{\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|\leqslant\sqrt{\eta}\right\}\right)\\
&\leqslant \mathrm{P}\left((\widehat{\theta},\widehat{\alpha})\notin\mathcal{N}_{NT}(\eta)\right)+\mathrm{P}\left(\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|>\sqrt{\eta}\right)\\
&\quad+\mathrm{P}\left(\left(\bigcup_{i=1}^{N}\{\widehat{g}_i(\widehat{\theta},\widehat{\alpha})\neq g_i\}\right)\cap\left\{(\widehat{\theta},\widehat{\alpha})\in\mathcal{N}_{NT}(\eta)\right\}\cap\left\{\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|\leqslant\sqrt{\eta}\right\}\right)
\end{aligned}$$

We bound the three terms of RHS. By applying the Lemma above, the third term is bounded by

$$\begin{aligned}
&\mathrm{P}\left(\left(\bigcup_{i=1}^{N}\{\widehat{g}_i(\widehat{\theta},\widehat{\alpha})\neq g_i\}\right)\cap\left\{(\widehat{\theta},\widehat{\alpha})\in\mathcal{N}_{NT}(\eta)\right\}\cap\left\{\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|\leqslant\sqrt{\eta}\right\}\right)\\
&\qquad\leqslant N\max_{i\in N}\mathrm{P}\left(\widehat{g}_i(\widehat{\theta},\widehat{\alpha})\neq g_i,\ (\widehat{\theta},\widehat{\alpha})\in\mathcal{N}_{NT}(\eta)\ \max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|\leqslant\sqrt{\eta}\right)\\
&\qquad= o(NT^{-\delta})
\end{aligned}$$

By Theorem 1, the first term $\mathrm{P}\left((\widehat{\theta},\widehat{\alpha})\notin\mathcal{N}_{NT}(\eta)\right)=o(1)$. Finally, using Assumption 4(h), we bound the second term

$$\mathrm{P}\left(\max_{\gamma\in G,t\in T}|\varepsilon_{\gamma t,N}|>\sqrt{\eta}\right)\leqslant GT\sup_{\gamma\in G,t\in T}\mathrm{P}\left(|\varepsilon_{\gamma t,N}|>\eta\right)\leqslant GTO(r(N))$$

which is $o(1)$ as $N,T\to\infty$ and $T=o(1/r(N))$. □

## B.3 Extension to model with group-specific coefficients

Consider the additional assumptions on the data-generating process, mostly taken from Assumption S.2. of Bonhomme & Manresa 2015.

**Assumption 6.**

*(a)* $\{\theta_{\gamma N}:\gamma\in G,N\in\mathbb{N}\}\subset\Theta$ *for bounded set* $\Theta\subset\mathbb{R}^d$.

*(b) There exists* $\hat{\rho}\to_p\rho>0$ *such that for every* $\gamma\in G$*, it holds*

$$\min_{\nu\in G}\max_{\tilde{\gamma}\in G}\hat{\rho}(\nu,\gamma,\tilde{\gamma})\geqslant\hat{\rho}$$

*where $\widehat{\rho}(\nu,\gamma,\tilde{\gamma})$ is the minimum eigenvalue of the matrix*

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{g_i=\gamma\}\cap\{\nu_i=\tilde{\gamma}\}}\begin{pmatrix}\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{it}' & \frac{1}{\sqrt{T}}x_{i1} & \cdots & \frac{1}{\sqrt{T}}x_{iT}\\ \frac{1}{\sqrt{T}}x_{iT}' & 1 & 0 & 0\\ \vdots & 0 & \ddots & 0\\ \frac{1}{\sqrt{T}}x_{iT}' & 0 & 0 & 1\end{pmatrix}.$$

*(c) For every $(\gamma,\tilde{\gamma})\in G^2$, $\gamma\neq\tilde{\gamma}$ there exists $c^{\dagger}_{\gamma,\tilde{\gamma}}>0$ such that*

$$\liminf_{N\to\infty}\min_{i\in N}\operatorname*{plim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}(x_{it}^{\top}(\theta_\gamma-\theta_{\tilde{\gamma}})+\alpha_{\gamma t,N}-\alpha_{\tilde{\gamma}t,N})^2\geqslant c^{\dagger}_{\gamma,\tilde{\gamma}},$$
$$\operatorname*{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}^{\top}(\theta_\gamma-\theta_{\tilde{\gamma}})+\alpha_{\gamma t,N}-\alpha_{\tilde{\gamma}t,N})^2\geqslant c^{\dagger}_{\gamma,\tilde{\gamma}},$$

*(d) For every $\eta,\delta>0$, it holds that*

$$\max_{i\in N}\mathrm{P}\left(\left\|\frac{1}{T}\sum_{i=1}^{T}v_{it}x_{it}\right\|_2\geqslant\eta\right)=o(T^{-\delta}).$$

Consider the following estimator:

$$(\widehat{\theta}^{\dagger},\widehat{\alpha}^{\dagger},\widehat{g}^{\dagger})=\arg\min_{(\vartheta,\mathrm{a},\nu)\in\Theta^G\times\mathcal{A}^{GT}\times G^N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}-x_{it}^{\top}\vartheta_{\nu_i}-\mathrm{a}_{\nu_i t}\right)^2.$$

**Theorem 3.** *If Assumptions 4, 5 and 6 hold, then*

$$\mathrm{P}\left(\max_{i\in N}|\hat{g}_i^{\dagger}-g_i|>0\right)=o(1)+o(NT^{-\delta})\quad\textit{for any }\delta>0.$$

Let

$$\begin{aligned}\check{Q}(\vartheta,\mathrm{a},\nu)&:=\sum_{i=1}^{N}\sum_{i=t}^{T}\left(y_{it,N}-x_{it}'\vartheta_{\nu_i}-\mathrm{a}_{\nu_i t}\right)^2\\&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\varepsilon_{g_i t,N}+v_{it}+x_{it}'(\theta_{g_i,N}-\vartheta_{\nu_i})+\alpha_{g_i t}-\mathrm{a}_{\nu_i t}\right)^2.\end{aligned}$$

and

$$\widehat{Q}(\vartheta, \mathrm{a}, \nu) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( v_{it} + x_{it}'(\theta_{g_i,N} - \vartheta_{\nu_i}) + \alpha_{g_i t} - \mathrm{a}_{\nu_i t} \right)^2$$

$$\widetilde{Q}(\vartheta, \mathrm{a}, \nu) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( x_{it}'(\theta_{g_i N} - \vartheta_{\nu_i}) + \alpha_{g_i t} - \mathrm{a}_{\nu_i t} \right)^2 + \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} {v_{it}}^2.$$

As in the proof of Theorem 1, we start by showing the following uniform convergence result.

**Lemma 3.** ***(Modified Lemma S.3)*** *Under Assumption 6,*

$$\sup_{(\vartheta, \mathrm{a}, \nu) \in \Theta^G \times \mathcal{A}^{GT} \times G} |\check{Q}(\vartheta, \mathrm{a}, \nu) - \widetilde{Q}(\vartheta, \mathrm{a}, \nu)| \to_p 0.$$

*as* $N, T \to \infty$ *and* $T = o(1/r(N))$.

*Proof.* We bound

$$|\check{Q}(\vartheta, \mathrm{a}, \nu) - \widetilde{Q}(\vartheta, \mathrm{a}, \nu)| \leqslant |\check{Q}(\vartheta, \mathrm{a}, \nu) - \widehat{Q}(\vartheta, \mathrm{a}, \nu)| + |\widehat{Q}(\vartheta, \mathrm{a}, \nu) - \widetilde{Q}(\vartheta, \mathrm{a}, \nu)|.$$

And note that the second term is uniformly $o_P(1)$ as $N, T \to \infty$ by Lemma S.3 in BM15. The first term writes as

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{g_i t, N}^2 + \frac{2}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{g_i t, N} \left( v_{it} + x_{it}'(\theta_{g_i} - \vartheta_{\nu_i}) + \alpha_{g_i t} - \mathrm{a}_{\nu_i t} \right).$$

The analogous argument as in the proof of Theorem 1 leads to

$$\begin{aligned}
&\sup_{(\vartheta, \mathrm{a}, \nu) \in \Theta^G \times \mathcal{A}^{GT} \times G} |\check{Q}(\vartheta, \mathrm{a}, \nu) - \widetilde{Q}(\vartheta, \mathrm{a}, \nu)| \\
&\leqslant 2\sqrt{G} \left( \max_{\gamma \in G, t \in T} |\varepsilon_{\gamma t, N}| \right) \sup_{(\vartheta, \mathrm{a}, \nu) \in \Theta^G \times \mathcal{A}^{GT} \times G^N} \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( v_{it} + x_{it}'(\theta_{g_i} - \vartheta_{\nu_i}) + \alpha_{g_i t} - \mathrm{a}_{\nu_i t} \right)} \\
&\quad + G \left( \max_{\gamma \in G, t \in T} |\varepsilon_{\gamma t, N}|^2 \right)
\end{aligned}$$

By Assumption 6, the sup term is $O_{\mathrm{P}}(1)$. By the union bound and $\mathrm{P}\left( \max_{\gamma \in G, t \in T} |\varepsilon_{\gamma t, N}| > \eta \right) \leqslant GTO(r(N))$, which is $o(1)$ when $T = o(1/r(N))$. Thus $\sup_{(\vartheta, \mathrm{a}, \nu) \in \Theta^G \times \mathcal{A}^{GT} \times G^N} |\check{Q}(\vartheta, \mathrm{a}, \gamma) -$

$\widetilde{Q}(\vartheta, \mathrm{a}, \gamma)| = o_{\mathrm{P}}(1)$, as desired. □

*Proof.* (of Theorem 3) The preceding lemma establishes uniform asymptotic equivalence between $\check{Q}$ and $\widetilde{Q}$. By Lemma S.4 in Bonhomme & Manresa (2015), which relies on Assumption6, it follows that for every $\gamma \in G$, that as $N, T \to \infty$

$$\|\hat{\theta}^{\dagger}_{\gamma} - \theta_{\gamma N}\|_2^2 + \frac{1}{T}\sum_{t=1}^{T}(\hat{\alpha}^{\dagger}_{\gamma tN} - \alpha_{\gamma tN})^2 \to_p 0.$$

The rest of the proof follows analogously as in Theorem 3; the only difference is to show

$$\mathrm{P}\left(\frac{1}{T}\sum_{t=1}^{T}\left(x_{it}'(\theta_{\tilde{\gamma}N} - \theta_{\gamma N}) + \alpha_{\tilde{\gamma}tN} - \alpha_{\gamma tN}\right)v_{it} \leqslant -\frac{c^{\dagger}_{\gamma,\tilde{\gamma}}}{8}\right) = o(T^{-\delta}),$$

rather than just $\mathrm{P}\left(T^{-1}\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}tN} - \alpha_{\gamma tN})v_{it} \leqslant -c_{\gamma,\tilde{\gamma}}/8)\right) = o(T^{-\delta})$ as in 3. The previous display holds due to Assumption 5(c), and Lemma B5 of Bonhomme & Manresa (2015), implying

$$\mathrm{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}(\alpha_{\tilde{\gamma}tN} - \alpha_{\gamma tN})v_{it}\right| \geqslant \frac{c^{\dagger}_{\gamma,\tilde{\gamma}}}{16}\right) = o(T^{-\delta}),$$

and also that by Assumption 6(a) and (d),

$$\mathrm{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T}v_{it}x_{it}^{\top}(\theta_{\tilde{\gamma}N} - \theta_{\gamma N})\right| \geqslant \frac{c^{\dagger}_{\gamma,\tilde{\gamma}}}{16}\right) = o(T^{-\delta}).$$

The rest of the proof follows analogously as in the proof of Theorem 3. □

## B.4 Verifying assumptions of the general theory for the nested logit model

We verify that a nested logit model satisfying Assumptions 1, 2 and 3 verifies Assumptions 4, 5 and 6 for an appropriate reduced-form equation. Thus, the general theory applies, implying that the clustering algorithm based on Bonhomme & Manresa (2015) applied to the reduced-form equation yields consistent classification.

**Theorem 4.** *If Assumptions 1, 2 and 3 holds, then the nested logit model admits the following*

*reduced-form equation*

$$\begin{aligned}\log \mathbb{P}_{jm,J} &= \left(\sigma_{k_j}^{-1}(\beta_x + \beta_p \pi_{k'_j,J})\right)^\top x_{jm} + (\sigma_{k_j}^{-1}\beta_p \gamma_{k'_j,J})^\top z_{jm} \\ &\quad + (\sigma_{k_j}^{-1}\beta_p \psi_{k'_j m,J} + \zeta_{k_j m,J}) + \sigma_{k_j}^{-1}(\beta_p \omega_{jm,J} + \xi_{jm}) \\ &= \tilde{x}_{jm}^\top \theta_{k_j,J} + \breve{\alpha}_{k_j m,J} + v_{jm,J} \\ \tilde{x}_{jm} &:= (x_{jm}^\top, z_{jm})^\top.\end{aligned} \tag{52}$$

*Moreover, the clustering algorithm Bonhomme & Manresa 2015 applied to this context yields*

$$\mathrm{P}\left(\max_{j \in J} \left|\widehat{k}_j(\widehat{\theta}, \widehat{\alpha}) - k_j\right| > 0\right) = o(1) + o(JM^{-\varsigma}), \text{ for every } \varsigma > 0,$$

*as* $J, M \to \infty$ *with* $M = o(J^q)$ *for any* $q > 0$.

The arguments verifying Assumptions 4, 5 and 6 from Assumptions 1, 2 and 3 are contained in the proofs of the following proposition.

**Proposition 1.** *If Assumptions 1, 2 and 3 holds, then the nested logit model admits the reduced-form equation (52). Moreover, Assumption 4 is satisfied for this equation, and Assumption 4(h) is satisfied for* $r(J) = J^{-q}$ *for any* $q > 0$.

*Proof.* **(Verification of Assumption 4).**

*Assumption 4(a).* Immediate from assumption 1(ii) and assumption 2(ii).

*Assumption 4(b).* Immediate from assumption 1(ii)(b).

*Assumption 4(c).* Immediate from assumption 1(ii)(a) and (b).

*Assumption 4(d).* By assumption 1(ii) and (iv),

$$\begin{aligned}\mathrm{E}[v_{jm} v_{in} \tilde{x}_{jm}^\top \tilde{x}_{in}] &= \mathrm{E}[v_{jm} \tilde{x}_{jm}^\top] \, \mathrm{E}[\tilde{x}_{in} v_{in}] \\ &= (\mathrm{E}[\sigma_{k_j} \xi_{jm} \tilde{x}_{jm}^\top] + \beta_p \, \mathrm{E}[\sigma_{k_j} \omega_{jm} \tilde{x}_{jm}^\top])^2 \\ &= 0\end{aligned}$$

*Assumption 4(e).* By identical distribution across $j, m$, we may simplify $M^{-1} \sum_{m=1}^M \mathrm{E}[v_{jm} v_{im}] = \mathrm{E}[v_{jm} v_{im}]$. Note that by Assumption 2(i), there is a simplification

$$\mathrm{E}[v_{jm} v_{im}] = \mathrm{E}[\sigma_{k_j}^{-1} \xi_{jm} \sigma_{k_i}^{-1} \xi_{im}] + \beta_p^2 \, \mathrm{E}[\sigma_{k_j}^{-1} \omega_{jm} \sigma_{k_i}^{-1} \omega_{im}].$$

Moreover, by the independence hypothesis of Assumption 1(i) and the conditional mean independence restriction in Assumption 1(ii)(a), it follows

$$\mathrm{E}[\sigma_{k_j}^{-1}\xi_{jm}\sigma_{k_i}^{-1}\xi_{im}] = \mathrm{E}[\sigma_{k_j}^{-1}\xi_{jm}]\,\mathrm{E}[\sigma_{k_i}^{-1}\xi_{im}] = (\mathrm{E}[\sigma_{k_j}\,\mathrm{E}[\xi_{jm}|k_j]])^2 = 0$$

We also have the bound

$$|\beta_p^2\,\mathrm{E}[\sigma_{k_j}^{-1}\omega_{jm}\sigma_{k_i}^{-1}\omega_{im}]| \leqslant \beta_p^2\underline{\sigma}^{-1}|\,\mathrm{E}[\omega_{jm}\omega_{im}]|.$$

Hence by assumption 2(iii)

$$\frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{J}|\,\mathrm{E}[v_{jm}v_{im}]| \leqslant \beta_p^2\underline{\sigma}^{-1}\frac{1}{J}\sum_{j=1}^{J}\sum_{i=1}^{J}|\,\mathrm{E}[\omega_{jm}\omega_{im}]| \leqslant \beta_p^2\underline{\sigma}^{-1}L < \infty$$

*Assumption 4(f).* Immediate because $\mathrm{cov}(v_{jm}v_{im}, v_{jn}v_{in}) = 0$ by iid sampling across $m$, assumption 1(iv).

*Assumption 4(g).* By iid sampling across $j, m$, it follow $\rho$ is the minimum eigenvalue of the $\mathrm{var}\,\tilde{x}_j$, which is positive by Assumption 1(ii)(b).

*Assumption 4(h).* From Proposition 2, we have that $\sup_{k\in K, m\in\mathbb{N}}\mathrm{P}(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta) \lesssim J^{-q}$, for every $q > 0$. Defining $\varepsilon_{km,J} := \zeta_{km,J} - \bar{\zeta}_{km,J}$ and $\alpha_{km,J} := \bar{\zeta}_{km,J} + \beta_p\sigma_k^{-1}\psi_{km,J}$ from Assumption 2(ii) yields the result, with $r(J) = J^{-q}$ for any $q > 0$.

**(Verification of Assumption 5).**

*Assumption 5(a).* Follows immediately from Assumption 1(i).

*Assumption 5(b).* Since sampling is iid across $m$ the limits simplifies to $\liminf_{J\to\infty}\mathrm{E}[(\alpha_{km,J} - \alpha_{\ell m,J})^2]$. The claim then follows from Proposition 3 proven below.

*Assumption 5(c).* Independence across $m$ follows immediately from independent sampling across markets, Assumption 1(iv). The orthogonality condition $\mathrm{E}[(\alpha_{km,J} - \alpha_{\ell m,J})v_{jm,J}] = 0$ is proven in Proposition 3 below.

*Assumption 5(d),(e).* Follows from uniform subgaussianity of $v$ and $x$, Assumption 1(ii)(c).

**(Verification of Assumption 6).**

*Assumption 6(a).* Immediate from assumption 1(ii) and assumption 2(ii).

*Assumption 6(b).* Under Assumption 1(ii) and (iii), the limit of the matrix is

$$\begin{pmatrix} \mathrm{E}[\tilde{x}_{jm}\tilde{x}_{jm}^\top] & 0 \\ 0 & I \end{pmatrix}.$$

Thus the condition is verified by Assumption 1(ii)(b).

*Assumption 6(c).* By iid sampling across $m$ and iid primitives across $j$, the limits simplifies $\liminf_{J\to\infty} \mathrm{E}[(\tilde{x}_{jm}^\top(\theta_k - \theta_\ell) + \alpha_{km,J} - \alpha_{\ell m,J})^2]$. The claim is then follows from Proposition 3(b) proven below.

*Assumption 6(d).* Note that

$$\mathrm{P}\left(\|\frac{1}{M}\sum_{m=1}^{M} v_{jm}\tilde{x}_{jm}\|_2 > \eta\right) \leqslant \mathrm{P}\left(\|\frac{1}{M}\sum_{m=1}^{M} \xi_{jm}\tilde{x}_{jm}\|_2 > \eta/2\right) + \mathrm{P}\left(\|\frac{1}{M}\sum_{m=1}^{M} \beta_p\omega_{jm}\tilde{x}_{jm}\|_2 > \eta/2\right)$$

Using subgaussianity of Assumptions 1 and 2 and orthogonality of Assumption 1(ii)(a) and Assumption 2(ii), both terms are $o(T^{-\delta})$ for any $\delta > 0$ by an Bernstein inequality. □

### B.4.1 Details on Assumption 4(b) and (h)

**Proposition 2.** *Recall the notation $\delta_{jm,J} := \beta_x' x_{jm} + \beta_p p_{jm,J} + \xi_{jm}$, and*

$$\zeta_{km,J} = \log \frac{\left(\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{k_j=k} e^{\delta_{jm,J}/\sigma_{kJ}}\right)^{\sigma_{kJ}-1}}{\sum_{\ell=1}^{K} J^{\sigma_{\ell J}-\sigma_{kJ}} \left(\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{k_j=\ell} e^{\delta_{jm,J}/\sigma_{\ell J}}\right)^{\sigma_{\ell J}}} \mathbb{1}\{\sum_{j=1}^{J} \mathbb{1}_{k_j=k} \geqslant 1\}$$

*If Assumptions 1 and 2 holds, then, for any $\eta > 0$, it holds that*

$$\sup_{k\in K, m\in\mathbb{N}} \mathrm{P}\left(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta\right) \lesssim \frac{1}{J^q}, \quad \textit{for every } q > 0$$

[25]*where*

$$\bar{\zeta}_{km,J} := \log \frac{\left(\varpi_{kJ}\, \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right]\right)^{\sigma_{kJ}-1}}{\sum_{\ell=1}^{K} J^{\sigma_{\ell J}-\sigma_{kJ}} \left(\varpi_{\ell J}\, \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{\ell J}}\right]\right)^{\sigma_{\ell J}}}$$

$$\varpi_{kJ} := \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{k_j=k},$$

*provided* $\sum_{j=1}^{J} \mathbb{1}_{k_j=k} \geqslant 1$. *Moreover* $\{\bar{\zeta}_{km,J} : k \in K, (m,J) \in \mathbb{N}^2\}$ *is bounded.*

*Proof.* Denote $S_{km,J} := \frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{k_j=k} e^{\delta_{jm,J}/\sigma_{kJ}}$ and $\mu_{km,J} := \varpi_{kJ}\, \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right]$, and note that $\mathrm{E}\, S_{km,J} = \mu_{km,J}$.

By Lemma 4(a), there is a positive number

$$c' := \min_{k\in K} \liminf_{J\to\infty} \mu_{km,J} > 0.$$

Consider the events $G_{mJ} := \{\max_{\ell\in K} |S_{\ell m,J} - \mu_{\ell m,J}| \leqslant c'/2\}$. Noting that

$$G_{mJ} \subset \{\min_{\ell\in K} S_{\ell m,J} > c'/2\},$$

and applying Lemma 8, which gives a Lipschitz constant for the map from $S_{mJ}, \mu_{mJ}$ to $\zeta_{km,J}$ and $\bar{\zeta}_{km,J}$, respectively, on domain bounded away from zero, we have that

$$\mathrm{P}\left(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta \;;\; G_{mJ}\right) \leqslant \mathrm{P}\left(\frac{4K}{c'} \max_{\ell\in K} |S_{\ell m,J} - \mu_{\ell m,J}| > \eta\right),$$

where $\frac{4K}{c'}$ is the uniform Lipschitz constant that applies for all $k, m$. We can thus upper-

[25] $a_J \lesssim J^{-q}$ for every $q$ means: for every $q$, there is a constant $C_q < \infty$ not depending on $J$ such that $a_J \leqslant C_q J^{-q}$.

bound $\mathrm{P}\left(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta\right)$ uniformly across $k, m$ by

$$\begin{aligned}
\mathrm{P}\left(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta\right) &\leqslant \mathrm{P}\left(|\zeta_{km,J} - \bar{\zeta}_{km,J}| > \eta \,;\, G_{mJ}\right) + P(G_{mJ}^c) \\
&\leqslant \mathrm{P}\left(\frac{4K}{c'} \max_{\ell \in K} |S_{\ell m,J} - \mu_{\ell m,J}| > \eta\right) + \mathrm{P}\left(\max_{\ell \in K} |S_{\ell m,J} - \mu_{\ell m,J}| > \frac{c'}{2}\right) \\
&\leqslant K\left(\sup_{\ell \in K, m \in \mathbb{N}} \mathrm{P}\left(|S_{\ell m,J} - \mu_{\ell m,J}| > \frac{c'}{4}\frac{\eta}{K}\right) + \sup_{\ell \in K, m \in \mathbb{N}} \mathrm{P}\left(|S_{\ell m,J} - \mu_{\ell m,J}| > \frac{c'}{2}\right)\right).
\end{aligned}$$

which is $J^{-q}$ by Lemma Lemma 4. The tail probability bound result then follows.

For boundedness, provided that $\sum_{j=1}^{J} \mathbb{1}_{k_j = k} \geqslant 1$, write

$$\bar{\zeta}_{km,J} = \log \frac{\left(\varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right]\right)^{\sigma_{kJ}}}{\sum_{\ell=1}^{K} J^{\sigma_{\ell J} - \sigma_{kJ}} \left(\varpi_{\ell J} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{\ell J}}\right]\right)^{\sigma_{\ell J}}} - \sigma_{kJ} \log\left(\varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right]\right).$$

Observe that by Lemma 4, there exists a bounded set bounded away form zero containing $\varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right]$. By uniform subgaussianity of $\delta_{jm,J}$ implied by assumption 1(ii), we have the uniform upper bound $\varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right] \leqslant \mathrm{E} \exp(\underline{\sigma}^{-1}\delta_{jm,J}) \leqslant \exp(\underline{\sigma}^{-1} \,\mathrm{E}\, \delta_{jm,J} + C\underline{\sigma}^{-2}\|\delta_{jm,J}\|_{\psi_2}^2) < \infty$, thus $\{\varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right] : k \in K, (m,J) \in \mathbb{N}^2\}$ is in a bounded subset of $\mathbb{R}_{++}$. Since $\sigma_{kJ} \to \sigma_*$ for all $k$, the weights $\{J^{\sigma_{\ell J} - \sigma_{kJ}} : J \in \mathbb{N}\}$ is also a bounded subset of $\mathbb{R}_{++}$. □

**Lemma 4.** *If assumption 1 holds, then*

*(a).* $\min_{k \in K} \liminf_{J \to \infty} \varpi_{kJ} \,\mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right] > 0$.

*(b).* $\mathrm{P}\left(|\frac{1}{J}\sum_{j=1}^{J} \mathbb{1}_{k_j=k}(e^{\delta_{jm,J}/\sigma_{kJ}} - \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right])| > u\right) \lesssim J^{-q}$ *for every* $q \geqslant 1$.

*Proof.* (a) Fix $k$. First note that $\liminf_{J\to\infty} \varpi_{kJ} = \varpi_k \in (0,1)$. Since $e^{\delta_{jm,J}/\sigma_{kJ}} > 0$ almost surely for every $J$ and $\sigma_{kJ} \to \sigma_* \in (0,1)$, it follows $\liminf_{J\to\infty} \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right] > 0$. Taking the minimum over $k \in K$ yields the result.

(b) The variables $\mathbb{1}_{k_j=k}(e^{\delta_{jm,J}/\sigma_{kJ}} - \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right])$ have $L^q$-norm $(q \geqslant 2)$ bounded by

$$\|\mathbb{1}_{k_j=k}(e^{\delta_{jm,J}/\sigma_{kJ}} - \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right])\|_{L^q} \leqslant 2\|e^{\delta_{jm,J}/\sigma_{kJ}}\|_{L^q}$$

Since $\sigma_{kJ} > \underline{\sigma} > 0$, and $\delta$ are subgaussian, we may further bound

$$\|e^{\delta_{jm,J}/\sigma_{kJ}}\|_{L^q}^q = \mathrm{E}\exp(\frac{q}{\sigma_{kJ}}\delta_{jm,J}) \leqslant \exp\left(\frac{q}{\underline{\sigma}}\,\mathrm{E}\,\delta_{jm,J} + C\left(\frac{q}{\underline{\sigma}}\right)^2 \|\delta_{jm,J}\|_{\psi_2}^2\right)$$

Hence

$$\|\mathbb{1}_{k_j=k}(e^{\delta_{jm,J}/\sigma_{kJ}} - \mathrm{E}\left[e^{\delta_{jm,J}/\sigma_{kJ}}\right])\|_{L^q}^q \leqslant 2^q \exp\left(\frac{q}{\underline{\sigma}}\,\mathrm{E}\,\delta_{jm,J} + C\left(\frac{q}{\underline{\sigma}}\right)^2 \|\delta_{jm,J}\|_{\psi_2}^2\right)$$

By applying lemma 7, the desired conclusion follows. Note that the bound is uniform across $m$ by Assumption 1(ii)(c) and Assumption 1(iii). □

### B.4.2 Details on verifying Assumptions 5(b), 5(c), and 6(c)

**Proposition 3.** *If Assumptions 1, 2 and 3 holds. For $\alpha_{km,J} := \beta_p\sigma_k^{-1}\psi_{km,J} + \bar{\zeta}_{km,J}$, It holds that for every $k \neq \ell$,*

*(a)* $\liminf_{J\to\infty} \mathrm{E}\left[(\alpha_{km,J} - \alpha_{\ell m,J})^2\right] > 0.$

*(b)* $\liminf_{J\to\infty} \mathrm{E}[(\tilde{x}_{jm}^\top(\theta_k - \theta_\ell) + \alpha_{km,J} - \alpha_{\ell m,J})^2] > 0.$

*(c)* $\mathrm{E}\left[(\alpha_{km,J} - \alpha_{\ell m,J})v_{jm,J}\right] = 0, \quad$ *for all $m$, $J$.*

*Proof.* Denote $\mu_{km,J} := \varpi_{kJ}\,\mathrm{E}[\mathbb{1}_{k_j=k}\,e^{\delta_{jm,J}/\sigma_{kJ}}]$, and observe that

$$\bar{\zeta}_{km,J} = \log\frac{J^{\sigma_{kJ}}{\mu_{km,J}}^{\sigma_{kJ}-1}}{\sum_{\ell=1}^K \left(J\mu_{\ell m,J}\right)^{\sigma_{\ell J}}},$$

so $\bar{\zeta}_{km,J} - \bar{\zeta}_{\ell m,J} = (\sigma_{kJ} - \sigma_{\ell J})\log J + (\sigma_{kJ} - 1)\log\mu_{km,J} - (\sigma_{kJ} - 1)\log\mu_{\ell m,J}$. Hence the difference in grouped FE is

$$\begin{aligned}\alpha_{km,J} - \alpha_{\ell m,J} := \beta_p(\sigma_{kJ}^{-1}\psi_{kmJ} - \sigma_{\ell J}^{-1}\psi_{\ell mJ}) \\ + (\sigma_{kJ} - \sigma_{\ell J})\log J + (\sigma_{kJ} - 1)\log\mu_{kmJ} - (\sigma_{\ell J} - 1)\log\mu_{\ell mJ}.\end{aligned}$$

**(a).** For the first claim, since it is assumed $\sigma_{kJ} \to \sigma_* \in (0,1)$ for all $k$, the previous display writes as

$$\alpha_{km,J} - \alpha_{\ell m,J} := \beta_p\sigma_*^{-1}(\psi_{kmJ} - \psi_{\ell mJ}) + (\sigma_* - 1)(\log\mu_{kmJ} - \log\mu_{\ell mJ}) + o(1).$$

Noting that $\mu_{kmJ}$ are non-stochastic, we have by assumption 2(iv)

$$\liminf_{J\to\infty} \mathrm{E}[(\alpha_{km,J} - \alpha_{\ell m,J})^2] \geqslant \beta_p \sigma_*^{-1} \liminf_{J\to\infty} \mathrm{var}(\psi_{kmJ} - \psi_{\ell mJ}) > 0.$$

**(b).** We have that

$$\theta_k - \theta_\ell = \sigma_*^{-1} \beta_p \begin{pmatrix} \pi_{k,J} - \pi_{\ell,J} \\ \gamma_{k,J} - \gamma_{\ell,J} \end{pmatrix} + o(1),$$

so

$$\begin{aligned} \tilde{x}_{jm}^\top(\theta_k - \theta_\ell) + \alpha_{km,J} - \alpha_{\ell m,J} = \beta_p \sigma_*^{-1} \big( (\psi_{kmJ} - \psi_{\ell mJ}) + x_{jm}^\top(\pi_{k,J} - \pi_{\ell,J}) + z_{jm}^\top(\gamma_{k,J} - \gamma_{\ell,J}) \big) \\ + (\sigma_* - 1)(\log \mu_{kmJ} - \log \mu_{\ell mJ}) + o(1). \end{aligned}$$

Therefore, by Assumption 2(iv),

$$\begin{aligned} \liminf_{J\to\infty} \mathrm{E}[(\tilde{x}_{jm}^\top(\theta_k - \theta_\ell) + \alpha_{km,J} - \alpha_{\ell m,J})^2] \geqslant \beta_p \sigma_*^{-1} \liminf_{J\to\infty} \mathrm{var} \left( \psi_{kmJ} - \psi_{\ell mJ} + x_{jm}^\top(\pi_{k,J} - \pi_{\ell,J}) + z_{jm}^\top(\gamma_{k,J} - \gamma \right. \\ > 0. \end{aligned}$$

**(c).** Expand

$$\begin{aligned} \mathrm{E}\left[(\alpha_{km,J} - \alpha_{\ell m,J}) v_{jm,J}\right] = \beta_p \sigma_{kJ}^{-1} \mathrm{E}\left[\psi_{kmJ} v_{jm,J}\right] - \beta_p \sigma_{\ell J}^{-1} \mathrm{E}\left[\psi_{\ell mJ} v_{jm,J}\right] \\ + \big( (\sigma_{kJ} - \sigma_{\ell J}) \log J + (\sigma_{kJ} - 1) \log \mu_{kmJ} - (\sigma_{\ell J} - 1) \log \mu_{\ell mJ} \big) \mathrm{E}\, v_{jm,J}. \end{aligned}$$

Now observe that $\mathrm{E}\left[\psi_{kmJ} v_{jm,J}\right] = \mathrm{E}\left[\psi_{kmJ}(\beta_p \sigma_{kjm,J}^{-1} \omega_{jm,J} + \xi_{jm})\right] = 0$ and $\mathrm{E}\, v_{jm,J} = 0$ by Assumption 2(i). Hence $\mathrm{E}\left[(\alpha_{km,J} - \alpha_{\ell m,J}) v_{jm,J}\right] = 0$ for all $k \neq \ell$ and $m, J$. □

## B.5 Technical Lemmata

**Lemma 5.** *If $X_1, X_2, \ldots$ have identical distributions. Then for every $n \in \mathbb{N}$,*

$$\log n + \mathrm{E}\, X_1 \leqslant \mathrm{E} \log \sum_{i=1}^n e^{X_i} \leqslant \log n + \log \mathrm{E}\, e^{X_1}.$$

*Proof.* $x \mapsto \log \sum_{i=1}^n \exp(x_i)$ is convex, thus by Jensen inequality and that $\mathrm{E}\, X_i = \mathrm{E}\, X_1$ for all

$i$,

$$\mathrm{E} \log \sum_{i=1}^{n} e^{X_i} \geqslant \log \sum_{i=1}^{n} e^{\mathrm{E}\, X_i} = \log n + \mathrm{E}\, X_1.$$

For the upper-bound, $x \mapsto \log x$ is concave, hence by Jensen and that $\mathrm{E}\, e^{X_i} = \mathrm{E}\, e^{X_1}$ for all $i$,

$$\mathrm{E} \log \sum_{i=1}^{n} e^{X_i} \leqslant \log \mathrm{E} \sum_{i=1}^{n} e^{X_i} = \log n + \log \mathrm{E}\, e^{X_1}.$$

□

**Lemma 6.** *Suppose $(Z_i, X_i)$ are independent across $i$, $Z_i$ are subgaussian random scalars and $X_i$ are subgaussian random vectors, then $Z_i X_i$ are sub-exponential random vectors.*

**Lemma 7.** *Let $X_1, X_2, \ldots$ be independent, mean-zero random variables with $\sup_{i \in \mathbb{N}} \mathrm{E}\, |X_i|^p < M < \infty$ for some $p \in [2, \infty)$. Then*

$$\mathrm{P}(|\frac{1}{n} \sum_{i=1}^{n} X_i| > \varepsilon) \leqslant \frac{(\sqrt{Cp}/\varepsilon)^p M}{n^{p/2}}$$

*Proof.* By the Marcinkiewicz-Zygmund inequality, for each $p \in [2, \infty)$ there exists a universal constant $C$ that does not depend on the distributions of $X_{1,} X_2, \ldots$, such that

$$\left( \mathrm{E}\, |\sum_{i=1}^{n} X_i|^p \right)^{2/p} \leqslant Cp \sum_{i=1}^{n} (\mathrm{E}\, |X_i|^p)^{2/p}.$$

By the uniform bound of $\mathrm{E}\, |X_i|^p$, we have that

$$\mathrm{E}\, |\frac{1}{n} \sum_{i=1}^{n} X_i|^p \leqslant \frac{(CpnM^{2/p})^{p/2}}{n^p} = \frac{(\sqrt{Cp})^p M}{n^{p/2}}.$$

By Markov's inequality, we conclude

$$\mathrm{P}(|\frac{1}{n} \sum_{i=1}^{n} X_i| > \varepsilon) = \mathrm{P}(|\frac{1}{n} \sum_{i=1}^{n} X_i|^p > \varepsilon^p) \leqslant \varepsilon^{-p}\, \mathrm{E}\, |\frac{1}{n} \sum_{i=1}^{n} X_i|^p \leqslant \frac{(\sqrt{Cp}/\varepsilon)^p M}{n^{p/2}}.$$

□

**Lemma 8.** *(Lipschitz properties). Let $c, c', c'' > 0$ and $C', C'' < \infty$.*

*(a) The function $f : (c', C')^K \to \mathbb{R}$ where $\alpha \in (c'', C'')^K, \sigma \in (c, 1)^K$ defined by*

$$f(x) = \log \frac{x_k^{\sigma_k - 1}}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}}$$

*satisfies*

$$|f(x) - f(y)| \leqslant \frac{2K}{c'} \|x - y\|_\infty.$$

*(b) The function $g : (c, 1)^K \to \mathbb{R}$ where $\alpha \in (c'', C'')^K, x \in (c', C')^K$ defined by*

$$g(\sigma) = \log \frac{x_k^{\sigma_k}}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}}$$

*satisfies*

$$|g(\sigma) - g(\tau)| \leqslant K |\log C'|' \|\sigma - \tau\|_\infty.$$

*(c) The function $h : (c'', C'')^K \to \mathbb{R}$ where $\sigma \in (c, 1)^K, x \in (c', C')^K$ defined by*

$$h(\alpha) = \log \sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}$$

*satisfies*

$$|h(\alpha) - h(\beta)| \leqslant \frac{1}{c''} \|\alpha - \beta\|_\infty.$$

*(d) For $\sigma, \tau \in (c, 1)^2$, and $X \in L^{2/c}$ such that $X \geqslant c'$ almost surely, it holds*

$$\left| \mathrm{E}\, X^{1/\sigma} - \mathrm{E}\, X^{1/\tau} \right| \leqslant \left( \frac{\log c'}{c} + \frac{\sqrt{\mathrm{E}\,|X|^{2/c}} \sqrt{\mathrm{E}(\log X)^2}}{c^2} \right) . |\sigma - \tau|.$$

*Proof.* (a) The partial derivatives of $f$ are

$$\partial_i f(x) = \begin{cases} \left( \frac{1 - \alpha_i x_i^{\sigma_i}}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}} \right) \frac{\sigma_i}{x_i} - \frac{1}{x_i} & i = k \\ - \frac{\alpha_i x_i^{\sigma_\ell}}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}} \frac{\sigma_i}{x_i} & i \neq k \end{cases}$$

Observe that $|\partial_\ell f(x)| \leqslant 2/x_\ell$, for every $\ell$, so by the mean value theorem, for some $\xi$ between $x$ and $y$ element-wise, $f(x) - f(y) = \nabla f(\xi)^\top (x - y)$. Applying Hölder inequality we get the

desired bound

$$|f(x) - f(y)| \leqslant \sup_{\xi \in (c', C')} \|\nabla f(\xi)\|_1' \|x - y\|_\infty \leqslant \frac{2K}{c'} \|x - y\|_\infty.$$

(b) Calculate

$$\partial_i g(\sigma) = \begin{cases} \log x_k - \frac{\alpha_i x_i^{\sigma_i} \log x_i}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}} & i = k \\ -\frac{\alpha_i x_i^{\sigma_i} \log x_i}{\sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell}} & i \neq k \end{cases},$$

so

$$\|\nabla g(\sigma)\|_1 = \sum_{\ell=1}^K w_\ell |\log x_j| + (1 - w_k) |\log x_k|$$

where the weights are $w_i = \alpha_i x_i^{\sigma_i} / \sum_{\ell=1}^K \alpha_\ell x_\ell^{\sigma_\ell} > 0$ and $\sum_i w_i = 1$. It is easy to see $\sup_{\xi \in (c,1)} \|\nabla g(\xi)\|_1 \leqslant K |\log C'|$ where $C'$ is the uniform component bound on $x$. Thus by the mean value theorem and Hölder's inequality,

$$|g(\sigma) - g(\tau)| \leqslant K |\log C'|' \|\sigma - \tau\|_\infty$$

(c) Calculate

$$\partial_k h(\alpha) = \frac{x_k^{\sigma_k}}{\sum_\ell \alpha_\ell x_\ell^{\sigma_\ell}},$$

so $\|\nabla h(\xi)\|_1 = \sum_\ell x_\ell^{\sigma_\ell} / \sum_\ell \alpha_\ell x_\ell^{\sigma_\ell} \leqslant 1/c''$ for every $\xi \in (c'', C'')$. By the mean value theorem and Hölder's inequality,

$$|h(\alpha) - h(\beta)| \leqslant \frac{1}{c''} \|\alpha - \beta\|_\infty.$$

(d) Define $\varphi(\sigma) = \mathrm{E}\, X^{1/\sigma} = \mathrm{E} \exp(\frac{1}{\sigma} \log X)$. Since $X^{1/\sigma} \leqslant X^{2/c}$ with $\mathrm{E}\, X^{2/c} < \infty$ by hypothesis, and $\sigma \mapsto \exp(\frac{1}{\sigma} \log x)$ is differentiable, by an application of dominated convergence we can interchange the order of integration and differentiation

$$\varphi'(\sigma) = \mathrm{E} \left[ \frac{1}{\sigma^2} X^{1/\sigma} |\log X| \right].$$

By the mean-value theorem,

$$|\varphi(\sigma) - \varphi(\tau)| \leqslant \sup_{\xi \in (c,1)} |\varphi'(\xi)|' |\sigma - \tau|.$$

We can bound the derivative for $\xi \in (c, 1)$ uniformly

$$\begin{aligned}
|\varphi'(\xi)| &= \mathrm{E}\left[\frac{1}{\xi^2} X^{1/\xi} |\log X|\right]. \\
&\leqslant \mathrm{E}\left[\frac{|\log X|}{c^2}\, \mathbb{1}_{\{c' \leqslant X \leqslant 1\}}\right] + \mathrm{E}\left[\frac{X^{1/c} |\log X|}{c^2}\, \mathbb{1}_{\{X > 1\}}\right] \\
&\leqslant \frac{\log c'}{c} + \frac{\sqrt{\mathrm{E}\, X^{2/c}}\sqrt{\mathrm{E}(\log X)^2}}{c^2}.
\end{aligned}$$

where $\mathrm{E}(\log X)^2 \leqslant \mathrm{E}\, X^{2/c} < \infty$. □

# C Appendix for the Empirical Applications

## C.1 Classification Algorithms

The classification problems in (12), (13), and (14) require optimizing not only over regression coefficients but also over partitions of products. The number of possible partitions grows exponentially with the number of products, making direct optimization computationally demanding and, in some cases, infeasible. To address this issue, we extend the algorithm proposed by Bonhomme and Manresa (2015) to accommodate the different empirical models considered in Section 2.

**Case 1: No price partition or identical partitions.** We first consider the case in which the partition of prices coincides with the partition of products, $k'(j) = k(j)$ for all $j$, or when there is effectively no price partition, i.e., $|\text{supp}(k')| = 1$. In this setting, we propose the following iterative algorithm to solve eq. (12):

1. Let $(\boldsymbol{\theta}^0, \boldsymbol{\alpha}^0) \equiv (\theta^{1,0}, \ldots, \theta^{K,0}, \alpha_1^{1,0}, \ldots, \alpha_M^{K,0})$ denote an initial set of parameter values.

2. Given $(\boldsymbol{\theta}^s, \boldsymbol{\alpha}^s)$, compute for all $j \in J$:

$$k(j)^{s+1} = \underset{k \in \{1,\ldots,K\}}{\arg\min} \sum_{m=1}^{M} \left( p_{jm} - (\theta_z^{k,s} z_{jm} + \theta_x^{k,s} \mathbf{x}_{jm} + \alpha_m^{k,s}) \right)^2,$$

   which determines the updated grouping structure $\mathcal{K}^{s+1}$.

3. Given the updated grouping structure $\mathcal{K}^{s+1}$, compute:

$$(\boldsymbol{\theta}^{s+1}, \boldsymbol{\alpha}^{s+1}) = \underset{\theta^1,\ldots,\theta^K,\alpha_1^1,\ldots,\alpha_M^K}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \left( \log \mathbb{P}_{jm} - (\theta_z^{k(j),s+1} z_{jm} + \theta_x^{k(j),s+1} \mathbf{x}_{jm} + \alpha_m^{k(j),s+1}) \right)^2.$$

4. Repeat steps 2 and 3 until convergence of the parameters.

In practice, this problem is not globally convex, as is common in clustering problems. We therefore recommend running the algorithm from multiple starting values and selecting the solution that achieves the lowest objective value.

**Case 2: Different partitions.** We now consider the case in which $k'(j) \neq k(j)$ for some $j$ and $|\text{supp}(k')| > 1$. In this case, we proceed in two stages. First, we estimate the partition of prices.

1. Let $(\boldsymbol{\gamma}^0, \boldsymbol{\kappa}^0) \equiv (\gamma^{1,0}, \ldots, \gamma^{K',0}, \kappa_1^{1,0}, \ldots, \kappa_M^{K',0})$ denote an initial set of parameter values.

2. Given $(\boldsymbol{\gamma}^s, \boldsymbol{\kappa}^s)$, compute for all $j \in J$:

$$k'(j)^{s+1} = \underset{k \in \{1,\ldots,K'\}}{\arg\min} \sum_{m=1}^{M} \Big( \log p_{jm} - (\gamma_z^{k,s} z_{jm} + \gamma_x^{k,s} \mathbf{x}_{jm} + \kappa_m^{k,s}) \Big)^2,$$

   which determines the updated grouping structure $\mathcal{K}'^{s+1}$.

3. Given the updated grouping structure $\mathcal{K}'^{s+1}$, compute

$$(\boldsymbol{\gamma}^{s+1}, \boldsymbol{\kappa}^{s+1}) = \underset{\gamma^1,\ldots,\gamma^{K'},\kappa_1^1,\ldots,\kappa_M^{K'}}{\arg\min} \sum_{j=1}^{J} \sum_{m=1}^{M} \Big( \log p_{jm} - (\gamma_z^{k'(j),s+1} z_{jm} + \gamma_x^{k'(j),s+1} \mathbf{x}_{jm} + \kappa_m^{k'(j),s+1}) \Big)^2.$$

4. Repeat until convergence of the parameters.

After obtaining the price partition and coefficients, we compute the predicted prices

$$\hat{p}_{jm} = \hat{\gamma}_z^{k'(j)} z_{jm} + \hat{\gamma}_x^{k'(j)} \mathbf{x}_{jm} + \hat{\kappa}_m^{k'(j)}.$$

We then estimate the grouping structure and coefficients of the demand equation using the following algorithm:

1. Let $(\boldsymbol{\beta}^0, \boldsymbol{\lambda}^0) \equiv (\beta^{1,0}, \ldots, \beta^{K,0}, \lambda_1^{1,0}, \ldots, \lambda_M^{K,0})$ denote an initial set of parameter values.

2. Given $(\boldsymbol{\beta}^s, \boldsymbol{\lambda}^s)$, compute for all $j \in J$:

$$k(j)^{s+1} = \underset{k \in \{1,\ldots,K\}}{\arg\min} \sum_{m=1}^{M} \Big( \log \mathbb{P}_{jm} - (\beta_p^{k,s} \hat{p}_{jm} + \beta_x^{k,s} \mathbf{x}_{jm} + \lambda_m^{k,s}) \Big)^2,$$

   which determines the updated grouping structure $\mathcal{K}^{s+1}$.

3. Given the updated grouping structure $\mathcal{K}^{s+1}$, compute

$$(\boldsymbol{\beta}^{s+1}, \boldsymbol{\lambda}^{s+1}) = \underset{\beta^1,\ldots,\beta^K,\lambda_1^1,\ldots,\lambda_M^K}{\arg\min} \sum_{j=1}^{J}\sum_{m=1}^{M} \Big( \log \mathbb{P}_{jm} - (\beta_p^{k(j),s+1}\hat{p}_{jm} + \beta_x^{k(j),s+1}\mathbf{x}_{jm} + \lambda_m^{k(j),s+1}) \Big)^2.$$

4. Repeat until convergence of the parameters.

As before, because these optimization problems are not globally convex, we recommend running the algorithm from multiple starting values and selecting the solution that attains the lowest objective value.

## C.2 Robustness Checks on Group Classification

### C.2.1 Results using Hausman instruments for prices

Table A1: Average group characteristics for $k^* = 4$ and Hausman instrument

| Characteristics | Unconditional values | Conditional values by group | | | |
|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 |
| Unit size (ml) | 354.882 | 354.882 | 473.176 | 354.882 | 354.882 |
| Unit quantity | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 |
| Unit price | 10.209 | 10.295 | 9.590 | 10.474 | 9.604 |
| Price per ml | 0.004 | 0.004 | 0.003 | 0.004 | 0.004 |
| Share of domestic | 0.794 | 0.767 | 0.836 | 0.833 | 0.879 |
| Share of ale | 0.158 | 0.102 | 0.226 | 0.287 | 0.339 |
| Share of regular beer | 0.442 | 0.437 | 0.437 | 0.496 | 0.358 |
| Share of stout and porter | 0.015 | 0.013 | 0.016 | 0.020 | 0.027 |
| Share of light beer | 0.369 | 0.437 | 0.297 | 0.171 | 0.256 |
| Share of other types | 0.016 | 0.011 | 0.024 | 0.026 | 0.020 |
| Share top-ten brands | 0.624 | 0.748 | 0.459 | 0.365 | 0.357 |
| # of products | 2806 | 271 | 692 | 1119 | 724 |

*Notes:* Rows 1–4 report median values of unit size, unit quantity, unit price, and price per ml, presented both unconditionally and conditional on group membership. Rows 5–11 report unconditional and conditional mean shares. Groups are estimated using Hausman instruments for prices instead of the baseline Gandhi-Houde instruments.

Table A2: Estimation Results for Nested Logit with $k^*$ = 4

| | OLS | IV |
|---|---|---|
| log unit size | 0.107 (0.006)*** | 0.431 (0.018)*** |
| log upc quantity | 0.121 (0.005)*** | 0.586 (0.020)*** |
| log price | -0.084 (0.006)*** | -0.581 (0.021)*** |
| $\sigma_1$ | 0.330 (0.002)*** | 0.962 (0.022)*** |
| $\sigma_2$ | 0.127 (0.001)*** | 0.662 (0.019)*** |
| $\sigma_3$ | -0.025 (0.001)*** | 0.469 (0.018)*** |
| $\sigma_4$ | -0.337 (0.001)*** | 0.178 (0.019)*** |
| Mean Own-price Elasticity | 1.193 (0.003) | -1.612 (0.004) |
| Mean Cross-price Elasticity | -0.003 (0.008) | 0.004 (0.011) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | hausman |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Note:* The table reports OLS and IV estimates of a nested logit demand model where products are grouped by beer type, yielding $k = 4$ nests. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and four group-specific nesting parameters ($\sigma_g$). The IV specification instruments for log price using Hausman IVs and for the within-group log share using the exogenous share projection. Market fixed effects are included. Mean own-price and cross-price elasticities are computed from the IV estimates and weighted by national unit sales. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

### C.2.2 Results for $k^* = 3$ and $k^* = 5$

Table A3: Characteristics of beer across groups ($k = 3$)

| Characteristics | Unconditional values | Conditional values by group 1 | 2 | 3 |
|---|---|---|---|---|
| Unit size (ml) | 354.882 | 354.882 | 354.882 | 354.882 |
| Unit quantity | 6.000 | 6.000 | 6.000 | 6.000 |
| Unit price | 10.209 | 10.180 | 10.232 | 11.138 |
| Price per ml | 0.004 | 0.004 | 0.004 | 0.004 |
| Share of domestic | 0.794 | 0.781 | 0.832 | 0.792 |
| Share of ale | 0.158 | 0.107 | 0.278 | 0.376 |
| Share of regular beer | 0.442 | 0.451 | 0.414 | 0.462 |
| Share of stout and porter | 0.015 | 0.012 | 0.018 | 0.039 |
| Share of light beer | 0.369 | 0.418 | 0.263 | 0.110 |
| Share of other types | 0.016 | 0.013 | 0.026 | 0.013 |
| Share top-ten brands | 0.624 | 0.725 | 0.372 | 0.332 |
| # of products | 2806 | 393 | 1119 | 1294 |

*Notes:* Rows 1–4 report median values of unit size, unit quantity, unit price, and price per ml, presented both unconditionally and conditional on group membership. Rows 5–11 report unconditional and conditional mean shares. Groups are estimated using Gandhi-Houde instruments.

Figure A1: Nested logit ($k^* = 3$)




Logos are trademarks of their respective owners

*Notes:* This figure shows the top ten most national sold beer brands at the national level for each group. The bubble size of each brand reflects the relative size of the market share in each particular group.

Table A4: Estimation Results for Nested Logit with $k = 3$

| | OLS | IV |
|---|---|---|
| log unit size | 0.059 (0.006)*** | 0.681 (0.040)*** |
| log upc quantity | 0.067 (0.004)*** | 0.824 (0.042)*** |
| log price | -0.067 (0.005)*** | -1.038 (0.056)*** |
| $\sigma_1$ | 0.272 (0.001)*** | 0.968 (0.024)*** |
| $\sigma_2$ | 0.047 (0.001)*** | 0.649 (0.022)*** |
| $\sigma_3$ | -0.204 (0.001)*** | 0.391 (0.022)*** |
| Mean Own-price Elasticity | -0.457 (0.001) | -2.010 (0.002) |
| Mean Cross-price Elasticity | 0.001 (0.003) | 0.002 (0.006) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Note:* The table reports OLS and IV estimates of a nested logit demand model where products are grouped by beer type (regular beer/lager, light beer, ale, stout/porter, and other), yielding $k = 3$ nests. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and five group-specific nesting parameters ($\sigma_g$). The IV specification instruments for log price using Gandhi-Houde differentiation IVs and for the within-group log share using the exogenous share projection. Market fixed effects are included. Mean own-price and cross-price elasticities are computed from the IV estimates and weighted by national unit sales. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

Table A5: Characteristics of beer across groups ($k = 5$)

| Characteristics | Unconditional values | Conditional values by group | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Unit size (ml) | 354.882 | 473.176 | 354.882 | 354.882 | 354.882 | 354.882 |
| Unit quantity | 6.000 | 6.000 | 6.000 | 6.000 | 6.000 | 4.000 |
| Unit price | 10.209 | 8.831 | 10.194 | 11.934 | 10.862 | 10.150 |
| Price per ml | 0.004 | 0.003 | 0.004 | 0.004 | 0.004 | 0.005 |
| Share of domestic | 0.794 | 0.838 | 0.756 | 0.866 | 0.778 | 0.816 |
| Share of ale | 0.158 | 0.188 | 0.090 | 0.318 | 0.337 | 0.411 |
| Share of regular beer | 0.442 | 0.438 | 0.446 | 0.434 | 0.449 | 0.364 |
| Share of stout and porter | 0.015 | 0.017 | 0.012 | 0.016 | 0.029 | 0.061 |
| Share of light beer | 0.369 | 0.333 | 0.441 | 0.209 | 0.172 | 0.146 |
| Share of other types | 0.016 | 0.024 | 0.011 | 0.023 | 0.013 | 0.017 |
| Share top-ten brands | 0.624 | 0.537 | 0.758 | 0.337 | 0.378 | 0.227 |
| # of products | 2806 | 445 | 197 | 727 | 878 | 559 |

*Notes:* Rows 1–4 report median values of unit size, unit quantity, unit price, and price per ml, presented both unconditionally and conditional on group membership. Rows 5–11 report unconditional and conditional mean shares. Groups are estimated using Gandhi-Houde instruments.




Figure A2: Nested logit ($k^* = 5$)

*Notes:* This figure shows the top ten most national sold beer brands at the national level for each group. The bubble size of each brand reflects the relative size of the market share in each particular group.

Table A6: Estimation Results for Nested Logit with $k = 5$

| | OLS | IV |
|---|---|---|
| log unit size | 0.087 (0.007)*** | 0.196 (0.027)*** |
| log upc quantity | 0.103 (0.005)*** | 0.241 (0.027)*** |
| log price | -0.037 (0.007)*** | -0.084 (0.036)** |
| $\sigma_1$ | 0.202 (0.002)*** | 0.698 (0.016)*** |
| $\sigma_2$ | 0.407 (0.002)*** | 1.000 (0.019)*** |
| $\sigma_3$ | 0.060 (0.001)*** | 0.516 (0.015)*** |
| $\sigma_4$ | -0.096 (0.001)*** | 0.340 (0.015)*** |
| $\sigma_5$ | -0.442 (0.002)*** | 0.029 (0.018) |
| Mean Own-price Elasticity | -0.057 (0.001) | -0.729 (0.004) |
| Mean Cross-price Elasticity | 0.000 (0.002) | 0.003 (0.011) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Note:* The table reports OLS and IV estimates of a nested logit demand model where products are grouped by beer type (regular beer/lager, light beer, ale, stout/porter, and other), yielding $k = 5$ nests. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and five group-specific nesting parameters ($\sigma_g$). The IV specification instruments for log price using Gandhi-Houde differentiation IVs and for the within-group log share using the exogenous share projection. Market fixed effects are included. Mean own-price and cross-price elasticities are computed from the IV estimates and weighted by national unit sales. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

### C.2.3 Classification including Beer Type Indicators

Table A7: Characteristics of beer across groups $k^* = 3$

| Characteristics | Unconditional values | Conditional values by group 1 | 2 | 3 |
|---|---|---|---|---|
| Unit size (ml) | 354.882 | 473.176 | 354.882 | 354.882 |
| Unit quantity | 6.000 | 6.000 | 6.000 | 6.000 |
| Unit price | 10.209 | 9.857 | 10.265 | 11.020 |
| Price per ml | 0.004 | 0.004 | 0.004 | 0.004 |
| Share of domestic | 0.794 | 0.799 | 0.793 | 0.777 |
| Share of ale | 0.158 | 0.227 | 0.126 | 0.336 |
| Share of regular beer | 0.442 | 0.448 | 0.437 | 0.506 |
| Share of stout and porter | 0.015 | 0.016 | 0.013 | 0.041 |
| Share of light beer | 0.369 | 0.283 | 0.411 | 0.102 |
| Share of other types | 0.016 | 0.026 | 0.013 | 0.016 |
| Share top-ten brands | 0.624 | 0.413 | 0.710 | 0.349 |
| # of products | 2806 | 1116 | 420 | 1270 |

*Notes:* This table shows the unconditional and conditional values of product characteristics across groups. Unit size, unit quantity, unit price, and price per ml (rows 1 to 4) are calculated as median. Rows 5 to 11 are unconditional and conditional (mean) shares.

Table A8: Estimation Results for Nested Logit with $k^* = 3$

| | OLS | IV |
|---|---|---|
| log unit size | 0.037 (0.006)*** | 0.088 (0.024)*** |
| log upc quantity | 0.040 (0.005)*** | 0.122 (0.024)*** |
| log price | -0.040 (0.006)*** | -0.146 (0.031)*** |
| $\sigma_1$ | 0.046 (0.001)*** | 0.303 (0.009)*** |
| $\sigma_2$ | 0.269 (0.001)*** | 0.578 (0.010)*** |
| $\sigma_3$ | -0.209 (0.001)*** | 0.031 (0.009)*** |
| Mean Own-price Elasticity | -0.276 (0.001) | -2.343 (0.006) |
| Mean Cross-price Elasticity | 0.000 (0.002) | 0.005 (0.015) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Note:* The table reports OLS and IV estimates of a nested logit demand model where products are grouped by beer type (regular beer/lager, light beer, ale, stout/porter, and other), yielding $k = 3$ nests. Importantly, these results extend for beer type indicators. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and five group-specific nesting parameters ($\sigma_g$). The IV specification instruments for log price using Gandhi-Houde differentiation IVs and for the within-group log share using the exogenous share projection. Market fixed effects are included. Mean own-price and cross-price elasticities are computed from the IV estimates and weighted by national unit sales. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$.

### C.2.4 K-means clustering only on characteristics

Table A9: Characteristics of beer across groups $k^* = 4$

| Characteristics | Unconditional values | Conditional values by group 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Unit size (ml) | 354.882 | 473.176 | 354.882 | 473.176 | 739.338 |
| Unit quantity | 6.000 | 6.000 | 12.000 | 4.000 | 1.000 |
| Unit price | 10.209 | 9.544 | 15.672 | 8.831 | 2.693 |
| Price per ml | 0.004 | 0.003 | 0.003 | 0.003 | 0.004 |
| Share of domestic | 0.794 | 0.913 | 0.748 | 0.863 | 0.741 |
| Share of ale | 0.158 | 0.182 | 0.113 | 0.279 | 0.124 |
| Share of regular beer | 0.442 | 0.449 | 0.395 | 0.443 | 0.524 |
| Share of stout and porter | 0.015 | 0.021 | 0.017 | 0.015 | 0.004 |
| Share of light beer | 0.369 | 0.316 | 0.469 | 0.246 | 0.324 |
| Share of other types | 0.016 | 0.032 | 0.006 | 0.017 | 0.024 |
| Share top-ten brands | 0.624 | 0.449 | 0.726 | 0.435 | 0.712 |
| # of products | 2806 | 740 | 463 | 1424 | 179 |

*Notes:* This table shows the unconditional and conditional values of product characteristics across groups. Unit size, unit quantity, unit price, and price per ml (rows 1 to 4) are calculated as median. Rows 5 to 11 are unconditional and conditional (mean) shares.

Table A10: Estimation Results for Nested Logit with $k^* = 4$

| | OLS | IV |
|---|---|---|
| log unit size | -0.019 (0.004)*** | 0.311 (0.018)*** |
| log upc quantity | 0.032 (0.003)*** | 0.335 (0.017)*** |
| log price | 0.004 (0.004) | -0.423 (0.024)*** |
| $\sigma_1$ | -0.044 (0.001)*** | 0.113 (0.005)*** |
| $\sigma_2$ | 0.146 (0.001)*** | 0.330 (0.006)*** |
| $\sigma_3$ | -0.079 (0.001)*** | 0.066 (0.005)*** |
| $\sigma_4$ | 0.047 (0.001)*** | 0.220 (0.006)*** |
| Mean Own-price Elasticity | -0.036 (0.000) | -4.566 (0.010) |
| Mean Cross-price Elasticity | 0.000 (0.000) | 0.011 (0.026) |
| Market Fixed Effects | ✓ | ✓ |
| IV Type | | $GH_{\text{units}}, GH_{\text{size}}$ |
| Number of Products | 2806 | |
| Number of Observations | 64497 | |

*Notes:* The table reports OLS and IV estimates of the nested logit demand model with $k^* = 4$ groups, where the first-step classification uses standard $k$-means clustering on mean observed product characteristics (unit size, quantity, price). The number of groups $k^* = 4$ is selected via the elbow method applied to the within-cluster sum of squares, using multiple random initializations. The dependent variable is $\log \mathbb{P}_{jm} - \log \mathbb{P}_{0m}$. Regressors include log unit size (ml), log UPC quantity, log price, and four group-specific nesting parameters $(\sigma_1, \ldots, \sigma_4)$. The IV specification instruments for log price using Gandhi-Houde differentiation IVs and for $\log \mathbb{P}_{j|g(j),m}$ using the exogenous share projection. Market (DMA-quarter) fixed effects are included. Standard errors are reported in parentheses. Mean own-price and cross-price elasticities are computed from the IV estimates, weighted by total unit state sales. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.